Swami Keshvanand Institute of Technology, Management &Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

A

Course File

on

(Machine Learning, 6CAI4-02)

Programme: B.Tech (CSE-AI)

Semester: VI

Session: 2023-24

(Abha Jain)
(Assistant Professor)
(Computer Science and Engineering)

Contents

Vision and Mission of Institute

Vision: "To promote higher learning in technology and industrial research to make our country a global player."

Mission: "To promote quality education, training and research in the field of engineering by establishing effective interface with industry and to encourage the faculty to undertake industry sponsored projects for the students."

Quality Policy

We are committed to **'achievement of quality'** as an integral part of our institutional policy by continuous self-evaluation and striving to improve ourselves.

Institute would pursue quality in

- All its endeavors like admissions, teaching- learning processes, examinations, extra and co-curricular activities, industry institution interaction, research & development, continuing education, and consultancy.
- Functional areas like teaching departments, Training & Placement Cell, library, administrative office, accounts office, hostels, canteen, security services, transport, maintenance section and all other services."

Vision of CSE Department

Vision of CSE department is to:

V1: Produce quality computer engineers trained in latest tools and technologies.

V2: Be a leading department in the region and country by imparting in-depth knowledge to the students in emerging technologies in computer science & engineering.

Mission of CSE Department

Mission of CSE department is to:

Deliver resources in IT enable domain through:

M1: Effective Industry interaction and project-based learning.

M2: Motivating our students for employability, entrepreneurship, research and higher education.

M3: Providing excellent engineering skills in a state-of-the art infrastructure.

RTU Scheme & Syllabus

RAJASTHAN TECHNICAL UNIVERSITY, KOTA

Syllabus

III Year-VI Semester: B.Tech. Computer Science and Engineering(AI)

6CAI4-02:Machine Learning

Credit: 3 **Max. Marks: 150(IA:30, ETE:120)**

3L+0T+0P **End Term Exam: 3 Hours**

SN	Contents	Hours
1	**Introduction:** Objective, scope and outcome of the course.	**01**
2	**Supervised learning algorithm:** Introduction, types of learning, application, Supervised learning: Linear Regression Model, Naive Bayes classifier Decision Tree, K nearest neighbor, Logistic Regression, Support Vector Machine, Random forest algorithm	**09**
3	**Unsupervised learning algorithm:** Grouping unlabelled items using k-means clustering, Hierarchical Clustering, Probabilistic clustering, Association rule mining, Apriori Algorithm, f-p growth algorithm, Gaussian mixture model.	**08**
4	**Introduction to Statistical Learning Theory**, Feature extraction - Principal component analysis, Singular value decomposition. Feature selection – feature ranking and subset selection, filter, wrapper and embedded methods, Evaluating Machine Learning algorithms and Model Selection.	**08**
5	**Semi supervised learning, Reinforcement learning:** Markov decision process (MDP), Bellman equations, policy evaluation using Monte Carlo, Policy iteration and Value iteration, Q-Learning, State- Action-Reward-State-Action (SARSA), Model-based Reinforcement Learning.	**08**
6	**Recommended system,** Collaborative filtering, Content-based filtering Artificial neural network, Perceptron, Multilayer network, Backpropagation, Introduction to Deep learning.	**08**
	Total	**42**

Prerequisite of Course

Mathematics:

- Linear Algebra: Vectors, matrices, matrix operations, eigenvalues, eigenvectors.
- Calculus: Differentiation, integration, partial derivatives.
- Probability and Statistics: Probability distributions, mean, median, mode, variance, standard deviation, hypothesis testing, probability density functions, cumulative distribution functions.

Basic Machine Learning Concepts:

- Familiarity with fundamental machine learning concepts such as supervised learning, unsupervised learning, and reinforcement learning.
- Understanding of common machine learning tasks like classification, regression, clustering, and dimensionality reduction.

Optional:

- Knowledge of basic algorithms and data structures can be beneficial.
- Experience with basic database querying and manipulation can also be helpful.
- Exposure to tools and libraries commonly used in machine learning such as scikit-learn, TensorFlow, PyTorch (Python), or caret, keras (R).

Problem-Solving Skills:

- Ability to think critically and analytically to solve problems.
- Strong logical reasoning skills.

List of Text and Reference Books

Machine Learning Text and Reference Books

1. A First Course in Machine Learning (Machine Learning & Pattern Recognition) 2nd Edition,by Simon Rogers, Mark Girolami; CRC Press
2. Applied Machine Learning , M.Gopla, McGrawHill
3. Machine Learning, Saikat Dutt, Shubramanian Chandramouli, Amit Kumar Das, Pearson Publication
4. Learning From Data by Yaser S. Abu-Mostafa , Malik Magdon-Ismail, Hsuan-Tien Lin; AMLBook
5. Fundamentals of Machine Learning, John D. Miller, MIT Press
6. An Introduction to Statistical Learning, Gareth James, Robert Tribshirani, Springer
7. Machine Learning An Algorithmic Perspective, II Edition, Stephan Marsland, CRC Press.

Time Table

Teacher :Abha Jain

MONDAY		9:00 AM-11:00 AM 6CS(AI)B - All Groups Campus Recruitment Training (Lab) 105		11:30 AM-2:30 PM 6CS(AI)B - G2 Machine Learning Lab (Lab) Computer Lab 10 (Ground Floor)	
TUESDAY	8:00 AM-9:00 AM 6CS(AI)B - All Groups Machine Learning (Lecture) 104		10:00 AM-11:00 AM 6CS(AI)A - All Groups Machine Learning (Lecture) 304		
WEDNESDAY				11:30 AM-12:30 PM 6CS(AI)B - All Groups Machine Learning (Lecture) 406	
THURSDAY		9:00 AM-10:00 AM 6CS(AI)B - All Groups Machine Learning (Lecture) 302		11:30 AM-2:30 PM 6CS(AI)B - G1 Machine Learning Lab (Lab) Computer Lab 10 (Ground Floor)	
FRIDAY					12:30 PM-1:30 PM 6CS(AI)A - All Groups Machine Learning (Lecture) 104
SATURDAY	8:00 AM-11:00 AM 6CS(AI)A - G1 Machine Learning Lab (Lab) Computer Lab 10 (Ground Floor)				12:30 PM-1:30 PM 6CS(AI)A - All Groups Machine Learning (Lecture) 302

Syllabus Deployment: Course Plan & Coverage

Lecture Number	Topics to be covered
Unit-1	**Introduction to Machine learning**
1	Objective, Scope, Outcome of the course
Unit-2	**Supervised learning algorithm**
2	Introduction
3	Types of learning and Application
4	Supervised learning: Linear Regression Model
5	Naive Bayes classifier
6	Decision Tree
7	K nearest neighbor
8	Logistic Regression
9	Support Vector Machine
10	Random forest algorithm
Unit-3	**Unsupervised learning algorithm**
11	Grouping unlabeled items using k-means clustering
12	Hierarchical Clustering, Probabilistic clustering
13	Association rule mining, Apriori Algorithm
14	F-P growth algorithm
15	Gaussian mixture model
Unit-4	**Introduction to Statistical Learning Theory**
16	Feature Extraction
17	Principal component analysis
18	Singular value decomposition
19	Feature selection: Feature ranking and Subset selection
20	Filter, Wrapper, Embedded methods
21	Evaluating Machine Learning algorithms, Model Selection
Unit-5	**Semi supervised learning& Reinforcement learning**
22	Introduction to Semi supervised learning & Reinforcement learning
23	Markov decision process (MDP), Bellman equations
24	Policy evaluation using Monte Carlo
25	Policy iteration and Value iteration
26	Q-Learning
27	State-Action-Reward-State-Action (SARSA)
28	Model-based Reinforcement Learning
Unit-6	**Recommended system**
29	Collaborative filtering: Memory Based
30	Model Based, Hybrid
31	Content-based filtering: Term Frequency (TF), Inverse Document Frequency(IDF)
32	Artificial neural network, Perceptron, Multilayer network,
33	Back propagation, A multilayer feed forward Neural Network
34	Introduction to Deep learning.

Swami Keshvanand Institute of Technology, Management &Gramothan, Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 3500300 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Program Outcomes/Program Specific Outcomes – Indicators - Competencies

PO 1: Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialisation for the solution of complex engineering problems.	
Competency	**Indicators**
1.1 Demonstrate competence in mathematical modelling.	1.1.1 Apply the knowledge of discrete structures, linear algebra, statistics and numerical techniques to solve problems.
	1.1.2 Apply the concepts of probability, statistics and queuing theory in modelling of computer-based system, data and network protocols.
1.2 Demonstrate competence in basic sciences.	1.2.1 Apply laws of natural science to an engineering problem.
1.3 Demonstrate competence in engineering fundamentals.	1.3.1 Apply engineering fundamentals.
1.4 Demonstrate competence in specialized engineering knowledge to the program.	1.4.1 Apply theory and principles of computer science and engineering to solve an engineering problem.
PO 2: Problem analysis: Identify, formulate, research literature, and analyse complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.	
Competency	**Indicators**
2.1 Demonstrate an ability to identify and formulate complex engineering problem	2.1.1 Evaluate problem statements and identifies objectives.
	2.1.2 Identify processes/modules/algorithms of a computer-based system and parameters to solve a problem.
	2.1.3 Identify mathematical algorithmic knowledge that applies to a given problem.
2.2 Demonstrate an ability to formulate a solution plan and methodology for an engineering problem	2.2.1 Reframe the computer-based system into interconnected subsystems.
	2.2.2 Identify functionalities and computing resources.
	2.2.3 Identify existing solution/methods to solve the problem, including forming justified approximations and assumptions.
	2.2.4 Compare and contrast alternative solution/ methods to select the best methods.
	2.2.5 Compare and contrast alternative solution processes to select the best process.
2.3 Demonstrate an ability to formulate and interpret a model	2.3.1 Able to apply computer engineering principles to formulate modules of a system with required applicability and performance.
	2.3.2 Identify design constraints for required performance criteria.
2.4 Demonstrate an ability to execute a	2.4.1 Apply engineering mathematics to implement the

solution process and analyze results	solution.
	2.4.2 Analyze and interpret results using contemporary tools.
	2.4.3 Identify the limitations of the solution and sources/ causes.
	2.4.4 Arrive at conclusions with respect to the objectives.
PO 3: Design/Development of Solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for public health and safety, and cultural, societal, and environmental considerations.	
Competency	**Indicators**
3.1 Demonstrate an ability to define a complex/ open-ended problem in engineering terms	3.1.1 Able to define a precise problem statement with objectives and scope.
	3.1.2 Able to identify and document system requirements from stake- holders.
	3.1.3 Able to review state-of-the-art literature to synthesize system requirements.
	3.1.4 Able to choose appropriate quality attributes as defined by ISO/IEC/IEEE standard.
	3.1.5 Explore and synthesize system requirements from larger social and professional concerns.
	3.1.6 Able to develop software requirement specifications (SRS).
3.2 Demonstrate an ability to generate a diverse set of alternative design solutions	3.2.1 Able to explore design alternatives.
	3.2.2 Able to produce a variety of potential design solutions suited to meet functional requirements.
	3.2.3 Identify suitable non-functional requirements for evaluation of alternate design solutions.
3.3 Demonstrate an ability to select an optimal design scheme for further development	3.3.1 Able to perform systematic evaluation of the degree to which several design concepts meet the criteria.
	3.3.2 Consult with domain experts and stakeholders to select candidate engineering design solution for further development.
3.4 Demonstrate an ability to advance an engineering design to defined end state	3.4.1 Able to refine architecture design into a detailed design within the existing constraints.
	3.4.2 Able to implement and integrate the modules.
	3.4.3 Able to verify the functionalities and validate the design.
PO 4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.	
Competency	**Indicators**
4.1 Demonstrate an ability to conduct investigations of technical issues	4.1.1 Define a problem for purposes of investigation, its scope and importance.

consistent with their level of knowledge and understanding.	4.1.2 Able to choose appropriate procedure/algorithm, dataset and test cases.
	4.1.3 Able to choose appropriate hardware/software tools to conduct the experiment.
4.2 Demonstrate an ability to design experiments to solve open-ended problems.	4.2.1 Design and develop appropriate procedures/methodologies based on the study objectives.
4.3 Demonstrate an ability to analyze data and reach a valid conclusion.	4.3.1 Use appropriate procedures, tools and techniques to collect and analyze data.
	4.3.2 Critically analyze data for trends and correlations, stating possible errors and limitations.
	4.3.3 Represent data (in tabular and/or graphical forms) so as to facilitate analysis and explanation of the data, and drawing of conclusions.
	4.3.4 Synthesize information and knowledge about the problem from the raw data to reach appropriate conclusions.
PO 5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.	
Competency	Indicators
5.1 Demonstrate an ability to identify/ create modern engineering tools, techniques and resources	5.1.1 Identify modern engineering, techniques and resources for engineering activities.
	5.1.2 Create/adapt/modify/extend tools and techniques to solve engineering problems.
5.2 Demonstrate an ability to select and apply discipline- specific tools, techniques and resources	5.2.1 Identify the strengths and limitations of tools for (i) acquiring information, (ii) modelling and simulating, (iii) monitoring system performance, and (iv) creating engineering designs.
	5.2.2 Demonstrate proficiency in using discipline-specific tools.
5.3 Demonstrate an ability to evaluate the suitability and limitations of tools used to solve an engineering problem	5.3.1 Discuss limitations and validate tools, techniques and resources.
	5.3.2 Verify the credibility of results from tool use with reference to the accuracy and limitations, and the assumptions inherent in their use.
PO 6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.	
Competency	**Indicators**
6.1 Demonstrate an ability to describe engineering roles in a broader context, e.g. pertaining to the environment, health, safety, legal and public welfare	6.1.1 Identify and describe various engineering roles; particularly as pertains to protection of the public and public interest at the global, regional and local level.
6.2 Demonstrate an understanding of professional engineering regulations, legislation and standards	6.2.1 Interpret legislation, regulations, codes, and standards relevant to your discipline and explain its contribution to the protection of the public.

PO 7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and the need for sustainable development.	
Competency	**Indicators**
7.1 Demonstrate an understanding of the impact of engineering and industrial practices on social, environmental and in economic contexts	7.1.1 Identify risks/impacts in the life-cycle of an engineering product or activity.
	7.1.2 Understand the relationship between the technical, socio-economic and environmental dimensions of sustainability.
7.2 Demonstrate an ability to apply principles of sustainable design and development	7.2.1 Describe management techniques for sustainable development.
	7.2.2 Apply principles of preventive engineering and sustainable development to an engineering activity or product relevant to the discipline.
PO 8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.	
Competency	**Indicators**
8.1 Demonstrate an ability to recognize ethical dilemmas	8.1.1 Identify situations of unethical professional conduct and propose ethical alternatives.
8.2 Demonstrate an ability to apply the Code of Ethics	8.2.1 Identify tenets of the ASME/IEEE/CSI/ACM professional code of ethics.
	8.2.2 Examine and apply moral & ethical principles to known case studies.
PO 9: Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.	
Competency	**Indicators**
9.1 Demonstrate an ability to form a team and define a role for each member	9.1.1 Recognize a variety of working and learning preferences; appreciate the value of diversity on a team.
	9.1.2 Implement the norms of practice (e.g. rules, roles, charters, agendas, etc.) of effective team work, to accomplish a goal.
9.2 Demonstrate effective individual and team operations- communication, problem- solving, conflict resolution and leadership skills.	9.2.1Demonstrate effective communication, problem-solving, conflict resolution and leadership skills.
	9.2.2 Treat other team members respectfully.
	9.2.3 Listen to other members.
	9.2.4Maintain composure in difficult situations.
9.3 Demonstrate success in a team-based project	9.3.1 Present results as a team, with smooth integration of contributions from all individual efforts.
PO 10: Communication: Communicate effectively on complex engineering activities with the engineering community and with the society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.	
Competency	**Indicators**
10.1 Demonstrate an ability to comprehend technical literature and document project work	10.1.1 Read, understand and interpret technical and non-technical information.
	10.1.2 Produce clear, well-constructed, and well-supported written engineering documents.
	10.1.3 Create flow in a document or presentation - a

	logical progression of ideas so that the main point is clear.
10.2 Demonstrate competence in listening, speaking, and presentation	10.2.1 Listen to and comprehend information, instructions, and viewpoints of others.
	10.2.2 Deliver effective oral presentations to technical and non-technical audiences.
10.3 Demonstrate the ability to integrate different modes of communication	10.3.1 Create engineering-standard figures, reports and drawings to complement writing and presentations.
	10.3.2 Use a variety of media effectively to convey a message in a document or a presentation.
PO 11: Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's work, as a member and leader in a team, to manage projects and in multidisciplinary environments.	
Competency	**Indicators**
11.1 Demonstrate an ability to evaluate the economic and financial performance of an engineering activity	11.1.1 Describe various economic and financial costs/benefits of an engineering activity.
	11.1.2 Analyze different forms of financial statements to evaluate the financial status of an engineering project.
11.2 Demonstrate an ability to compare and contrast the costs/benefits of alternate proposals for an engineering activity	11.2.1 Analyze and select the most appropriate proposal based on economic and financial considerations.
11.3 Demonstrate an ability to plan/manage an engineering activity within time and budget constraints	11.3.1 Identify the tasks required to complete an engineering activity, and the resources required to complete the tasks.
	11.3.2 Use project management tools to schedule an engineering project, so it is completed on time and on budget.
PO 12: Life-long learning: Recognise the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.	
Competency	**Indicators**
12.1 Demonstrate an ability to identify gaps in knowledge and a strategy to close these gaps	12.1.1 Describe the rationale for the requirement for continuing professional development.
	12.1.2 Identify deficiencies or gaps in knowledge and demonstrate an ability to source information to close this gap.
12.2 Demonstrate an ability to identify changing trends in engineering knowledge and practice	12.2.1 Identify historic points of technological advance in engineering that required practitioners to seek education in order to stay current.
	12.2.2 Recognize the need and be able to clearly explain why it is vitally important to keep current regarding new developments in your field.
12.3 Demonstrate an ability to identify and access sources for new information	12.3.1 Source and comprehend technical literature and other credible sources of information.
	12.3.2 Analyze sourced technical and popular information for feasibility, viability, sustainability, etc.
PSO 1: Core Engineering Skills: Exhibit fundamental concepts of Data Structures, Databases, Operating Systems, Computer Network, Theory of Computation, Advanced Programming and Software Engineering.	
PSO1.1.1	Possess the concepts of Data Structure and Database Management System.

PSO1.1.2	Possess the concepts of core engineering subjects including Operating System, Computer Networks and Software Engineering.
PSO1.1.3	Apply basic programming skills to solve real world problems.
PSO 2: Standard Software Engineering practices: Demonstrate an ability to design, develop, test, debug, deploy, analyze, troubleshoot, maintain, manage and secure a software.	
PSO2.1.1	Apply fundamental software engineering concepts to solve real world problem.
PSO2.1.2	Possess conceptual knowledge for designing, analysing and testing a software.
PSO2.1.3	Estimate and evaluate the cost related to a Software.
PSO 3: Future Endeavours: Recognize the need to have knowledge of higher education institutions/ organizations/ companies related to computer science & engineering.	
PSO3.1.1	Explore the need of current technology being practised by computer science industry/ institutions.
PSO3.1.2	Identify the requirement of continuing education through post graduation like M.Tech., MS, MBA etc.
PSO3.1.3	List various higher education institutes and organizations related to computer science & engineering.

Bloom's Taxonomy (Revised)

Level	Descriptor	Level of Attainment	Keywords
1	Remembering	Recalling from memory	List, define, tell, describe, recite, recall, identify, show, label, tabulate, quote, name, who, when, where, etc.
2	Understanding	Explaining ideas or concepts	Describe, explain, paraphrase, restate, associate, contrast, summarize, differentiate interpret, discuss.
3	Applying	Using information in another familiar situation	Calculate, predict, apply, solve, illustrate, use, demonstrate, determine, model, experiment, show, examine, modify
4	Analysing	Breaking information into part to explore understandings and relationships	Classify, outline, break down, categorize, analyze, diagram, illustrate, infer, select.
5	Evaluating	Justifying a decision or course of action	Assess, decide, choose, rank, grade, test, measure, defend, recommend, convince, select, judge, support, conclude, argue, justify, compare, summarize, evaluate.
6	Creating	Generating new ideas, products or views to do things	Design, formulate, build, invent, create, compose, generate, derive, modify, develop, integrate.

** It may be noted that some of the verbs in the above table are associated with multiple Bloom's Taxonomy level. These verbs are actions that could apply to different activities. We need to keep in mind that it's the skill, action or activity we need out students to demonstrate that will determine the contextual meaning of the verb used in the assessment question.

Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur

Programme: B.Tech. (Computer Science & Engineering(AI))

Semester: VI

Course Name (Course Code): Machine Learning(6CAI4-02)

Course Outcomes

After completion of this course, students will be able to –

6CAI4-06.1	***Demonstrate*** a variety of supervised machine-learning models and showcase their applications.
6CAI4-06.2	***Illustrate*** various applications of unsupervised machine learning models.
6CAI4-06.3	***Recognize*** diverse machine-learning algorithms through statistical learning techniques.
6CAI4-06.4	***Discuss*** the methodologies of semi-supervised learning and reinforcement learning.
6CAI4-06.5	***Employ*** the concepts of Neural Networks within the realm of Machine Learning.

Name of Faculty: Abha Jain
(Signature)

Verified by Course Coordinator

Signature
(Name: .....)

Verified by Verification and Validation Committee, DPAQIC

Signature
(Name: .....)
Mamta Saloyal

Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur

COURSE:Machine Learning(6CAI4-02)

CO	Outcomes	Bloom'sLevel	POIndicators	PSOIndicators
Upon successful completion of this course,students should be able to:				
6CAI4-06.1	***Demonstrate*** a variety of supervised machine-learning models and showcase their applications.	3	1.1.1, 1.1.2, 1.2.1, 1.3.1, 1.4.1, 2.1.1, 2.1.2, 2.1.3, 2.2.3, 2.2.4, 2.2.5, 2.3.1, 2.4.1, 2.4.3, 2.4.4, 3.1.1, 3.1.4, 3.2.3, 4.1.2, 4.3.1, 5.1.1, 5.1.2, 8.1.1, 8.2.2, 12.1.1, 12.2.1, 12.2.2, 12.3.1, 12.3.2	1.1.1, 1.1.2, 1.1.3, 2.1.3, 3.1.1, 3.1.2, 3.1.3
6CAI4-06.2	***Illustrate*** various applications of unsupervised machine learning models.	3	1.1.1, 1.1.2, 1.2.1, 1.3.1, 1.4.1, 2.1.1, 2.1.2, 2.1.3, 2.2.3, 2.2.4, 2.2.5, 2.3.1, 2.4.1, 2.4.3, 2.4.4, 3.1.1, 3.1.4, 3.2.3, 4.1.2, 4.3.1, 5.1.1, 5.1.2, 8.1.1, 8.2.2, 12.1.1, 12.2.1, 12.2.2, 12.3.1, 12.3.2	1.1.1, 1.1.3, 2.1.3, 3.1.1, 3.1.2, 3.1.3
6CAI4-06.3	***Recognize*** diverse machine-learning algorithms through statistical learning techniques.	2	1.1.1, 1.1.2, 1.2.1, 1.3.1, 1.4.1, 2.1.1, 2.1.2, 2.1.3, 2.2.3, 2.2.4, 2.2.5, 2.3.1, 2.4.1, 2.4.2, 2.4.3, 2.4.4, 4.1.2, 4.1.3, 4.2.1, 4.3.1, 4.3.2, 4.3.3, 4.3.4, 12.1.1, 12.2.1, 12.2.2, 12.3.1, 12.3.2	1.1.1, 1.1.2, 1.1.3, 2.1.3, 3.1.1, 3.1.2, 3.1.3
6CAI4-06.4	***Discuss*** the methodologies of semi-supervised learning and reinforcement learning.	2	1.1.1, 1.1.2, 1.2.1, 1.3.1, 1.4.1, 2.1.1, 2.1.2, 2.1.3, 2.2.3, 2.2.4, 2.2.5, 2.3.1, 2.4.1, 2.4.3, 2.4.4, 5.1.1, 5.1.2, 12.1.1, 12.1.2, 12.2.1, 12.2.2, 12.3.1, 12.3.2	1.1.1, 1.1.3, 2.1.3, 3.1.1, 3.1.2, 3.1.3
6CAI4-06.5	***Employ*** the concepts of Neural Networks within the realm of Machine Learning.	3	1.1.1, 1.1.2, 1.2.1, 1.3.1, 1.4.1, 2.1.1, 2.1.2, 2.1.3, 2.2.3, 2.2.4, 2.2.5, 2.3.1, 2.4.1, 2.4.3, 2.4.4, 3.1.1, 3.1.4, 3.2.3, 4.1.2, 4.2.1, 4.2.3, 4.3.1, 4.3.2, 4.3.3, 4.3.4, 5.1.1, 5.3.1, 5.3.2,8.1.1, 8.2.2, 12.1.1, 12.1.2, 12.2.1, 12.2.2, 12.3.1, 12.3.2	1.1.1, 1.1.2, 1.1.3, 2.1.3, 3.1.1, 3.1.2, 3.1.3

Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur

CO-PO/PSO Mapping: Formulation and Justification

The CO-PO/PSO mapping is based on the correlation of course outcome (CO) with Program Outcome Indicators. These indicators are the breakup statements of broad Program Outcome statement.

The correlation is calculated as number of correlated indicators of a PO/PSO mapped with CO divided by total indicators of a PO/PSO. The calculated value represents the correlation level between a CO & PO/PSO. Detailed formulation and mathematical representation can be seen below in equation 1:

Input: CO_i: *The* i^{th} *course outcome of the course*

PO_j: *The* j^{th} *Program Outcome*

Ij_k: *The* k^{th} *indicator of the* j^{th} *Program Outcome*

$\alpha(I_{jk}, CO_i)$: *level of CO-PO mapping*

$= 1$, *if,* $0 < \alpha \leq 0.33$

$= 2$, *if,* $0.33 < \alpha \leq 0.66$

$= 3$, *if,* $0.66 < \alpha \leq 1$

$$\alpha\,(I_{jk},CO_i) = (count(\lambda(I_{jk},CO_i))/\ count\,(I_k,PO_j)$$

λ = *Degree of correlation*

Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur

CO-PO/PSO Mapping
Programme: B.Tech. (Computer Science & Engineering(AI))
Semester: VI
Course Name (Course Code): Machine Learning(6CAI4-02)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO 1	PSO 2	PSO 3
CO1	3	3	2	1	2			2				3	3	1	3
CO2	3	3	2	1	2			2				3	2	1	3
CO3	3	3		3								3	3	1	3
CO4	3	3			1							3	2	1	3
CO5	3	3	2	3	2			2				3	3	1	3
Weighted Average	**3**	**3**	**2**	**3**	**2**			**2**				**3**	3	1	3

Name of Faculty (Signature) Abha Jain

Name of Faculty (Signature)

Verified by Course Coordinator

Signature (Name: .....)

Verified by Verification and Validation Committee, DPAQIC

Signature (Name: Mamta, Jaipur.)

Teaching-Learning Methodology

Teaching-Learning Methodology for Machine Learning Course:

1. **Interactive Lectures**: Engaging lectures covering theoretical foundations, algorithms, and practical applications of machine learning concepts. Encourages active participation through discussions, questions, and real-world examples.
2. **Hands-on Practice**: Extensive hands-on sessions with programming exercises using popular machine learning libraries (e.g., scikit-learn, TensorFlow). Students apply theoretical knowledge to real datasets, fostering practical skills in data preprocessing, model implementation, and evaluation.
3. Project-Based Learning**: Structured projects requiring students to solve real-world problems using machine learning techniques. Projects enable students to work independently or in groups, enhancing problem-solving abilities and promoting creativity and innovation.**
4. Case Studies and Use Cases**: Analysis of case studies and industry use cases demonstrates the application of machine learning in various domains such as healthcare, finance, and marketing. Helps students understand real-world challenges, solutions, and best practices.**
5. Guest Lectures and Workshops**: Inviting experts from academia and industry to deliver specialized lectures or conduct workshops on advanced topics, emerging trends, and practical insights. Provides diverse perspectives and facilitates networking opportunities for students.**
6. Peer Learning and Collaboration**: Encourages peer learning through group discussions, peer code reviews, and collaborative projects. Fosters teamwork, communication skills, and mutual support among students, enriching the learning experience.**
7. Assessment and Feedback**: Regular assessments, quizzes, and assignments to evaluate understanding and progress. Constructive feedback from instructors and peers helps students identify strengths, weaknesses, and areas for improvement, promoting continuous learning and growth.**
8. Online Resources and Support**: Access to supplementary materials, tutorials, and online forums for self-paced learning and additional support. Facilitates independent exploration and reinforces key concepts covered in class.**

The teaching-learning methodology for the machine learning course combines theoretical foundations, practical applications, hands-on practice, collaborative projects, and continuous assessment to provide a comprehensive and engaging learning experience for students, preparing them for success in the dynamic field of machine learning.

Roll No. ..... Total Page No. : 5

610402/610902/611802

610402/610902/611802

B.TECH. VI SEM MAIN/BACK EXAM AUGUST 2023

COMPUTER SCIENCE AND ENGINEERING

(6CS4-02) - MACHINE LEARNING

COMMON WITH CSE & IT, CSE(DS)

Time : 3 Hours] **[Max. Marks :** 120

[Min. Passing Marks :

Instructions to Candidates : Part – A : Short answer type questions (up to 25 words) 10 × 2 marks = 20 marks. All ten questions are compulsory.

Part – B: Analytical/Problem Solving questions 5 × 8 marks = 40 marks. Candidates have to answer 5 questions out of 7.

Part – C: Descriptive/Analytical/Problem Solving questions 4 × 15 marks = 60 marks. Candidates have to answer 4 questions out of 5.

Schematic diagrams must be shown wherever necessary. Any data you feel missing may suitably be assumed and stated clearly. Units of quantities used/calculated must be stated clearly.

Use of following supporting materials is permitted during examination. (Mentioned in form No. 205)

1 : NIL **2 : NIL**

PART A (2 marks each)

1. What is the difference between classification and regression ?
2. What is 'Naive' in a Naive Bayes ?

3. Explain overfitting and underfitting ?

4. What do you mean by Associative Rule Mining ?

5. What is PCA ? When do you use it ?

6. Define SVD. Also write their applications.

7. Compare Reinforced Learning and Supervised Learning.

8. What is Markov Decision Process ?

9. Describe the structure of Artificial Neural Networks ?

10. Name and define techniques used to find similarities in the recommendation system.

PART B

1. Explain Decision Tree algorithm in detail. What are Entropy and Information gain in Decision tree algorithm ? [4+4]

2. What are Recommender Systems ? What is Collaborative filtering and Content Based filtering ? [4+4]

3. What is the motivation behind the Gaussian Mixture Model ? What is the relationship between k-means and PCA ? [4+4]

4. What is curse of dimensionality and how can unsupervised learning help with it ? Discuss various dimensionality reduction techniques. [4+4]

5. Explain the purpose of slack variables in SVM problem formulation. When would you use SVM over random forest and vice-versa ? [4+4]

6. Differentiate Q-learning and SARSA and when you would use each one explain with example. [4+4]

7. Transform the following data into decision tree based on Gini impurity criteria. [4+4]

Deadline ?	Is there a party ?	Lazy ?	Activity
Urgent	Yes	Yes	Party
Urgent	No	Yes	Study
Near	Yes	Yes	Party
None	Yes	No	Party
None	No	Yes	Pub
None	Yes	No	Party
Near	No	No	Study
Near	No	Yes	TV
Near	Yes	Yes	Party
Urgent	No	No	Study

PART C (15 marks each)

1. The following dataset will be used to learn a decision tree for predicting whether a person is happy (H) or sad (S) based on the color of their shoes, whether they wear a wig and number of masks they have.

Color	Wig	No. of Marks	(Output) Emotion)
G	Y	2	S
G	N	2	S
G	N	2	S
B	N	2	S
B	N	2	H
R	N	2	H
R	N	2	H
R	N	2	H
R	Y	3	H

(i) What is the H (Emotion | Wig=Y) ? [Show your calculations] [2]

(ii) What is the H (Emotion | Masks=3) ? [Show your calculations] [2]

(iii) Which attribute would the decision-tree building algorithm choose to use for the root of the tree (Assume no pruning). [Show your calculations] [2]

(iv) Draw the full decision tree that would be learned for this data (assume no pruning). [2]

(v) What would be the training set error for this dataset ? Express your answer as the percentage of records that would be misclassified. [Justify your answer with proper reasoning] [2]

The following two questions does not use the previous example dataset.

(vi) Assuming that the output attribute can take two values. What is the maximum training set error that any dataset could possibly have ? [2]

(vii) Construct an example dataset that achieves this maximum percentage training set error. (It must have two or fewer inputs and five or fewer records). [3]

2. Explain why precision or recall alone is not a sufficient measure to measure classifier performance in the case of binary classification. Discuss AUC-ROC curve used in classification problems.

3. Illustrate Markov Decision Problem in reinforcement learning with the help of an example.

4. Discuss Association rule mining in detail. A database has nine transactions. Let min-support = 22% and min- confidence = 70%. Find all frequent item sets by using Apriori Algorithm and generate association rule on this. TID is the transaction ID. (7+8)

TID	List of Items
T100	L1, L2, L5
T100	L2, L4
T100	L2, L3
T100	L1, L2, L4
T100	L1, L3
T100	L2, L3
T100	L1, L3
T100	L1, L2, L3, L5
T100	L1, L2, L3

5. Train a perceptron model to learn the following truth table :

X	Y	Output (Target)
1	1	0
1	0	0
0	1	1
0	0	1

Take the initial parameters as $w_0 = 0.03$, $w_1 = -0.04$, $w_2 = 0.04$, $n = 0.4$ and bias as -1. Use threshold activation function [$g(h) = 1$, if $h > 0$; $g(h) = 0$, if h < =0] and sequential update strategy. Consider input samples in the same order as given in the above table. Show iteration steps and show in graph how the decision boundary changes.

6E7102

Roll No. _____ [Total No. of Pages : 3]

6E7102

B.Tech. VI sem. (Main) Examination, July - 2023
Computer Science and Engineering (Artificial Intelligence)
6CAI4-02 Machine Learning
CS,IT,AID, CAI

Time : 3 Hours **Maximum Marks : 70**

Instructions to Candidates:

Attempt all ten questions From Part A, five questions out of seven questions from Part B and three questions out of five questions from Part C .

Schematic diagrams must be shown wherever necessary. Any data you feel missing suitably be assumed and states clearly. Units of quantities used/calculated must be stated clearly.

Use of following supporting material is permitted during examination (Mentioned in form No. 205)

PART - A

(Answer should be given up to 25 words only)

All questions are compulsory. **(10×2=20)**

1. Express the Markov property mathematically.
2. Give clear difference between episodic and *continuous* tasks of Markov process.
3. Why dimensionality reduction is required for a dataset?
4. Which cost function is used in logistic regression and why?
5. Write names of different types of clustering methods.
6. What is the use of attribute selection measure in decision tree classifier.
7. Define singular value decompositions.
8. What is Deep learning?
9. What is support vector in SVM?
10. Give name of u-filter feature selection methods.

PART - B

(Analytical/Problem solving questions)

Attempt any five questions. **(5×4=20)**

1. What do you understand about " Bellman equation for value function". Give example.
2. Give merits and demerits of filter and wrapper feature selection methods.
3. Discuss about frequent pattern, support and confidence of a association rule with example.
4. Explain following with respect to multilayer network
   a) Weights and Biases
   b) Use of Activation functions.
5. What is the use of confusion matrix. Define all the related terms of a confusion matrix.
6. Discuss various types of splits of a attribute in a decision tree classification algorithm.
7. What is overfitting problem in Machine learning algorithm. Give solutions for it.

PART - C

(Descriptive/Analytical/Problem solving/Design questions))

Attempt any three questions. **(3×10=30)**

1. Explain K-nearest neighbor method. Consider a binary classification problem with two classes C1 and C2. Class lables of ten other training set instances sorted in increasing order of their distance to an instance.

   x is as follows: $\{C1, C2, C1, C2, C2, C2, C1, C2, C1, C2\}$.

   How will a $K=3$ nearest neighbour classifier classify the instance x

2. Suppose you are given following set of training examples. Each attribute can take on one of three nominal values: $a b, or\ c$.

A1	*A2*	*A3*	*Class*
a	*c*	*a*	*C1*
c	*a*	*c*	*C1*
a	*a*	*c*	*C2*
b	*c*	*a*	*C2*
c	*c*	*b*	*C2*

   a) How would a Naive Bayes classifier classify the example $A1 = a, A2 = c, A3 = b$? Show all steps.

   b) How would a Name Bayes classifier classify the example $A1 = c, A2 = c, A3 = a$? Show all steps.

3. Explain f-p Growth algorithm for frequent pattern generation. Give suitable example and all computational steps with diagrams.

4. A neural network takes two binary values inputs, $x1, x2 \in \{0,1\}$ and activation function is the binary threshold functions $\begin{pmatrix} h(z)=1 & if & z>0 \\ 0 & & otherwise \end{pmatrix}$

Design a neural network to compute the AND Boolean function. Consider the truth table for of AND Boolean functions. weights are {2,2) and Biase is -3.

5. Write short notes on following

a) Model based reinforcement learning.

b) K - means clustering algorithm

c) Single linkage and complete linkage clustering algorithm with example.

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

I -Mid Term Examination, March-2024

Prog./Semester:	**B. Tech.-VI**	**Branch: CS/AI/DS**	**CSE**
Course Title:	**Machine Learning**	**Course Code:**	**6CS4-02/ 6CAI4-02/ 6CSD4-02**
Duration:	**1.5 Hours**	**Maximum Marks:**	**20**
Session: First/Second/Third		**Roll No.:**	
Instructions if any:		***Neat diagrams must be drawn wherever necessary***	

PART -A (short answer type questions)

(All questions are compulsory)

Q. No.	Question	Max. Marks	CO(s)	Bloom's Level
1	Differentiate Supervised and Unsupervised Machine Learning technique (3 differences).	2	CO1	L1
2	Explain hyperplane, support vector and margin in SVMs algorithm with suitable diagram.	2	CO1	L1
3	Explain Hierarchical Clustering and its type with the help of any example.	2	CO2	L2

PART -B (Analytical/Problem solving questions)

(All questions are compulsory)

Q. No.	Question	Max. Marks	CO(s)	Bloom's Level
4	Utilize the dataset provided below to predict whether a new individual, identified by a high income and male gender, will make a purchase using the Naïve Bayes algorithm.	4	CO1	L3

Income	Gender	Buy Product
HIGH	MALE	YES
HIGH	FEMALE	YES
MEDIUM	FEMALE	NO
LOW	MALE	YES
LOW	MALE	YES
MEDIUM	FEMALE	NO

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

I -Mid Term Examination, March-2024

Prog./Semester:	**B. Tech.-VI**	**Branch: CS/AI/DS**	**CSE**
Course Title:	**Machine Learning**	**Course Code:**	**6CS4-02/ 6CAI4-02/ 6CSD4-02**
Duration:	**1.5 Hours**	**Maximum Marks:**	**20**
Session: First/Second/Third		**Roll No.:**	
Instructions if any:		***Neat diagrams must be drawn wherever necessary***	

PART -A (short answer type questions)

(All questions are compulsory)

Q. No.	Question	Max. Marks	CO(s)	Bloom's Level
1	Differentiate Supervised and Unsupervised Machine Learning technique (3 differences).	2	CO1	L1
2	Explain hyperplane, support vector and margin in SVMs algorithm with suitable diagram.	2	CO1	L1
3	Explain Hierarchical Clustering and its type with the help of any example.	2	CO2	L2

PART -B (Analytical/Problem solving questions)

(All questions are compulsory)

Q. No.	Question	Max. Marks	CO(s)	Bloom's Level
4	Utilize the dataset provided below to predict whether a new individual, identified by a high income and male gender, will make a purchase using the Naïve Bayes algorithm.	4	CO1	L3

Income	Gender	Buy Product
HIGH	MALE	YES
HIGH	FEMALE	YES
MEDIUM	FEMALE	NO
LOW	MALE	YES
LOW	MALE	YES
MEDIUM	FEMALE	NO

5	Illustrate why the decision tree algorithm is the most widely adopted method for classification and explain the concepts of entropy and information gain within this algorithm.	**4**	**CO1**	**L4**

PART- C (Descriptive/Analytical/Problem solving/Design questions)
(Attempt any one Question)

Q. No.	Question	Max. Marks	CO(s)	Bloom's Level
6	A database has 6 transactions. Let min support =50% and min Confidence =60%.Find all the frequent item sets using Apriori algorithm and generate association rule on this. **Transaction Id** / **Items Bought** T1 — I1, I2, I3 T2 — I2, I3, I4 T3 — I4, I5 T4 — I1, I2, I4 T5 — I1, I2, I3, I5 T6 — I1,I2,I3,I4	**6**	**CO2**	**L3**
7	Apply K- Means Clustering for the following data sets for two clusters. Use any two random points as initial centroids. **S. No** / **Height** / **Weight** 1 / 185 / 72 2 / 170 / 56 3 / 168 / 60 4 / 179 / 68 5 / 182 / 72 6 / 188 / 77	**6**	**CO2**	**L3**

5	Illustrate why the decision tree algorithm is the most widely adopted method for classification and explain the concepts of entropy and information gain within this algorithm.	**4**	**CO1**	**L4**

PART- C (Descriptive/Analytical/Problem solving/Design questions)
(Attempt any one Question)

Q. No.	Question	Max. Marks	CO(s)	Bloom's Level
6	A database has 6 transactions. Let min support =50% and min Confidence =60%. Find all the frequent item sets using Apriori algorithm and generate association rule on this. **Transaction Id** / **Items Bought** T1 — I1, I2, I3 T2 — I2, I3, I4 T3 — I4, I5 T4 — I1, I2, I4 T5 — I1, I2, I3, I5 T6 — I1,I2,I3,I4	**6**	**CO2**	**L3**
7	Apply K- Means Clustering for the following data sets for two clusters. Use any two random points as initial centroids. **S. No** / **Height** / **Weight** 1 / 185 / 72 2 / 170 / 56 3 / 168 / 60 4 / 179 / 68 5 / 182 / 72 6 / 188 / 77	**6**	**CO2**	**L3**

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

II -Mid Term Examination, May-2024

Prog./Semester:	**B. Tech.-VI**	**Branch: CS/AI/DS**	**CSE**
Course Title:	**Machine Learning**	**Course Code:**	**6CS4-02/ 6CAI4-02/ 6CSD4-02**
Duration:	**1.5 Hours**	**Maximum Marks:**	**20**
Session: First/Second/Third		**Roll No.:**	
Instructions if any:		***Neat diagrams must be drawn wherever necessary***	

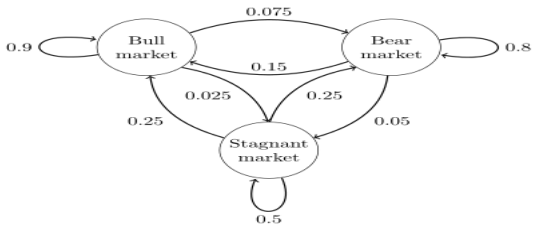
PART -A (short answer type questions)

(All questions are compulsory)

Q. No.	Question	Max. Marks	CO(s)	Bloom's Level
1	What is the difference between Dynamic programming, Monte Carlo and Temporal methods of Reinforcement learning?	2	CO4	L3
2	Differentiate Feature Selection and Feature Extraction methods.	2	CO3	L3
3	Derive gradient descent algorithm and give reasons for using it in Neural Networks.	2	CO5	L4

PART -B (Analytical/Problem solving questions)

(All questions are compulsory)

Q. No.	Question	Max. Marks	CO(s)	Bloom's Level
4	Differentiate Policy iteration and value iteration. Draw transition matrix for the problem described below to be used in morkov model. 	4	CO4	L4

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

II -Mid Term Examination, May-2024

Prog./Semester:	**B. Tech.-VI**	**Branch: CS/AI/DS**	**CSE**
Course Title:	**Machine Learning**	**Course Code:**	**6CS4-02/ 6CAI4-02/ 6CSD4-02**
Duration:	**1.5 Hours**	**Maximum Marks:**	**20**
Session: First/Second/Third		**Roll No.:**	
Instructions if any:		***Neat diagrams must be drawn wherever necessary***	

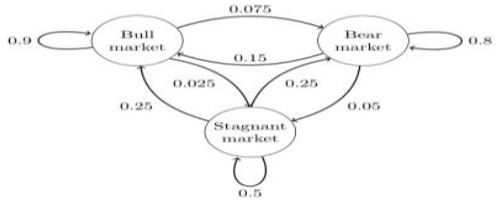
PART -A (short answer type questions)

(All questions are compulsory)

Q. No.	Question	Max. Marks	CO(s)	Bloom's Level
1	What is the difference between Dynamic programming, Monte Carlo and Temporal methods of Reinforcement learning?	2	CO4	L3
2	Differentiate Feature Selection and Feature Extraction methods.	2	CO3	L3
3	Derive gradient descent algorithm and give reasons for using it in Neural Networks.	2	CO5	L4

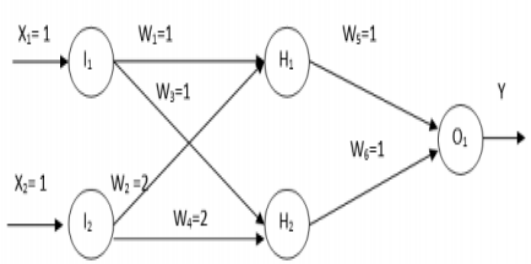
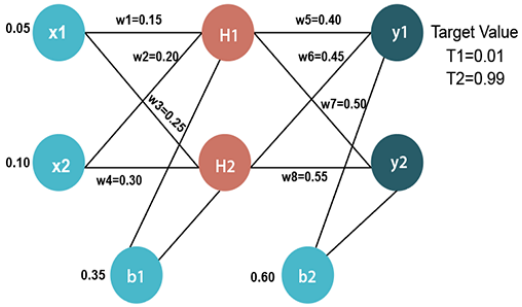
PART -B (Analytical/Problem solving questions)

(All questions are compulsory)

Q. No.	Question	Max. Marks	CO(s)	Bloom's Level
4	Differentiate Policy iteration and value iteration. Draw transition matrix for the problem described below to be used in morkov model. 	4	CO4	L4

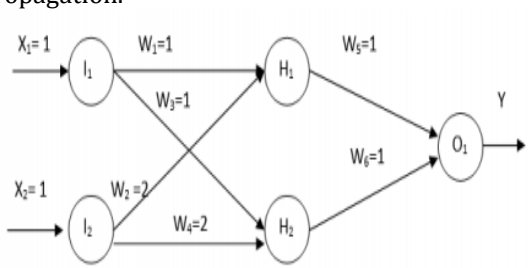
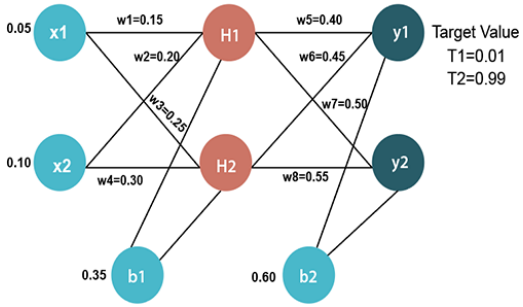
5	A Confusion Matrix of a model applied to a set of 200 observations is given in table below. Determine Accuracy, Recall, Precision, Specificity and F1 score of model. Predicted C1 Predicted C2 Actual C1 55 25 Actual C2 35 85	4	CO3	L3

PART- C (Descriptive/Analytical/Problem solving/Design questions)
(Attempt any one Question)

Q. No.	Question	Max. Marks	CO(s)	Bloom's Level
6	Consider the following Neural Network with alpha = 0.5, desired o/p = 1 and sigmoid activation function. 1. Perform one forward pass and calculate the error. 2. Calculate the updated weights for w5 using back propagation. 	6	CO5	L4
7	Calculate Mean Square error for given neural network using feed forward neural network. 	6	CO5	L4

5	A Confusion Matrix of a model applied to a set of 200 observations is given in table below. Determine Accuracy, Recall, Precision, Specificity and F1 score of model. Predicted C1 Predicted C2 Actual C1 55 25 Actual C2 35 85	4	CO3	L3

PART- C (Descriptive/Analytical/Problem solving/Design questions)
(Attempt any one Question)

Q. No.	Question	Max. Marks	CO(s)	Bloom's Level
6	Consider the following Neural Network with alpha = 0.5, desired o/p = 1 and sigmoid activation function. 1. Perform one forward pass and calculate the error. 2. Calculate the updated weights for w5 using back propagation. 	6	CO5	L4
7	Calculate Mean Square error for given neural network using feed forward neural network. 	6	CO5	L4

Machine Learning Quiz 2

* Indicates required question

1. Email *

2. Name *

3. Roll No *

4. * 1 point

1. Type of matrix decomposition model is

Mark only one oval.

- "1. predictive model
- 2. descriptive model
- 3. logical model
- 4. None"

5. * 1 point

2. PCA is

Mark only one oval.

- "1. backward feature selection
- 2. forward feature selection
- 3. feature extraction
- 4. None of these"

6. * 1 point

3. Supervised learning and unsupervised clustering both require which is correct according to the statement.

Mark only one oval.

- "1. input attribute
- 2. hidden attribute
- 3. output attribute
- 4. categorical attribute"

7. 4. Following are the types of supervised learning * 1 point

Mark only one oval.

- 1. regression
- 2. classification
- 3. subgroup discovery
- 4. All of above

8. 5. A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college. Here feature type is * 1 point

Mark only one oval.

- 1. ordinal
- 2. nominal
- 3. categorical
- 4. Boolean

9. 6.Following is powerful distance metrics used by Geometric model * 1 point

Mark only one oval.

- 1. Manhattan distance
- 2. Euclidean distance
- 3. All of above
- 4. None of above

10. 7. The output of training process in machine learning is * 1 point

Mark only one oval.

- 1. machine learning algorithm
- 2. machine learning model
- 3. null
- 4. accuracy

11. 8. Which of the following is a good test dataset characteristic? * 1 point

Mark only one oval.

- 1. is representative of the dataset as a whole
- 2. large enough to yield meaningful results
- 3. All of above
- 4. None of above

12. 9. Which of the following techniques would perform better for reducing dimensions of a data set? * 1 point

Mark only one oval.

- 1. removing columns which have high variance in data
- 2. removing columns which have too many missing value
- 3. removing columns with dissimilar data trends
- 4. None of the above

13. 10. What characterize is hyper plane in geometrical model of machine learning? * 1 point

Mark only one oval.

- 1. a plane with 1 dimensional fewer than number of input attributes
- 2. a plane with 1 dimensional more than number of input attributes
- 3. a plane with 2 dimensional more than number of input attributes
- 4. a plane with 2 dimensional fewer than number of input attributes

14. 11. You are given reviews of few Netflix series marked as positive, negative and neutral. Classifying reviews of a new Netflix series is an example of * 1 point

Mark only one oval.

- 1. unsupervised learning
- 2. semi supervised learning
- 3. supervised learning
- 4. reinforcement learning

15. 12. Like the probabilistic view, the view allows us to associate a probability of membership with each classification * 1 point

Mark only one oval.

- 1. deductive
- 2. exampler
- 3. classical
- 4. inductive

16. 13. The problem of finding hidden structure in unlabeled data is called * 1 point

Mark only one oval.

- 1. unsupervised learning
- 2. reinforcement learning
- 3. supervised learning
- 4. None

17. 14. If machine learning model output involves target variable then that model is called as * 1 point

Mark only one oval.

- 1. predictive model
- 2. descriptive model
- 3. reinforcement learning
- 4. All of above

18. 15. Database query is used to uncover this type of knowledge. * 1 point

Mark only one oval.

- 1. hidden
- 2. shallow
- 3. deep
- 4. multidimensional

19. 16. Data used to build a data mining model. * 1 point

Mark only one oval.

- 1. training data
- 2. hidden data
- 3. test data
- 4. validation data

20. 17. Application of machine learning methods to large databases is * 1 point

Mark only one oval.

- 1. big data computing
- 2. artificial intelligence
- 3. data mining
- 4. internet of things

21. 18. Which learning Requires Self-Assessment to identify patterns within data? * 1 point

Mark only one oval.

- 1. supervised learning
- 2. unsupervised learning
- 3. semi supervised learning
- 4. reinforced learning

22. 19. In simple term, machine learning is * 1 point

Mark only one oval.

- 1. prediction to answer a query
- 2. training based on historical data
- 3. All of above
- 4. None of above

23. 20. Of the Following Examples, Which would you address using an supervised learning Algorithm? * 1 point

Mark only one oval.

- 1. given a set of news articles found on the web, group them into set of articles about the same story
- 2. given email labeled as spam or not spam, learn a spam filter
- 3. given a database of customer data, automatically discover market segments and group customers into different market segments
- 4. find the patterns in market basket analysis

24. 21. If machine learning model output doesn't involves target variable then that model is called as * 1 point

Mark only one oval.

- 1. predictive model
- 2. descriptive model
- 3. reinforcement learning
- 4. all of the above

25. 22. In what type of learning labelled training data is used * 1 point

Mark only one oval.

- 1. supervised learning
- 2. unsupervised learning
- 3. reinforcement learning
- 4. active learning

26. 23.In the example of predicting number of babies based on stork's population ,Number of babies is * 1 point

Mark only one oval.

- 1. feature
- 2. observation
- 3. outcome
- 4. attribute

27. 24. Following are the descriptive models * 1 point

Mark only one oval.

- 1. classification
- 2. clustering
- 3. association rule
- 4. Both 1 and 2

28. 25. In following type of feature selection method we start with empty feature set * 1 point

Mark only one oval.

- 1. backward feature selection
- 2. forward feature selection
- 3. All of above
- 4. None of above

29. 26. A person trained to interact with a human expert in order to capture their knowledge. * 1 point

Mark only one oval.

- 1. knowledge developer
- 2. knowledge programmer
- 3. knowledge engineer
- 4. knowledge extractor

30. 27. What characterize unlabeled examples in machine learning * 1 point

Mark only one oval.

- 1. there is plenty of confusing knowledge
- 2. there is prior knowledge
- 3. there is no confusing knowledge
- 4. there is no prior knowledge

31. 28. What does dimensionality reduction reduce? * 1 point

Mark only one oval.

- 1. collinearity
- 2. stochastic
- 3. entropy
- 4. performance

32. 29. Some telecommunication company wants to segment their customers into distinct groups ,this is an example of * 1 point

Mark only one oval.

- 1. supervised learning
- 2. unsupervised learning
- 3. data extraction
- 4. reinforcement learning

33. 30. Which of the following is the best machine learning method? * 1 point

Mark only one oval.

- 1. accuracy
- 2. scalable
- 3. fast
- 4. All of above

34. 31. In multiclass classification number of classes must be * 1 point

Mark only one oval.

- 1. equals to two
- 2. less than two
- 3. greater than two
- 4. None

35. 32. Which of the following can only be used when training data are linearly separable? * 1 point

Mark only one oval.

- 1. linear logistic regression
- 2. linear hard-margin svm
- 3. linear soft margin svm
- 4. parzen windows

36. 33. Which of the following can only be used when training data are linearly separable? * 1 point

Mark only one oval.

- 1. linear logistic regression
- 2. linear soft margin svm
- 3. linear hard-margin svm
- 4. the centroid method

37. 34. You are given seismic data and you want to predict next earthquake , this is an example * 1 point

Mark only one oval.

- 1. supervised learning
- 2. unsupervised learning
- 3. reinforcement learning
- 4. dimensionality reduction
- Other: _____

38. 35. Prediction is * 1 point

Mark only one oval.

- 1. discipline in statistics used to find projections in multidimensional data
- 2. value entered in database by expert
- 3. the result of application of specific theory or rule in a specific case
- 4. independent of data

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

Department of Computer Science & Engineering

Assignment Sheet-I

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CAI4-02)
Max Marks- 40

(Set-I)

<table>
<tr><th>Q.No</th><th>Question</th><th>BL</th><th>CO</th><th>MM</th></tr>
<tr><td>1</td><td>Build a decision tree using ID3 algorithm for the given training data in the table (Buy Computer data), and predict the class of the following new example: age<=30, income=medium, student=yes, credit-rating=fair
<table>
<tr><th>age</th><th>income</th><th>student</th><th>Credit rating</th><th>Buys computer</th></tr>
<tr><td><=30</td><td>high</td><td>no</td><td>fair</td><td>no</td></tr>
<tr><td><=30</td><td>high</td><td>no</td><td>excellent</td><td>no</td></tr>
<tr><td>31...40</td><td>high</td><td>no</td><td>fair</td><td>yes</td></tr>
<tr><td>>40</td><td>medium</td><td>no</td><td>fair</td><td>yes</td></tr>
<tr><td>>40</td><td>low</td><td>yes</td><td>fair</td><td>yes</td></tr>
<tr><td>>40</td><td>low</td><td>yes</td><td>excellent</td><td>no</td></tr>
<tr><td>31...40</td><td>low</td><td>yes</td><td>excellent</td><td>yes</td></tr>
<tr><td><=30</td><td>medium</td><td>no</td><td>fair</td><td>no</td></tr>
<tr><td><=30</td><td>low</td><td>yes</td><td>fair</td><td>yes</td></tr>
<tr><td>>40</td><td>medium</td><td>yes</td><td>fair</td><td>yes</td></tr>
<tr><td><=30</td><td>medium</td><td>yes</td><td>excellent</td><td>yes</td></tr>
<tr><td>31...40</td><td>medium</td><td>no</td><td>excellent</td><td>yes</td></tr>
<tr><td>31...40</td><td>high</td><td>yes</td><td>fair</td><td>yes</td></tr>
<tr><td>>40</td><td>medium</td><td>no</td><td>excellent</td><td>no</td></tr>
</table></td><td>3</td><td>1</td><td>10</td></tr>
<tr><td>2</td><td>Find the class label for the given test data point as given in the table using Naive Bayes learning algorithm
<table>
<tr><th></th><th>Doc</th><th>Class</th></tr>
<tr><td rowspan="4">Training</td><td>Chinese Beijing Chinese</td><td>c</td></tr>
<tr><td>Chinese Chinese Shanghai</td><td>c</td></tr>
<tr><td>Chinese Macao</td><td>c</td></tr>
<tr><td>Tokyo Japan Chinese</td><td>j</td></tr>
<tr><td>Test</td><td>Chinese Chinese Chinese Tokyo Japan</td><td>?</td></tr>
</table>
Here p(w|class) = (count(w,c) + 1)/ (count(c) +|V|) where |V| Number of unique words in the documents and ‘w’ is the word in the document.</td><td>3</td><td>1</td><td>10</td></tr>
<tr><td>3</td><td>How to use K-Means clustering algorithm to divide the given dataset into different cluster. Given data points are A1(2,10),A2(2,5),A3(8,4),B1(5,8),B2(7,5),B3(6,4),C1(1,2),c4(4,9),use</td><td>3</td><td>2</td><td>10</td></tr>
</table>

	Euclidean Distance function, and assume that A1,B1,C1 are the center for each cluster.			
4	What is Gaussian Mixture Model or Mixture of Gaussian GMM)?Compare GMM with Clustering (any 4 points).	3	2	10

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

Department of Computer Science & Engineering

Assignment Sheet-I

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CAI4-02)
Max Marks- 40

SET-II

<table>
<tr><th>Q.No</th><th>Question</th><th>BL</th><th>CO</th><th>MM</th></tr>
<tr><td>1</td><td>Explain the concepts of Entropy and Information Gain in Decision Tree Learning with example</td><td>3</td><td>1</td><td>10</td></tr>
<tr><td>2</td><td>How to Apply K nearest neighbor classifier to predict the diabetic patient with the given features BMI, Age. If the training examples are as given in table
<table>
<tr><th>BMI</th><th>Age</th><th>Sugar</th></tr>
<tr><td>33.6</td><td>50</td><td>1</td></tr>
<tr><td>26.6</td><td>30</td><td>O</td></tr>
<tr><td>23.4</td><td>40</td><td>O</td></tr>
<tr><td>43.1</td><td>67</td><td>O</td></tr>
<tr><td>35.3</td><td>23</td><td>1</td></tr>
<tr><td>35.9</td><td>67</td><td>1</td></tr>
<tr><td>36.7</td><td>45</td><td>1</td></tr>
<tr><td>25.7</td><td>46</td><td>O</td></tr>
<tr><td>23.3</td><td>29</td><td>O</td></tr>
<tr><td>31</td><td>56</td><td>1</td></tr>
</table></td><td>4</td><td>1</td><td>10</td></tr>
<tr><td>3</td><td>Consider the following data point :p1(0,2,3), p2(0, 1, 2), p3(3,0,5), p4(4,1,3),and p5(5,0,1) .Apply the K-means clustering algorithm ,to group these data points into 2 clusters using L1 distance measure.</td><td>3</td><td>2</td><td>10</td></tr>
<tr><td>4</td><td>How to appl FP growth algorithm to Find the frequent dataset and association rule for the given transition data set (assume that the minimum support is 3)
<table>
<tr><th>Transaction ID</th><th>Items</th></tr>
<tr><td>T1</td><td>{E,K,M,N,O,Y}</td></tr>
<tr><td>T2</td><td>{D,E,K,N,O,Y}</td></tr>
<tr><td>T3</td><td>{A,E,K,M}</td></tr>
<tr><td>T4</td><td>{C,K,M,U,Y}</td></tr>
<tr><td>T5</td><td>{C,E,I,K,O,O}</td></tr>
</table></td><td>4</td><td>2</td><td>10</td></tr>
</table>

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

Department of Computer Science & Engineering

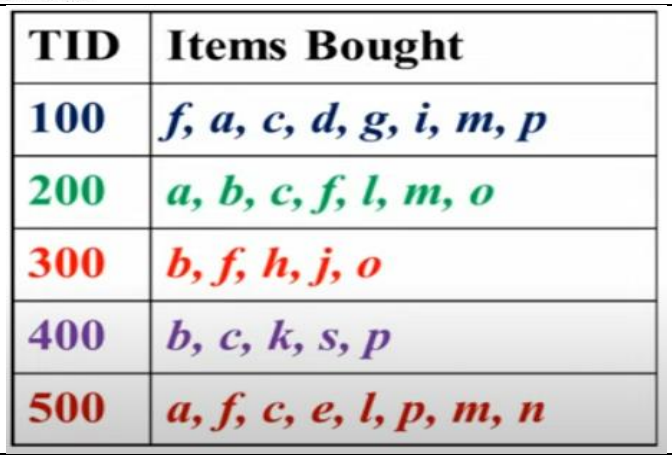
Assignment Sheet-I

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CAI4-02)
Max Marks- 40

SET-III

<table>
<tr><th>Q.No</th><th>Question</th><th>BL</th><th>CO</th><th>MM</th></tr>
<tr><td>1</td><td>how to apply Classification And Regression Trees (CART) decision tree algorithm (Solved Example 3) to construct and find the optimal decision tree for the given Data set with City Size, Avg. Income, Local Investors, LOHAS Awareness attributes. Also, predict the class label for the given example…?
<table>
<tr><th>City Size</th><th>Avg. Income</th><th>Local Investors</th><th>LOHAS Awareness</th><th>Decision</th></tr>
<tr><td>Big</td><td>High</td><td>Yes</td><td>High</td><td>Yes</td></tr>
<tr><td>Medium</td><td>Med</td><td>No</td><td>Med</td><td>No</td></tr>
<tr><td>Small</td><td>Low</td><td>Yes</td><td>Low</td><td>No</td></tr>
<tr><td>Big</td><td>High</td><td>No</td><td>High</td><td>Yes</td></tr>
<tr><td>Small</td><td>Med</td><td>Yes</td><td>High</td><td>No</td></tr>
<tr><td>Med</td><td>High</td><td>Yes</td><td>Med</td><td>Yes</td></tr>
<tr><td>Med</td><td>Med</td><td>Yes</td><td>Med</td><td>No</td></tr>
<tr><td>Big</td><td>Med</td><td>No</td><td>Med</td><td>No</td></tr>
<tr><td>Med</td><td>High</td><td>Yes</td><td>Low</td><td>No</td></tr>
<tr><td>Small</td><td>High</td><td>No</td><td>High</td><td>Yes</td></tr>
<tr><td>Small</td><td>Med</td><td>No</td><td>High</td><td>No</td></tr>
<tr><td>Med</td><td>Heigh</td><td>No</td><td>Med</td><td>No</td></tr>
<tr><th>City Size</th><th>Avg. Income</th><th>Local Investors</th><th>LOHAS Awareness</th><th>Decision</th></tr>
<tr><td>Med</td><td>Med</td><td>No</td><td>Med</td><td>?</td></tr>
</table></td><td>3</td><td>1</td><td>10</td></tr>
<tr><td>2</td><td>How to apply Linear regression model to predict the glucose level given the age .
<table>
<tr><th>SUBJECT</th><th>AGE X</th><th>GLUCOSE LEVEL Y</th></tr>
<tr><td>1</td><td>43</td><td>99</td></tr>
<tr><td>2</td><td>21</td><td>65</td></tr>
<tr><td>3</td><td>25</td><td>79</td></tr>
<tr><td>4</td><td>42</td><td>75</td></tr>
<tr><td>5</td><td>57</td><td>87</td></tr>
<tr><td>6</td><td>59</td><td>81</td></tr>
</table></td><td>3</td><td>1</td><td>10</td></tr>
<tr><td>3</td><td>How to appl FP growth algorithm to Find the frequent dataset and association rule for the given transition data set (assume that the minimum support is 3)</td><td>4</td><td>2</td><td>10</td></tr>
</table>

				
4	What is Gaussian Mixture Model or Mixture of Gaussian GMM)?Compare GMM with Clustering (any 4 points).	3	2	10

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

Department of Computer Science & Engineering

Subject: Machine Learning (6CAI4-02)
Max Marks- 40

SET-IV

<table>
<tr><th>Q.No</th><th>Question</th><th>BL</th><th>CO</th><th>MM</th></tr>
<tr><td>1</td><td>Explain the concepts of Entropy and Information Gain in Decision Tree Learning with example
<table>
<tr><th></th><th>Pepper</th><th>Ginger</th><th>Chilly</th><th>Liked</th></tr>
<tr><td>A</td><td>True</td><td>True</td><td>True</td><td>False</td></tr>
<tr><td>B</td><td>True</td><td>False</td><td>False</td><td>True</td></tr>
<tr><td>C</td><td>False</td><td>True</td><td>True</td><td>False</td></tr>
<tr><td>D</td><td>False</td><td>True</td><td>False</td><td>True</td></tr>
<tr><td>E</td><td>True</td><td>False</td><td>False</td><td>True</td></tr>
</table></td><td>3</td><td>1</td><td>10</td></tr>
<tr><td>2</td><td>“Restaurant A” sells burgers with optional flavors: Pepper, Ginger, and Chilly. Every day this week you have tried a burger (A to E) and kept a record of which you liked.

Using Hamming distance, show how the 3NN classifier with majority voting would classify { pepper: false, ginger: true, chilly: true}.</td><td>4</td><td>1</td><td>10</td></tr>
<tr><td>3</td><td>Explain how the market basket analysis uses the concept of association analysis.</td><td>3</td><td>2</td><td>10</td></tr>
<tr><td>4</td><td>Consider A1,B1 and C1 as initial cluster, if you want to run a second iteration then what will be the cluster centriods using k-means algorithm.

A1(2, 10), A2(2, 5), A3(8, 4), B1(5, 8), B2(7, 5), B3(6, 4), C1(1, 2), C2(4, 9).</td><td>4</td><td>2</td><td>10</td></tr>
</table>

Assignment Sheet-I

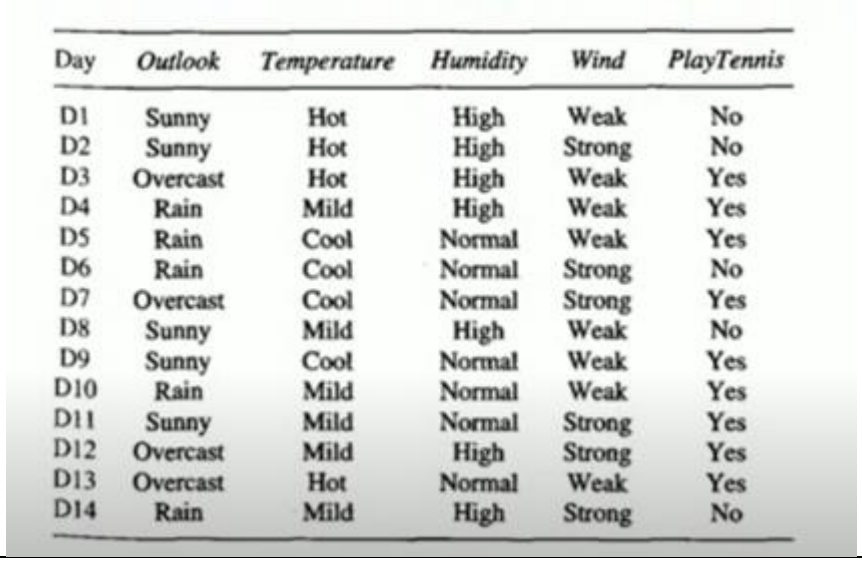
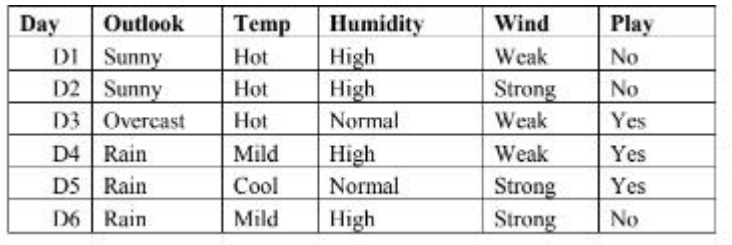
Session 2023_24 (Even Semester)

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

Department of Computer Science & Engineering

Subject: Machine Learning (6CAI4-02)
Max Marks- 40

SET-V

Q.No	Question	BL	CO	MM
1	How to use the naive Bayes classifier and the training data from this table to classify the following novel instance: (Outlook=Sunny, Temp=Cool, Humidity=High, Wind= Strong) 	4	1	10
2	Consider the following dataset in the table. Make the decision tree using GiniIndex. (Play is target class). 	4	1	10
3	Explain how the market basket analysis uses the concept of association analysis.	3	2	10
4	1. Use Apriori algorithm to make the rules when minimum support threshols =33.34% andconfidence threshold =60% given	4	2	10

Day	*Outlook*	*Temperature*	*Humidity*	*Wind*	*PlayTennis*
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Day	Outlook	Temp	Humidity	Wind	Play
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	Normal	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Strong	Yes
D6	Rain	Mild	High	Strong	No

T. ID	Items
T1	HotDogs, Buns, Ketchup
T2	HotDogs, Buns
T3	HotDogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	HotDogs, Coke, Chips

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

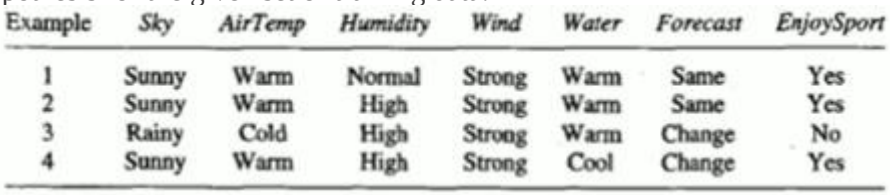
Department of Computer Science & Engineering

Assignment Sheet-I

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CAI4-02)
Max Marks- 40

SET-VI

Q.No	Question	BL	CO	MM
1	Estimate conditional probabilities of each attributes (color, leg, height, smelly) for the species class{M, H} using the data given in the table. using these probabilities estimate the probability value for the new instance (color=Green, Legs=2,Height= Tall, Smelly =NO} 	4	1	10
2	How to apply candidate Elimination algorithm to find the consistent hypothesis for the given set of training data. 	4	1	10
3	Explain the concept of clustering with a neat diagram. What is dendrogram? explain its use.	3	2	10
4	How to use the association rule mining algorithm to find the affinities between products ,that is which product sell together often.Assume that the support level is 33% and confidence level is 50 %. 	3	2	10

No	Color	Legs	Height	Smelly	Species
1	White	3	Short	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	Yes	H

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

1	Milk	Egg	Bread	Butter
2	Milk	Butter	Egg	Ketchup
3	Bread	Butter	Ketchup	
4	Milk	Bread	Butter	
5	Bread	Butter	Cookies	
6	Milk	Bread	Butter	Cookies
7	Milk	Cookies		
8	Milk	Bread	Butter	
9	Bread	Butter	Egg	Cookies
10	Milk	Butter	Bread	
11	Milk	Bread	Butter	
12	Milk	Bread	Cookies	Ketchup

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

Department of Computer Science & Engineering

Assignment Sheet-I

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CAI4-02)
Max Marks- 40

SET-VII

<table>
<tr><th>Q.No</th><th>Question</th><th>BL</th><th>CO</th><th>MM</th></tr>
<tr><td>1</td><td>What is inductive bias in a decision tree? How pruning apply in decision tree to avoid overfitting?</td><td>3</td><td>1</td><td>10</td></tr>
<tr><td>2</td><td>Consider the following dataset in the table. Make the decision tree using GiniIndex. (Play is target class).
<table><tr><th>Day</th><th>Outlook</th><th>Temp</th><th>Humidity</th><th>Wind</th><th>Play</th></tr><tr><td>D1</td><td>Sunny</td><td>Hot</td><td>High</td><td>Weak</td><td>No</td></tr><tr><td>D2</td><td>Sunny</td><td>Hot</td><td>High</td><td>Strong</td><td>No</td></tr><tr><td>D3</td><td>Overcast</td><td>Hot</td><td>Normal</td><td>Weak</td><td>Yes</td></tr><tr><td>D4</td><td>Rain</td><td>Mild</td><td>High</td><td>Weak</td><td>Yes</td></tr><tr><td>D5</td><td>Rain</td><td>Cool</td><td>Normal</td><td>Strong</td><td>Yes</td></tr><tr><td>D6</td><td>Rain</td><td>Mild</td><td>High</td><td>Strong</td><td>No</td></tr></table></td><td></td><td>1</td><td>10</td></tr>
<tr><td>3</td><td>How to apply the apriori algorithm to generate the strong association rule for the given transition assumethat min_support =40% and min_confidence =70%
<table><tr><th>Transaction ID</th><th>Items Bought</th></tr><tr><td>1</td><td>{Bread, Butter, Milk}</td></tr><tr><td>2</td><td>{Bread, Butter}</td></tr><tr><td>3</td><td>{Beer, Cookies, Diapers}</td></tr><tr><td>4</td><td>{Milk, Diapers, Bread, Butter}</td></tr><tr><td>5</td><td>{Beer, Diapers}</td></tr></table></td><td>4</td><td>2</td><td>10</td></tr>
<tr><td>4</td><td>How apriori principal help in reducing the calculation overhead for a market basket analysis? Provide an example to explain.</td><td>3</td><td>2</td><td>10</td></tr>
</table>

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

Department of Computer Science & Engineering

ssignment Sheet-I

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CAI4-02)
Max Marks- 40

(Set-VIII)

Q.No	Question	BL	CO	MM
1	Elaborate are the important objectives of machine learning?	3	1	10
2	Give decision trees to represent the following boolean functions: (a) A ∧˜B (b) A V [B ∧ C] (c) A XOR B (d) [A ∧ B] v [C ∧ D]	6	1	10
3	Assign the response variable (Target Class) to the testing object x(1,1) using K Nearest Neighbor algorithm. Observations are depicted below: X1(3, 0, A), X2(4, 1, A), X3(3, 2, A), X4(1,-1, B), X5(1,-1.5,B) where A and B are the target class variables. Assume K=3.Write down algorithm to implement addition and transposition operation of sparse matrix.	3	2	10
4	Use Apriori algorithm to make the rules when minimum support thresholds =33.34% and confidence threshold =60% given 	3	2	10

T. ID	Items
T1	HotDogs, Buns, Ketchup
T2	HotDogs, Buns
T3	HotDogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	HotDogs, Coke, Chips

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

Department of Computer Science & Engineering

Assignment Sheet-I

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CDS4-02)
Max Marks- 40

(Set-IX)

<table>
<tr><th>Q.No</th><th>Question</th><th>BL</th><th>CO</th><th>MM</th></tr>
<tr><td>1</td><td>Illustrate the concept of Bagging and OOB error rate with respect to Random Forest Classifier.</td><td>4</td><td>1</td><td>10</td></tr>
<tr><td>2</td><td>Explain the broad three categories of clustering techniques? Explain the characteristicsof each.</td><td>3</td><td>1</td><td>10</td></tr>
<tr><td>3</td><td>How to apply the apriori algorithm to generate the strong association rule for the given transition assumethat min_support =40% and min_confidence =70%

<table><tr><th>Transaction ID</th><th>Items Bought</th></tr><tr><td>1</td><td>Bread, Butter, Milk</td></tr><tr><td>2</td><td>Bread, Butter</td></tr><tr><td>3</td><td>Beer, Cookies, Diapers</td></tr><tr><td>4</td><td>Milk, Diapers, Bread, Butter</td></tr><tr><td>5</td><td>Beer , Diapers</td></tr></table></td><td>6</td><td>2</td><td>10</td></tr>
<tr><td>4</td><td>Consider the following dataset in the table. Make the decision tree using GiniIndex. (Play is target class).

<table><tr><th>Day</th><th>Outlook</th><th>Temp</th><th>Humidity</th><th>Wind</th><th>Play</th></tr><tr><td>D1</td><td>Sunny</td><td>Hot</td><td>High</td><td>Weak</td><td>No</td></tr><tr><td>D2</td><td>Sunny</td><td>Hot</td><td>High</td><td>Strong</td><td>No</td></tr><tr><td>D3</td><td>Overcast</td><td>Hot</td><td>Normal</td><td>Weak</td><td>Yes</td></tr><tr><td>D4</td><td>Rain</td><td>Mild</td><td>High</td><td>Weak</td><td>Yes</td></tr><tr><td>D5</td><td>Rain</td><td>Cool</td><td>Normal</td><td>Strong</td><td>Yes</td></tr><tr><td>D6</td><td>Rain</td><td>Mild</td><td>High</td><td>Strong</td><td>No</td></tr></table></td><td>4</td><td>2</td><td>10</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

Department of Computer Science & Engineering

Assignment Sheet-I

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CDS4-02)
Max Marks- 40

(Set-X)

<table>
<tr><th>Q.No</th><th>Question</th><th>BL</th><th>CO</th><th>MM</th></tr>
<tr><td>1</td><td>Illustrate the Random Forest algorithms.</td><td>3</td><td>1</td><td>10</td></tr>
<tr><td>2</td><td>Define the Association rule mining in detail.</td><td>4</td><td>1</td><td>10</td></tr>
<tr><td>3</td><td>Find the class label for the given test data point as given in the table using Naive Bayes learning algorithm.
<table><tr><th></th><th>Doc</th><th>Class</th></tr><tr><td rowspan="4">Training</td><td>Chinese Beijing Chinese</td><td>c</td></tr><tr><td>Chinese Chinese Shanghai</td><td>c</td></tr><tr><td>Chinese Macao</td><td>c</td></tr><tr><td>Tokyo Japan Chinese</td><td>j</td></tr><tr><td>Test</td><td>Chinese Chinese Chinese Tokyo Japan</td><td>?</td></tr></table>Here p(w|class) = (count(w,c) + 1)/ (count(c) +|V|) where |V| Number of unique words in the documents and ‘w’ is the word in the document.</td><td>4</td><td>2</td><td>10</td></tr>
<tr><td>4</td><td>Consider A1,B1 and C1 as initial cluster, if you want to run a second iteration then what will be the cluster centroids using k-means algorithm.
A1(2, 10), A2(2, 5), A3(8, 4), B1(5, 8), B2(7, 5), B3(6, 4), C1(1, 2), C2(4, 9).</td><td>4</td><td>2</td><td>10</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

Department of Computer Science & Engineering

Assignment Sheet-I

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CDS4-02)
Max Marks- 40

(Set-XI)

Q.No	Question	BL	CO	MM
1	Explain Naïve Bayes Classifier with an Example	3	1	10
2	Define the following terms with respect to K - Nearest Neighbour Learning : i) Regression ii) Residual iii) Kernel Function.	3	1	10
3	How to apply candidate Elimination algorithm to find the consistent hypothesis for the given set of training data.	3	2	10
4	Given one dimension data set {1,5,8,10,2},use the agglomerative clustering algorithm with the complete link with Euclidean distance to establish a hierarchical grouping relation.	3	2	10

Example	*Sky*	*AirTemp*	*Humidity*	*Wind*	*Water*	*Forecast*	*EnjoySport*
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

Department of Computer Science & Engineering

Assignment Sheet-I

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CDS4-02)
Max Marks- 40

(Set-XII)

<table>
<tr><th>Q.No</th><th>Question</th><th>BL</th><th>CO</th><th>MM</th></tr>
<tr><td>1</td><td>Differentiate between Supervised, Unsupervised and Reinforcement Learning</td><td>3</td><td>1</td><td>10</td></tr>
<tr><td>2</td><td>Write the steps of ID3Algorithm</td><td>3</td><td>1</td><td>10</td></tr>
<tr><td>3</td><td>How to apply the apriori algorithm to generate the strong association rule for the given transition assumethat min_support =40% and min_confidence =70%
<table>
<tr><th>Image ID</th><th>Associated Tags</th></tr>
<tr><td>1</td><td>Beach, Sunshine, Holiday</td></tr>
<tr><td>2</td><td>Sand, Beach</td></tr>
<tr><td>3</td><td>Sunshine, Beach, Ocean</td></tr>
<tr><td>4</td><td>Ocean, People, Beach, Sunshine</td></tr>
<tr><td>5</td><td>Holiday, Sunshine</td></tr>
</table></td><td>4</td><td>2</td><td>10</td></tr>
<tr><td>4</td><td>Calculate various distances between three records A, B and C, with two features.
<table>
<tr><td></td><td>Feature1</td><td>Feature2</td></tr>
<tr><td>A</td><td>45</td><td>0.05</td></tr>
<tr><td>B</td><td>60</td><td>0.05</td></tr>
<tr><td>C</td><td>52</td><td>0.09</td></tr>
</table></td><td>4</td><td>2</td><td>10</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
</table>

.

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

Department of Computer Science & Engineering

Assignment Sheet-I

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CDS4-02)
Max Marks- 40

(Set-XIII)

<table>
<tr><th>Q.No</th><th>Question</th><th>BL</th><th>CO</th><th>MM</th></tr>
<tr><td>1</td><td>Explain the following a)Linear regression b) Logistic Regression</td><td>3</td><td>1</td><td>10</td></tr>
<tr><td>2</td><td>Define clustering. What are the different types of clustering explain in detail?</td><td>3</td><td>1</td><td>10</td></tr>
<tr><td>3</td><td>Suppose you have a database on four customers. You know their gender, income and whether or not they bought your product.
<table>
<tr><th>S. No</th><th>Income</th><th>Gender</th><th>Buy Product</th></tr>
<tr><td>1</td><td>HIGH</td><td>MALE</td><td>YES</td></tr>
<tr><td>2</td><td>HIGH</td><td>FEMALE</td><td>NO</td></tr>
<tr><td>3</td><td>MEDIUM</td><td>FEMALE</td><td>NO</td></tr>
<tr><td>4</td><td>LOW</td><td>MALE</td><td>YES</td></tr>
</table>
Using Naïve Baye’s algorithm predicts the probability of buying product given that a customer has a high income and is male.</td><td>4</td><td>2</td><td>10</td></tr>
<tr><td>4</td><td>Apply K-mean algorithm on given data for K=2 use C1(4) and C2(12) as initial cluster centers Data : 2,3,4,10,11,12,20,25,30,35</td><td>4</td><td>2</td><td>10</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
</table>

.

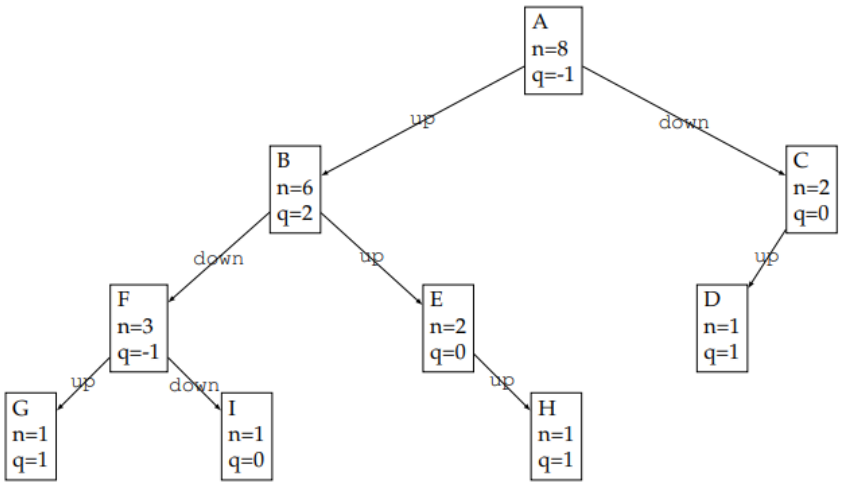
.

Assignment Sheet-II

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CAI-02)
Max Marks- 40

(Set-I)

Q.No	Question	BL	CO	MM
1	What is the curse of Dimensionality in machine learning? How PCA is used for dimensionality reduction?	3	3	10
2	6. You are using Monte Carlo Tree Search to decide on the next action for a two-person competitive game with 2 actions at each state (up and down). It is player 1's turn to play in state A. The state of the tree so far is as follows (each node consists of state identifier, n value, and q value):  Remember that the formula for the UCT value for a node, v, is: $$UCT(v) = \frac{q(v)}{n(v)} + c\sqrt{\frac{\ln n(v.parent)}{n(v)}}$$ Assume the constant c in the UCT formula is 0.5. a. What is the node that is next selected (show your work)? b. What are the details of the node that gets expanded from the selected node? c. Assuming that the simulation (rollout) from the expanded node gives a value of 1 (that is, player 1 wins), backup that value to all of the affected nodes. d. Assume that after this final rollout, we've run out of time to run the MCTS simulation and must now choose an action for player 1 from state A. What action will be chosen and why?	4	4	10
3	(a)What is the difference between Forward propagation and Backward Propagation in Neural Networks? (b)Explain any four activation functions used in a neural network.	3	5	10
4	Find the Singular Value Decomposition (SVD) of A, UΣV T , **where A = ([3, 2, 2], [2, 3, −2], [5, 2, 9])**	3	3	10

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

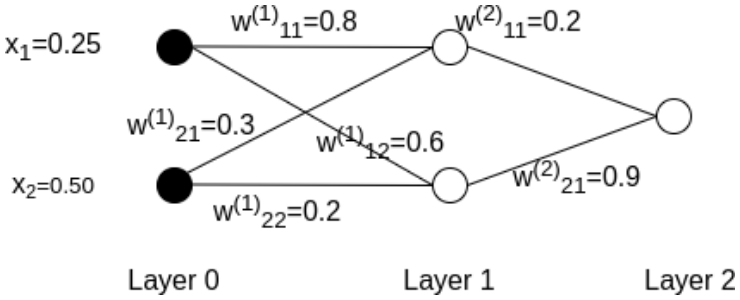
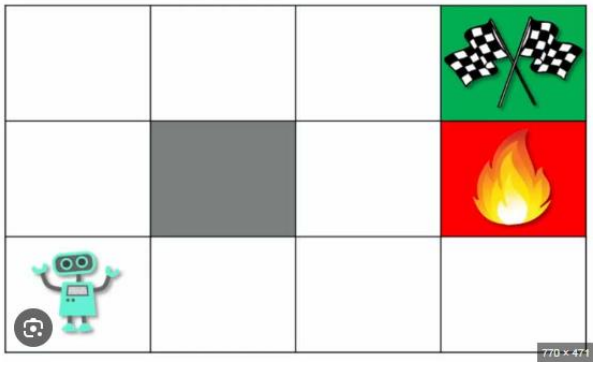
Department of Computer Science & Engineering

Assignment Sheet-I

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CAI4-02)
Max Marks- 40

(Set-II)

Q.No	Question	BL	CO	MM
1	Compare Feature Extraction and Feature Selection techniques. Explain how dimensionality can be reduced using a subset selection procedure	3	3	10
2	What is HMM? Explain HMM with suitable diagram	3	4	10
3	Consider the following Neural Network diagram and find the updated parameters of the network using oneiteration. Sigmoid activation function is used at hidden and output layer. Consider only the parameters given inthe network. (use η = 0.1 and target output = 0.7). 	3	5	10
4	find the optimal solution using the Bellman equation. Also update the table and find value at each state of the given environment. Here Flag represent the Reward(+1) and Fire(-1). 	3	4	10

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

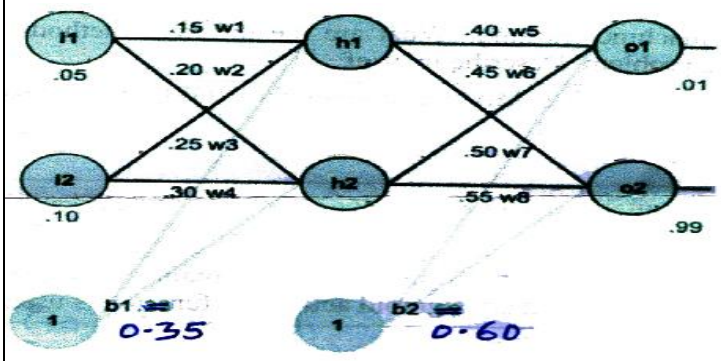
Department of Computer Science & Engineering

Assignment Sheet-I

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CAI4-02)
Max Marks- 40

(Set-III)

Q.No	Question	BL	CO	MM
1	(a)What is training, testing and cross validation of machine learning models (b) What is overfitting and underfitting? What are the catalysts of overfitting? c) Elaborate Bias Variance dilemma.	3	3	10
2	What s Temporal Learning Reinforcement? When using Temporal Difference learning, why is it better to learn action values (Qvalues) rather than state values (V -values)?	3	4	10
3	(a) Explain the differences between Collaborative filtering and Content based filtering (b)What is TF and IDF? How document is represented with TF-IDF explain with an example.	3	5	10
4	Refer the neural network given in the figure. Determine the updated weight w5 using backpropagation algorithm. Consider the learning rate $\alpha = 0.5$. (Refer given figure).  figure	4	5	10

Assignment Sheet-I

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CAI4-02)
Max Marks- 40

(Set-IV)

Q.No	Question	BL	CO	MM
1	Compare Feature Extraction and Feature Selection techniques. Explain how dimensionality can be reduced using a subset selection procedure	3	3	10
2	How exactly a Markov Decision Process work in reinforcement learning?	3	4	10
3	How deep learning overcomes the challenges in conventional machine learning techniques? Draw and explain the architecture of different Neural networks.	3	5	10
4	**Compute the principal component of following data-** **CLASS 1: X = 2 , 3 , 4 Y = 1 , 5 , 3** **CLASS 2: X = 5 , 6 , 7. Y = 6 , 7 , 8**	3	3	10

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

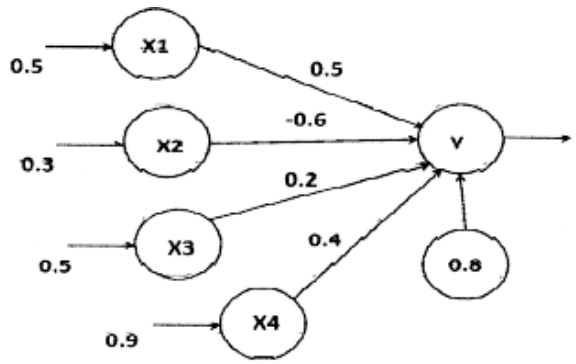
Department of Computer Science & Engineering

Assignment Sheet-I

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CAI4-02)
Max Marks- 40

(Set-V)

Q.No	Question	BL	CO	MM
1	**Explain the following terms:** (i) Precision (ii) Recall (iii) F-score (iv) Sensitivity (v) Specificity	3	3	10
2	You are given an environment with 1 state, x, and 2 actions, b and c. T is the terminal state. Your TD algorithm generates the following episode using the policy π when interacting with its environment: **Timestep | Reward | State | Action** 0 | | x | b 1 | 16 | x | c 2 | 12 | x | b 3 | 16 | T | • The policy π is given by: $\pi(b\|x) = 0.9, \pi(c\|x) = 0.1$ • The current values of q are: $q(x, b) = 1$ and $q(x, c) = 2$. • the discount factor, γ, is $\frac{1}{2}$. • the step size, α, is 0.1 Show the values of $q(x, b)$ and $q(x, c)$ after their *first* update using 1-step Sarsa, 2-step Sarsa, 2-step Expected Sarsa, and 2-step Tree Backup. Note: you should update $q(x, b)$ and $q(x, c)$ only once per learning algorithm. **Show your work** and carry out your calculations to *two* decimal places.	3	4	10
3	What is the significance of bias in perceptron networks? Obtain the output of neuron Y for the given network if binary sigmoid activation function is used. (Refer figure 1).  figure – 1	4	5	10
4	List various environments for an agent. Compare between following environments. [7] a) Fully observable versus partially observable b) Deterministic versus stochastic	4	4	10

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

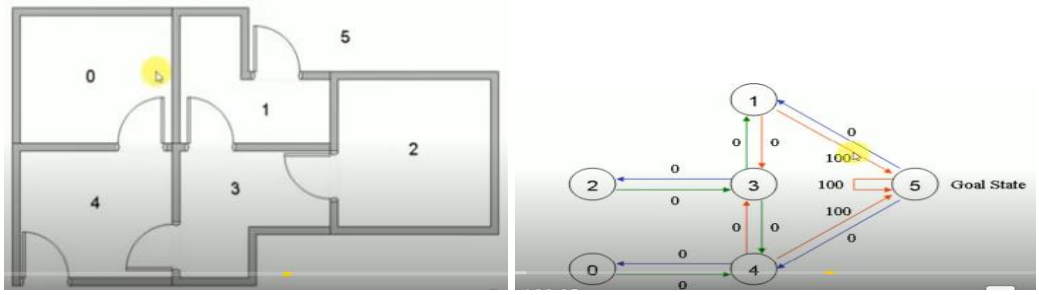
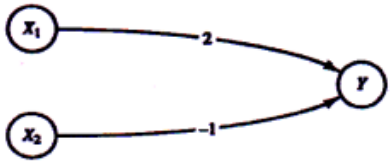
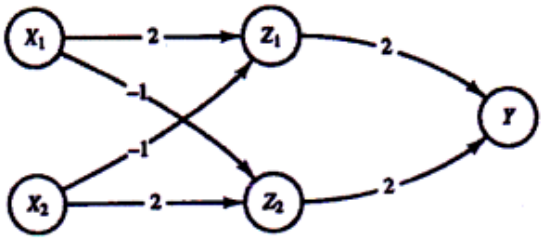
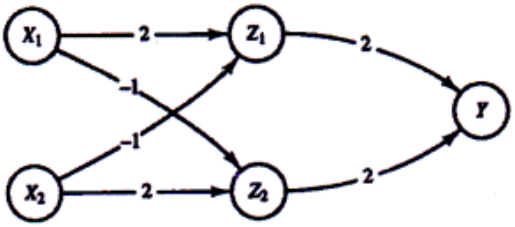
Department of Computer Science & Engineering

Assignment Sheet-I

Session 2023_24 (Even Semester)

Subject: Machine Learning (6CAI4-02)
Max Marks- 40

(Set-VI)

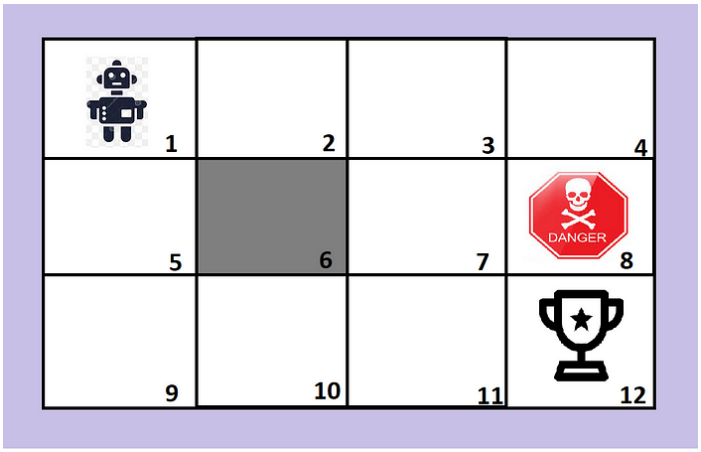
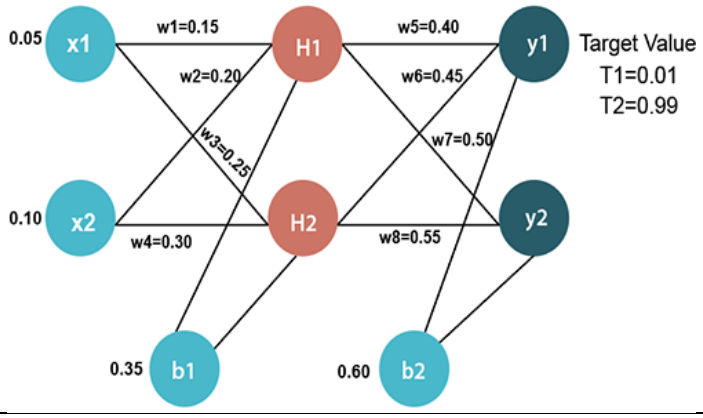
Q.No	Question	BL	CO	MM
1	What is Feature Engineering? How do you apply Feature Transformation and feature selection in the machine Learning	3	3	10
2	Suppose we have 5 rooms in a building connected by doors as shown in the fig below. Note that door 1 and 4 leads into the building from room 5(outside).Goal state is Room 5(outside). Find the optimal path for an agent using Q learning Algorithm. the reward from door 1to 5 and 4 to 5 is hundred(100), expect that all reward are equivalent to zero(0) 	4	4	10
3	Explain the architecture of feed forward neural network. Give its limitations. [8]	3	5	10
4	Fill in the blanks/Answer the following. (i) Equation for Sigmoidal basis function is _____ (ii) Following figure implement which logic  (iii) Following figure implement which logic?  (iv) Logical Expression for z1 is _____ 	4	5	10

Assignment Sheet-II

Session 2023_24 (Even Semester)

Subject: MACHINE LEARNING (6CAI4-02)
Max Marks- 40

(Set-VII)

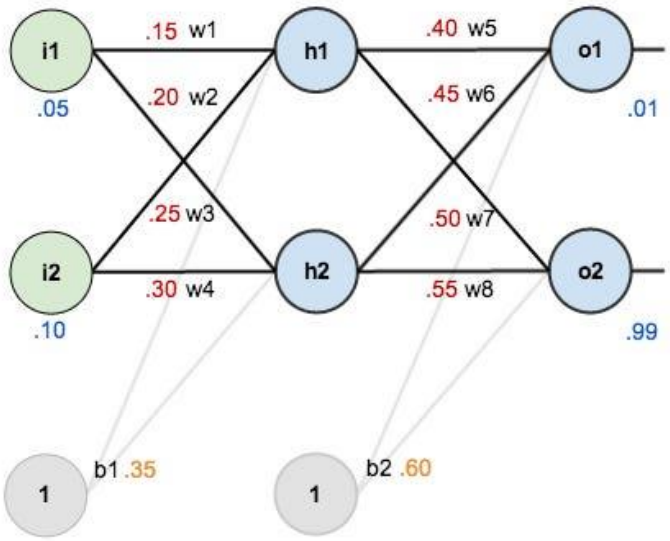
Q. No.	Question	BL	CO
1.	Illustrate Principal component analysis. Demonstrate how PCA is applied in ML.	3	3
2.	Apply bellman equation and calculate value functions for all states of the environment given below.  Rewards: 1: (for goal state), 0: (for non-goal state) -1: (for DANGER state)	3	4
3.	Apply Backpropagation algorithm for following scenario: 	3	5
4.	Illustrate Markov Decision Process (MDP). Demonstrate how MDP is applied in Reinforcement Learning.	3	4

Assignment Sheet-II

Session 2023_24 (Even Semester)

Subject: MACHINE LEARNING (6CAI4-02)
Max Marks- 40

(Set-VIII)

Q. No.	Question	BL	CO
1.	Find the Singular Value Decomposition (SVD) of A, UΣV T , where A = ([3, 2, 2], [2, 3, −2], [5, 2, 9])	3	3
2.	Apply Backpropagation algorithm for following scenario: 	3	5
3.	Differentiate between Policy evaluation and Value evaluation in Reinforcement Learning. Illustrate how these are applied.	3	4
4.	How Machine Learning algorithms are evaluated ? Demonstrate the evaluation measures required for regression problems.	3	3

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

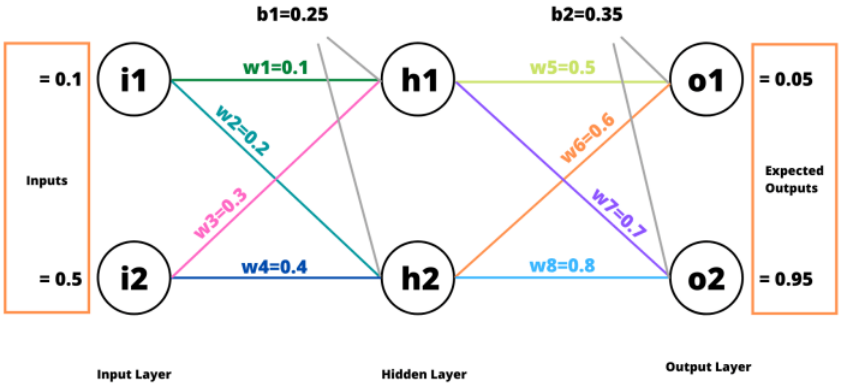
Department of Computer Science & Engineering

Assignment Sheet-II

Session 2023_24
(Even Semester)

Subject: MACHINE LEARNING (6CAI4-02)
Max Marks- 40

(Set-IX)

Q. No.	Question	BL	CO
1.	How Machine Learning algorithms are evaluated? Demonstrate the evaluation measures required for classification problems.	3	3
2.	Apply Backpropagation algorithm for following scenario: 	3	5
3.	Find the Singular Value Decomposition (SVD) of X, UΣV T , where X = ([6, -8, 2, 2], [-1, 5, 9, 0], [4, 5, -2, 9])	3	4
4.	Illustrate Q-Learning in Reinforcement Learning? Demonstrate how Q-function is evaluated?	3	4

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

Department of Computer Science & Engineering

Assignment Sheet-II

Session 2023_24

Subject: MACHINE LEARNING (6CAI4-02)
Max Marks- 40

(Set-X)

Q. No.	Question	BL	CO
1.	In a binary classification task for detecting spam emails, a model correctly predicted 450 instances as spam (True Positive), correctly predicted 850 instances as non-spam (True Negative), mispredicted 30 instances as spam (False Positive), and mispredicted 20 instances as non-spam (False Negative). Evaluate the model's performance and discuss its effectiveness in distinguishing between spam and non-spam emails.	3	3
2.	Consider a 3x3 grid maze where the agent starts at position (0, 0) and the goal is at position (2, 2). The agent can move in four directions: up, down, left, and right. The numbers in the maze represent the rewards: 100 for reaching the goal, -1 for each step taken, and -100 for hitting an obstacle. The agent aims to learn the optimal policy to navigate from the start to the goal while maximizing its total reward. Determine the optimal action based on its current state using Q-Learning.	3	4

3.	Consider a simple neural network with one input layer, hidden layer, and output layer. The input layer has two neurons, the hidden layer has two neurons, and the output layer has one neuron. We'll denote the weights connecting the neurons as w_{ij} where i represents the neuron in the previous layer and j represents the neuron in the current layer. The biases for each neuron are denoted as b_j.
	Consider the followings:
	• Input data: $x_1 = 0.5, x_2 = 0.7$
	• True label: $y = 1$ (binary classification)
	• Initial weights and biases (randomly initialized):
	  • $w_{11} = 0.3, w_{12} = 0.4, w_{21} = -0.1, w_{22} = 0.2$
	  • $b_1 = -0.2, b_2 = 0.1, b_3 = 0.05$
	• Learning rate: $\alpha = 0.1$
	Illustrate the iterative process of backpropagation.
		3	5

4.	Imagine you have a dataset containing measurements of 3 different houses, with features including: Square footage (sq ft), Number of bedrooms (bedrooms), Average income in the neighborhood ($10,000s). **House** | **Square Footage (sq ft)** | **Bedrooms** | **Average Income ($10,000s)** 1 | 2000 | 3 | 5 2 | 1500 | 2 | 7 3 | 2500 | 4 | 8 Apply PCA to reduce the dimensionality of the data while retaining most of the variance.	3	3

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

Department of Computer Science & Engineering

Subject: MACHINE LEARNING (6CAI4-02)

Max Marks- 40

(Set-XI)

<table>
<tr><th>Q. No.</th><th>Question</th><th>BL</th><th>CO</th></tr>
<tr><td>1.</td><td>An image recognition model is deployed for identifying cats and dogs in images. The model accurately classified 900 images of cats (True Positive), correctly classified 800 images without cats (True Negative), misclassified 30 images of dogs as cats (False Positive), and failed to identify 40 images of cats (False Negative). Evaluate the model's performance and discuss its reliability in animal recognition tasks.</td><td>3</td><td>3</td></tr>
<tr><td>2.</td><td>Consider a 3x3 grid maze where the agent starts at position (0, 0) and the goal is at position (2, 2). The agent can move in four directions: up, down, left, and right. The numbers in the maze represent the rewards: 100 for reaching the goal, -1 for each step taken, and -100 for hitting an obstacle. The agent aims to learn the optimal policy to navigate from the start to the goal while maximizing its total reward.

Determine the optimal action based on its current state using Q-Learning.</td><td>3</td><td>4</td></tr>
<tr><td>3.</td><td>Consider a simple neural network with one input layer, hidden layer, and output layer. The input layer has two neurons, the hidden layer has two neurons, and the output layer has one neuron. We'll denote the weights connecting the neurons as w_{ij} where i represents the neuron in the previous layer and j represents the neuron in the current layer. The biases for each neuron are denoted as b_j.

Consider the followings:

• Input data: $x_1 = 0.5, x_2 = 0.7$
• True label: $y = 1$ (binary classification)
• Initial weights and biases (randomly initialized):
  • $w_{11} = 0.3, w_{12} = 0.4, w_{21} = -0.1, w_{22} = 0.2$
  • $b_1 = -0.2, b_2 = 0.1, b_3 = 0.05$
• Learning rate: $\alpha = 0.1$

Illustrate the iterative process of backpropagation.</td><td>3</td><td>5</td></tr>
<tr><td>4.</td><td>Imagine you have a dataset containing measurements of 5 students, with features including: Hours spent studying (hours), Scores on Math exam (out of 100), Scores on English exam (out of 100)

<table>
<tr><th>Student</th><th>Hours Studied</th><th>Math Score</th><th>English Score</th></tr>
<tr><td>1</td><td>5</td><td>80</td><td>75</td></tr>
<tr><td>2</td><td>8</td><td>90</td><td>85</td></tr>
<tr><td>3</td><td>3</td><td>70</td><td>65</td></tr>
<tr><td>4</td><td>6</td><td>85</td><td>80</td></tr>
<tr><td>5</td><td>9</td><td>95</td><td>90</td></tr>
</table>
</td><td>3</td><td>3</td></tr>
</table>

	Apply PCA to reduce the dimensionality of the data while retaining most of the variance, aiming to identify underlying factors that might influence both Math and English scores.		

Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur

Department of Computer Science & Engineering

Subject: MACHINE LEARNING (6CAI4-02)
Max Marks- 40

(Set-XII)

Q. No.	Question	BL	CO
1.	A medical diagnostic model is trained to classify patients as having a certain disease or not. The model correctly identified 180 patients with the disease (True Positive), correctly identified 750 patients without the disease (True Negative), misdiagnosed 20 healthy patients as having the disease (False Positive), and missed 50 patients with the disease (False Negative). Assess the model's performance and its suitability for medical diagnosis.	3	3
2.	Consider a 3x3 grid maze where the agent starts at position (0, 0) and the goal is at position (2, 2). The agent can move in four directions: up, down, left, and right. The numbers in the maze represent the rewards: 100 for reaching the goal, -1 for each step taken, and -100 for hitting an obstacle. The agent aims to learn the optimal policy to navigate from the start to the goal while maximizing its total reward. Determine the optimal action based on its current state using Q-Learning.	3	4

3.	Consider a simple neural network with one input layer, hidden layer, and output layer. The input layer has two neurons, the hidden layer has two neurons, and the output layer has one neuron. We'll denote the weights connecting the neurons as w_{ij} where i represents the neuron in the previous layer and j represents the neuron in the current layer. The biases for each neuron are denoted as b_j.
	
	Consider the followings:
	• Input data: $x_1 = 0.5, x_2 = 0.7$
	• True label: $y = 1$ (binary classification)
	• Initial weights and biases (randomly initialized):
	  • $w_{11} = 0.3, w_{12} = 0.4, w_{21} = -0.1, w_{22} = 0.2$
	  • $b_1 = -0.2, b_2 = 0.1, b_3 = 0.05$
	• Learning rate: $\alpha = 0.1$
	
	Illustrate the iterative process of backpropagation.
		3	5

4.	Imagine you have a dataset containing measurements of 4 different fruits, with features including: Weight (grams), Sugar content (grams), Vitamin C content (mg), Here's the data: **Fruit** | **Weight** | **Sugar Content** | **Vitamin C** Apple | 180 | 10 | 8 Banana | 130 | 22 | 85	3	3

Orange	150	18	70
Mango	300	40	60

Apply PCA to reduce the dimensionality of the data while retaining most of the variance, aiming to identify potential relationships between size, sweetness, and vitamin content.

Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur

Department of Computer Science & Engineering

B.Tech./Semester: CS(AI)VI Semester **Session:** 2023-2024

Subject: Machine Learning (6CAI4-02)

Assignment Sheet: II

Part A

(Short Questions)

1. What is Machine Learning?
2. List the basic design issues to machine learning.
3. State version space representation theorem.
4. How to use entropy as evaluation function?
5. What are ensemble learning methods in machine learning?
6. Overfitting vs Underfitting (Bias vs variance tradeoff).
7. Write functions of supervised and unsupervised learning.
8. Differentiate classification and regression.
9. Define Bayes Rule.Why is "Naive" Bayes naïve?
10. What is support vectors, margin and hyperplane in SVM?
11. Compare hierarchical and probabilistic clustering.
12. What is dendograms? Explain its use.
13. What are the applications of machine learning?
14. Explain Conditional Probability, Joint Probability, Probability Density Function.
15. Write a note on Sigmoid, Tanh and ReLU Neurons. or What is activation function? State and explain any two activation functions.
16. Draw the preceptron network with the notation. Derive an equation of gradient descent rule to minimize the error.
17. What are the components of reinforcement learning?
18. What are the different ways to evaluate the classification problem in machine learning?

Part B

(Analytical/ Problem Solving Questions)

19. Explain linear regression with suitable example
20.) What are the support vectors and margins? Explain soft SVM and hard SVM
21. What are various classification models?
22. Describe classification tree with suitable example
23. Explain the hierarchical method of clustering.
24. Explain two methods for reducing dimensionality.
25. Define and explain Minority Class, Gini Index, Entropy.
26. What is Deep Learning and how it is different from Machine Learning?
27. What is ensemble learning? Explain bagging and boosting.
28. Explain Linear Perceptron as Neurons.
29. Explain the role of machine learning algorithms in following applications.
30. i) Spam filtering
31. ii) Natural Language Processing
32. What is neural network? Differentiate single layer perceptron with multilayer perceptron?
33. Explain Recurrent Networks.
34. Explain with suitable example K-medoids algorithm.
35. Explain steps of Apriori algorithm.
36. Describe Gaussian Mixtures.

37. What is cluster analysis? What are the requirements of clustering?
38. Write short note on :
39. Markov Random Fields
40. Hidden Markov Model
41. What is HMM? Explain HMM with suitable diagram
42. What do you mean by Hierarchical clustering methods? What are the advantages of Agglomerative Hierarchical clustering technique?
43. What is Machine Learning? Describe Unsupervised Learning. Name two algorithms of it.
44. What is Likelihood? Describe the concept of Maximum Likelihood Estimation (MLE) for linear regression.
45. What is Overfitting and Underfitting in machine learning model? Explain with an example.
46. Differentiate with an example Label encoder and One Hot encoder for managing categorical data
47. Explain Elbow method for finding optimal number of clusters.
48. Compare Feature Extraction and Feature Selection techniques. Explain how dimensionality can be reduced using a subset selection procedure.
49. What is the role of PCA in machine learning? How PCA can be implemented on a 2D Dataset.
50. Explain the differences between Collaborative filtering and Content based filtering.
1.

Part C

(Descriptive/ Analytical/ Problem Solving/ Design Questions)

51. Consider following 8 points P1 = [0.1, 0.6], P2 = [0.15, 0.71], P3 = [0.08, 0.9], P4 = [0.16, 0.85], P5 = [0.2, 0.3], P6 = [0.25, 0.5], P7 = [0.24, 0.1], P8 = [0.3, 0.2]. Apply K-Means clustering with initial centroids m1 & m2 where m1 = P1, m2 = P8 and clusters are C1 & C2. Which cluster point P6 belongs to?
52. Consider following splits having four features:
    a. Length = [3, 4, 5] [2+, 0–] [1+, 3–] [2+, 2 –]
    b. Gills = [Yes, No], [0+, 4 –] [5+, 1 –]
    c. Beak = [Yes, No], [5+, 3 –] [0+, 2 –]
    d. Teeth = [many, few] [3+,4 –] [2+, 1–]

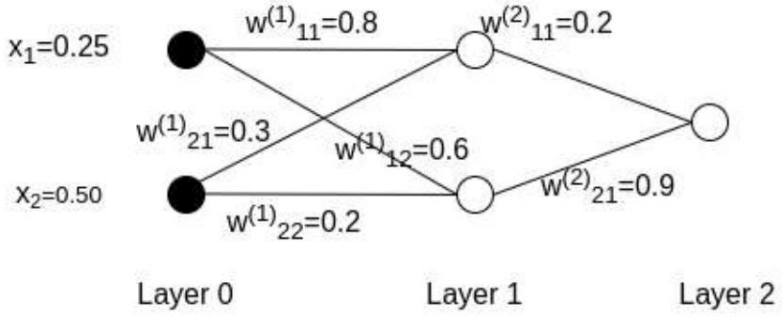
    Find Total weighted Entropy & Gini-index of all Features.
53. Draw and explain multi-layer perception in detail
54. Explain following measures used in association rule mining. Also give example of each. [8] i) Support ii) Confidence
55. What is TF and IDF? How document is represented with TF-IDF explain with an example
56. Use Iterative Dichotomiser 3 algorithm to formulate decision tree for the dataset shared below. The target variable is to take decision whether to play foot ball or not.

Outlook	Temperature	Humidity	Wind	Played football (yes/no)
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

57. The following table shows the results of a recently conducted study on the correlation of the number of hours spent driving with the risk of developing acute back-ache. Find the equation of the best fit line for this data

No. of hours spent driving (x)	10	9	2	15	10	16	11	16
Risk Score on a scale of 0-100	95	80	10	50	45	98	38	93

58. What is temporal difference Learning. Explain Sarsa and Q-Learning?
59. What is Bellman equation? How Markov Decision Method work in Reinforcement Learning? Also explain how does Q-Table gets updated in Q-Learning.
60. Consider the following Neural Network diagram and find the updated parameters of the network using one iteration. Sigmoid activation function is used at hidden and output layer. Consider only the parameters given in the network. (Use $\eta = 0.1$ and target output = 0.7).



****

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

Unit-1

Surpervised learning Algorithm: - Introduction, types of learning, application, Supervised learning: Linear Regression Model, Naive Bayes classifier, Decision tree, k nearest neighbor, Logistic Regression, Support Vector Machine, Random forest algorithm.

Unit 1 :-

Machine learning is a branch of AI & Computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

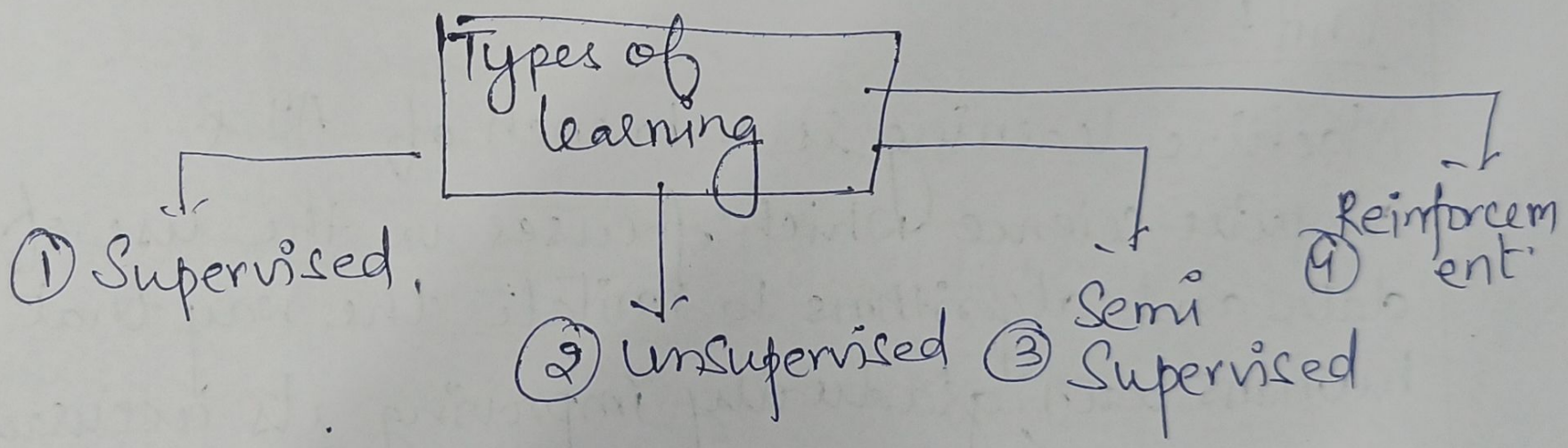
Machine learning (ML) is a subdomain of AI that focus on developing system that learn - ~~or~~ improve performance -based on the data they ingest.

All ML is AI
Not All AI is ML

A machine is said to be learning from past Experience with respect to some class of tasks if its performance in a given task improves with the experience.

Ex:- m/c has to predict whether a customer will buy a "Anti Virus" this year or not.
m/c will look at previous experience of that person, if he buys an Antivirus every year, then there is a high probability that he will buy.

- By using an algorithm or model in m/c we can predict the data.

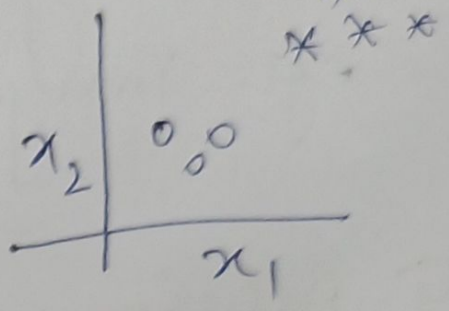


Supervised learning:-

→ Model or algo presented with inputs and output. It find pattern and conn" b/w the input & op.
→ Goal is to learn a general rule that train the model continue until desired level of accuracy. achieves.

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

eg:- Image classification
- Market Prediction/Regression.



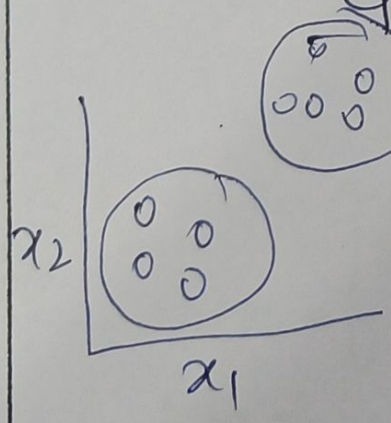
(2) Unsupervised learning :→

- No labels are given to the learning algorithm.
- It is used for clustering population and learn by itself-

e.g:- Clustering = Similar data in same cluster

High-Dimension Visualization →

Generative Models:→ Used to create data based on learning

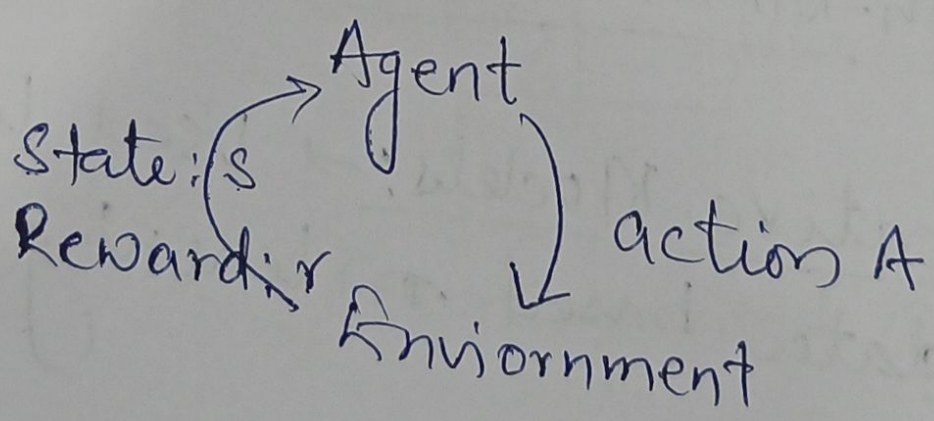


③ Semi-Supervised learning :-

- Have huge amt of data but some of the data is labeled.
- Between supervised and unsupervised learning.
- eg: images given few with label

④ Reinforcement learning :-

- Computer program interact with dynamic environment and must perform certain goal.
- Program is provided feedback in terms of reward and punishment



Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

(5)

ML	Traditional	AI
① Subset of AI, focus on learning from data to develop an algo that can predict	① Rule based code by developers depend on prob. statement	① AI involve making m/c intelligent like humanbeing.
② ML use historical data for training and make prediction on new data	② Self learning features	② Include many diff techniques ML & DL.
③ find pattern, diff for human	③ Depend on developer Intelligence	③ find good accuracy which seem possible to humans.

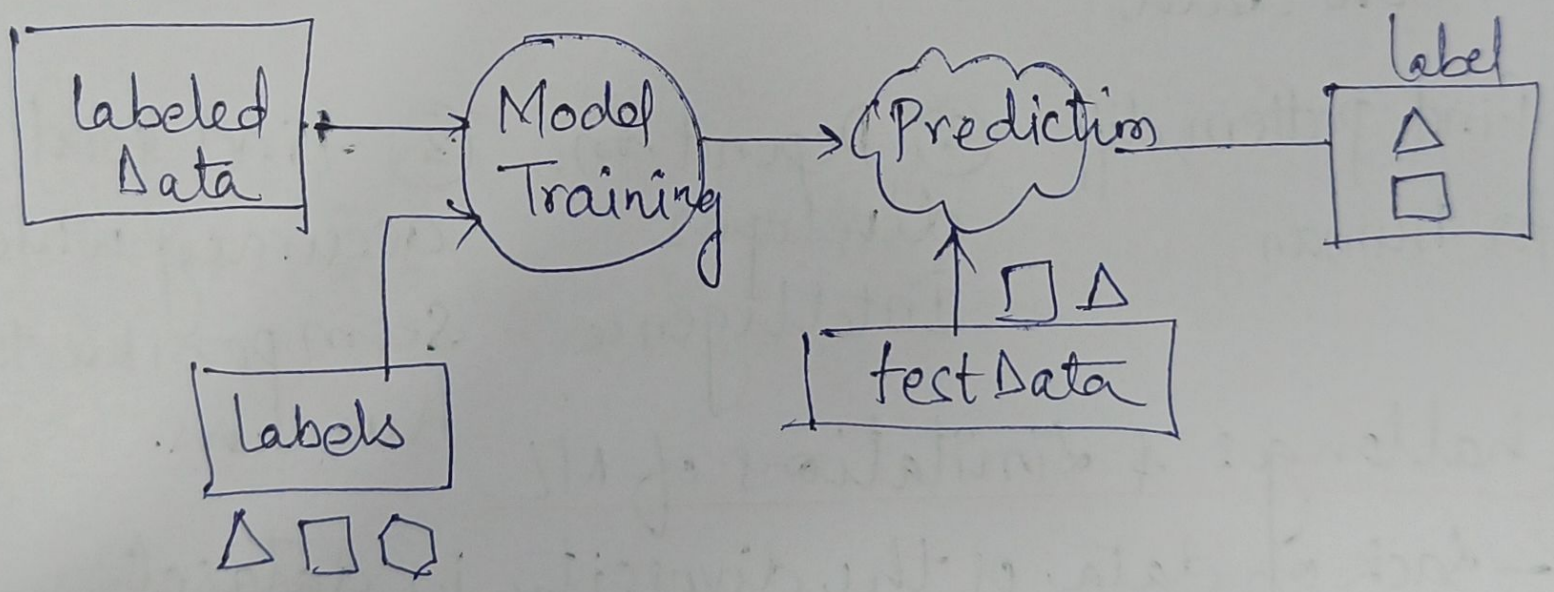
Challenges & Limitations of ML

④
- Lack of data or the diversity in dataset
- Machine can't learn if data sets are not available
-

Supervised Machine Learning:→

- Supervised ~~m/c~~ learning is the types of m/c learning in which machines are trained using well "labelled" training data, and on basis of that data, m/c predict the o/p.
- Supervised learning is a process of providing input data as well as correct output Data to the machine learning model. The aim of a supervised learning algorithm is to find a mapping func to map the input variable (x) with the output variable (y).

eg: Risk assesment, image classification, frad detection, spam filtering etc.



It's of two type

1. Regression
2. Classification

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Regression :→

- It is used for the prediction of continuous variable such as Weather forcasting, Market trend etc.
- Regression algo are used if there is a relationship b/w the input variable & o/p variable.
  - 1) Linear Regression
  2) Regression Trees
  3) Non-Liear Regression
  4) Bayesian Liear Regression
  5) Polynomial Regression.

1) Linear Regression
2) Logistic Regression
3) Decision Tree Regression

Classification: -

Classification algos are used when the o/p variable is categorical, which means there are two classes such as Yes-No, Male-female, True-false etc.

Random forest
Decision Tree
Logistic Regression
Support Vector Machine

Advantages of Supervised learning :-

- labeled training data benefits to models for better accuracy
- Can accurately predict and classify new data
- It include wide range of applications.
- Performace measured by accuracy, precision, recall and f1-score.

Disadvantages of supervised learning

- Not suitable for handling the complex tasks.
- Can't predict the correct o/p if test data is different from the training data.
- Training required lots of computation times.
- We need enough knowledge about the classes of object {feature Extraction}
- Overfitting cause poor performance

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Terminologies of Machine learning:-

1) **Model** – It's a specific representation learned from data by applying some machine learning algorithm. It is also known as hypothesis.

2) **Feature**:- Measurable property of data. A set of numeric features can be conventional described by a feature vector. feature vectors are fed as input to the model.

eg: ~~Choose~~ Predict fruit

features: color, smell, taste etc.

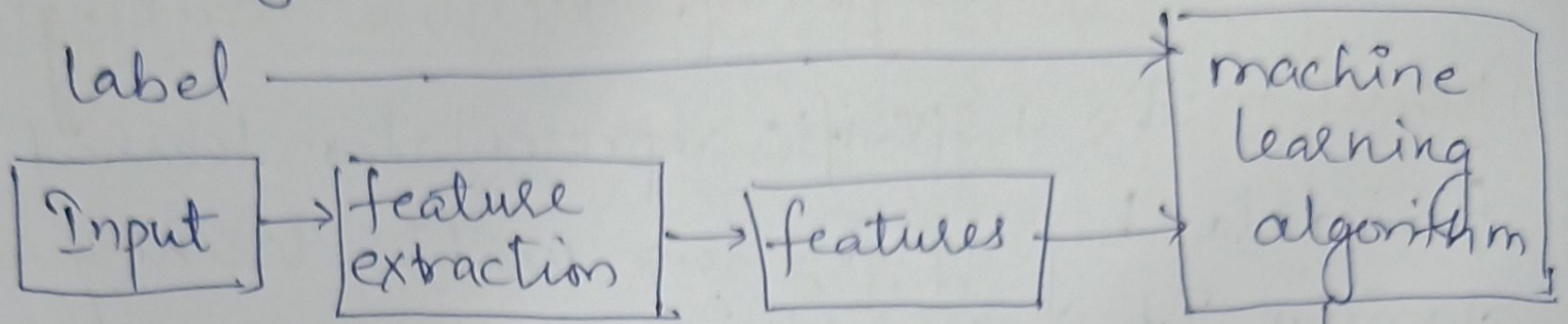
- feature extraction is used to extract the revelant features from raw data.

3) **Target (label)**:- Value to be predicted by model.
. As in above ex. find the name of fruit like apple, mango, banana.

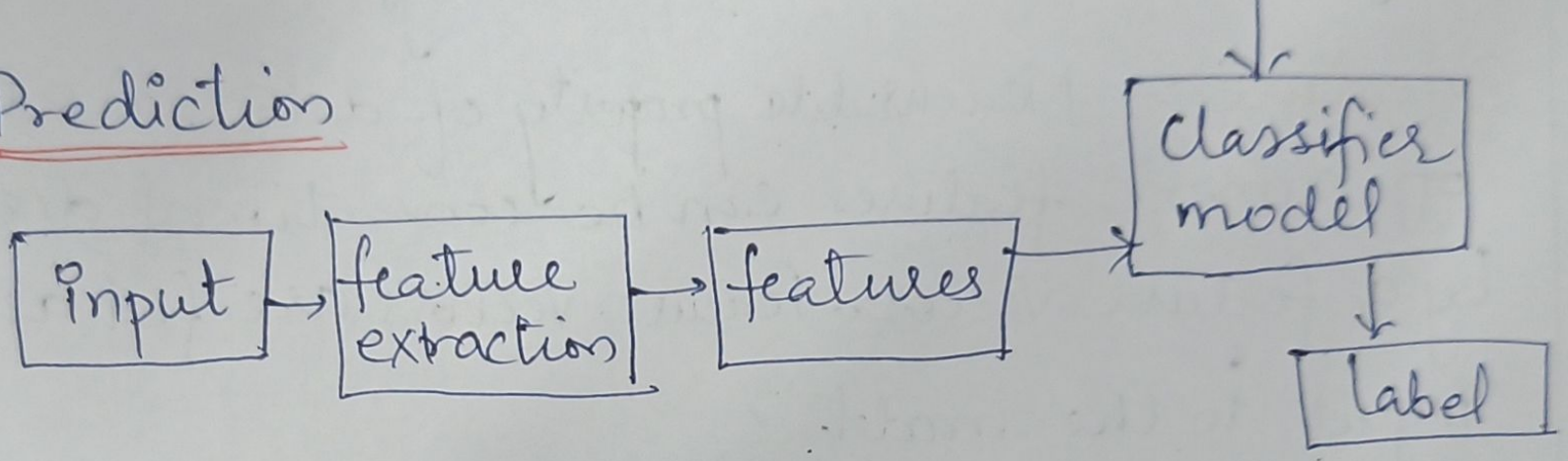
4) **Training**:- Give a set of inputs (features) and expected output (label), so after training we will have a model (hypothesis) that will then map new data to one of the categories trained on.

5 **Prediction**:- After training our model is ready. Now it can be fed a set of i/p and provide predicted o/p. If it performing well on non-training data means our model is well trained.

a) **Training**



b) **Prediction**



Steps in Machine learning:-

1. Define the problem.
2. Collect data
3. Explore the data (Use data visualization and statistical methods to understand the structure & relationship with your data)
4. Pre-process the data (normalizing, transforming, cleaning)
5. Split data (Training & Test)

, Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

6. Choose model
7. Train the model
8. Evaluate the model (performance of model)
9. fine tune the model (According to evaliation process until the desired level of accuracy is achived.)

10. Deploy the Model (Integrate the model into your application or system, making it available for use by others.)

11. Monitor the Model:-
(Continuously monitor the performance of the model to ensure that it continues to provide accurate result over time)

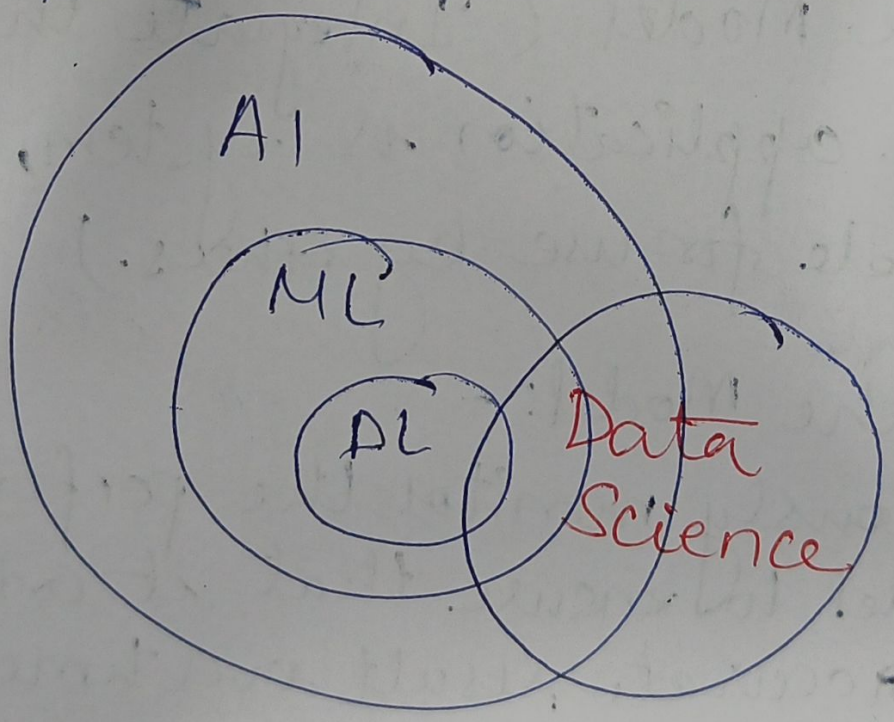
ML:- field of study that gives computers the capability to learn without being explicitly programmed.

Machine learning applications

1) Image & speech recognition to NLP.
2) Recommendation Systems
3) fraud detection
4) Portfolio optimization
5) Automated task
6) Drones & robots

Netflix:- Use combination of collaborative filtering and content-based filtering to recommend movies and TV shows to users - based on history, rating

- Reinforcement learning is another type of ML used to improve recommendation-based systems.



Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

Evolution of machine learning → 1950.

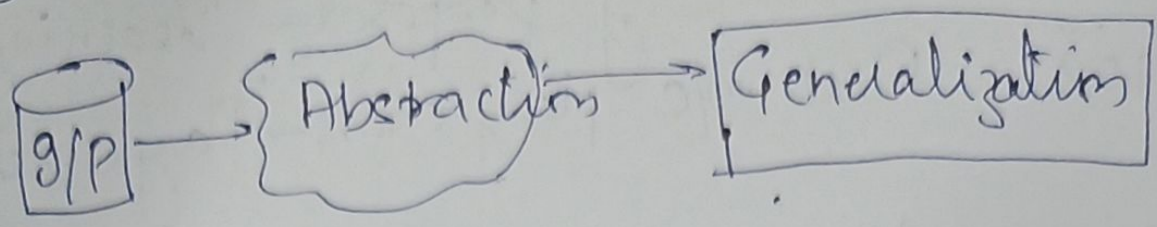
1. 1950 — Alan Turing proposes — "learning m/c"
2. 1952 → first m/c play chess → "Play checkers" by Arthur Samuel.
3. 1957 → first Neural N/w prog.
4. 1979 → Self driving cart {in room} by Stanford Univ.
5. 1982 → RNN
6. 1989 - Reinforcement learning.
7. 1995 — Random forest & SVM
8. 1997 - IBM Deep Blue beats
9. 2006 → Netflix {first ML learning competitions} Deeplearning
10. 2010 → Kaggle → ML Competitions
11. 2011 → IBM Watson beats → 2 human Champion.
12. 2016 → Google AlphaGo beats → human players.

Application of ML: →

1. Banking & finance
2. Insurance
3. Healthcare

How do machine learn

1. Data Input {Past data}
2. Abstraction {Data in broader way}
3. Generalization {make framwork for making decisio



Basic Types of Data in Machine learning:-

Dataset is a collection of related information or record.

attributes

eg:	Roll-no	Name	Gender	Age
	1	Geeta	F	14 → Record (Row)
	2	Pallavi	F	15
	3	Mihir	M	14
	4	Rahul	M	14

Data → Qualitative data / Categorical data
→ Provide Info abt the quality of object
eg: good, poor, avg

→ Quantitative data

eg: Student name, Rollno, can't be measured, so it is qualitative data. It is Two type

① Nominal data → no numeric value, but a named value.
eg: Blood group → A, B, AB, O
Nationality, Indian, American
Gender, Male, female other

② Ordinal data
Assign named values to attributes but they can be arranged in seq of increasing & decreasing

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

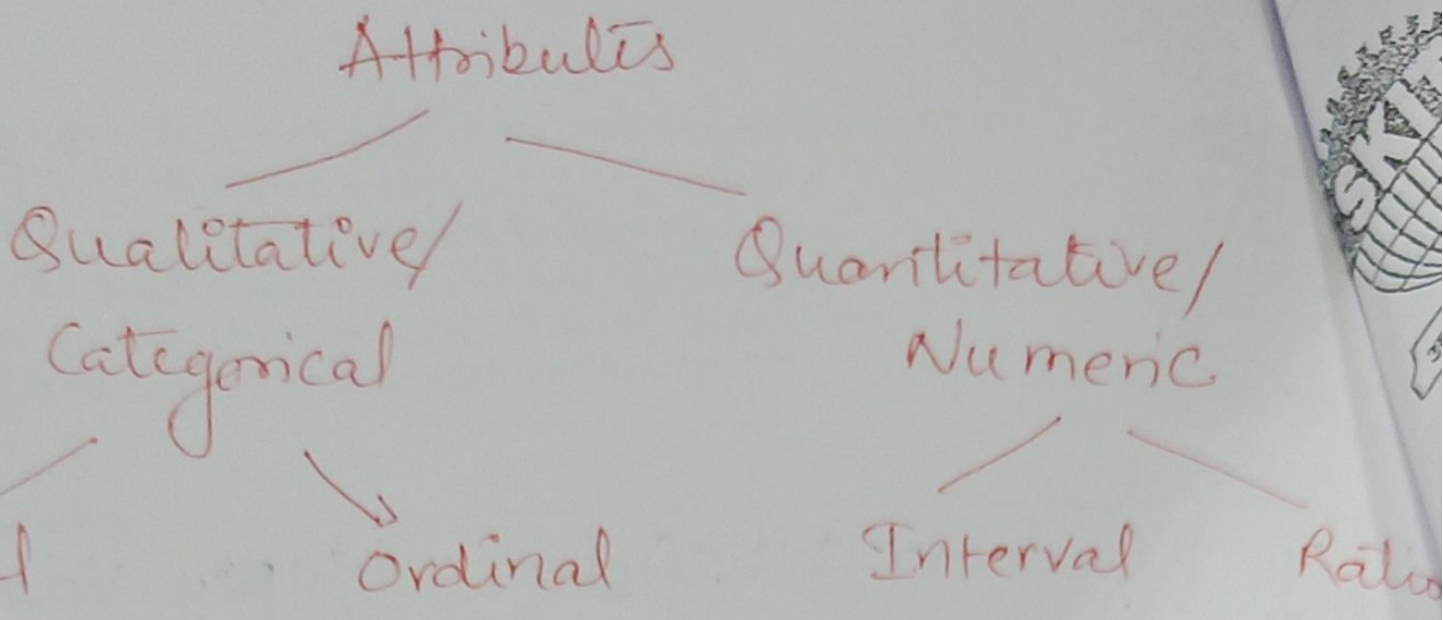
eg: 1. Customer satisfaction Happy, Very Happy, Unhappy.
2. Grades: A, B, C, D
3. Hardness of metal : Very Hard
Hard
Soft.

On qualitative data we can't apply mathematics like mean, variance, add, sub but basic count is possible. {Median is possible in Ordinal data}

Quantitative Data:—

Quantity of object: It can be measured.
eg: marks:

Interval	Ratio
– Order is known but the exact diff b/w values is also known Date, time, temp → Mathematical operation possible. → Zero is not there.	· Numeric data for which exact value can be measured. → Zero is there → height, weight, age, salary



→ Attributes can also be categorized into discrete and continous based.

→ Discrete attributes → finite/countable
name, gender, rank of students

→ Continous attribute → real no/ length, height, weight etc.

Exploring numerical data —

Mean :- sum of data values divided by count of data elements.

-21, 89, 34, 67, 96 → 61.4

Median :- value in the middle of an orderded list

-21, 34, [67] 89, 96

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Govemment of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

Measuring data dispersion:-

Attribute 1	44	46	48	45	47	→ mean? → median?
Attribute 2	34	46	59	39	52	↓ 46

Mean/Median → 46

Attribute 1 values are more close to it but in attribute 2 it is more spread.

→ find spread out / Dispersion

↓,

Variance is used.

$$\text{Variance}(x) = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2 = \frac{\sum(x-\bar{x})}{n}$$

$$\text{Stand dev}(x) = \sqrt{\text{Variance}(x)}$$

- Large value of Variance / Stad dev show more dispersion.

Variance (Attribute 1) = 2

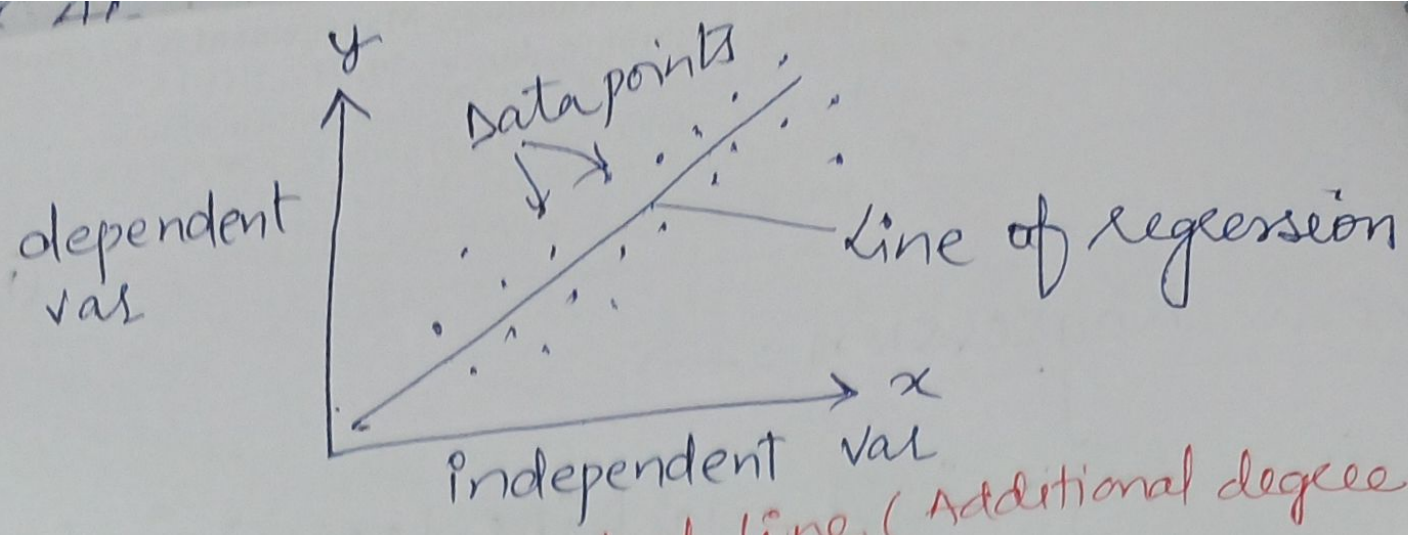
Variane (Attribute 2) = 79.6

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Linear Regression :→

- It is a supervised machine learning algo. that learns from the labelled datasets and maps the data points to the most optimized linear functions.
- Which can be used for prediction on new datasets.
- Linear regression is one of the easiest and most popular m/c learning algo.
- It's a statistical method that is used for predictive analysis.
- It makes predictions for continuous/real or numeric variable such as
  - Sales, salary, age, product price.
- It shows a linear relationship between a dependent (y) and one or more independent (x) variable, hence called as linear regression.
- It provides a sloped straight line representing the relationship b/w the variables.

eg: Rain → Independent
Crop → dependent



intercept of line (Additional degree of freedom)
Linear regression coefficient

$$Y = a_0 + a_1 x + \varepsilon$$

ε → random error

Dependent var

independent

Types of Linear Regression

- Simple

-Single independent variable is used

$$y = \beta_0 + \beta_1 x$$

- Multiple

more independent variables

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_p$$

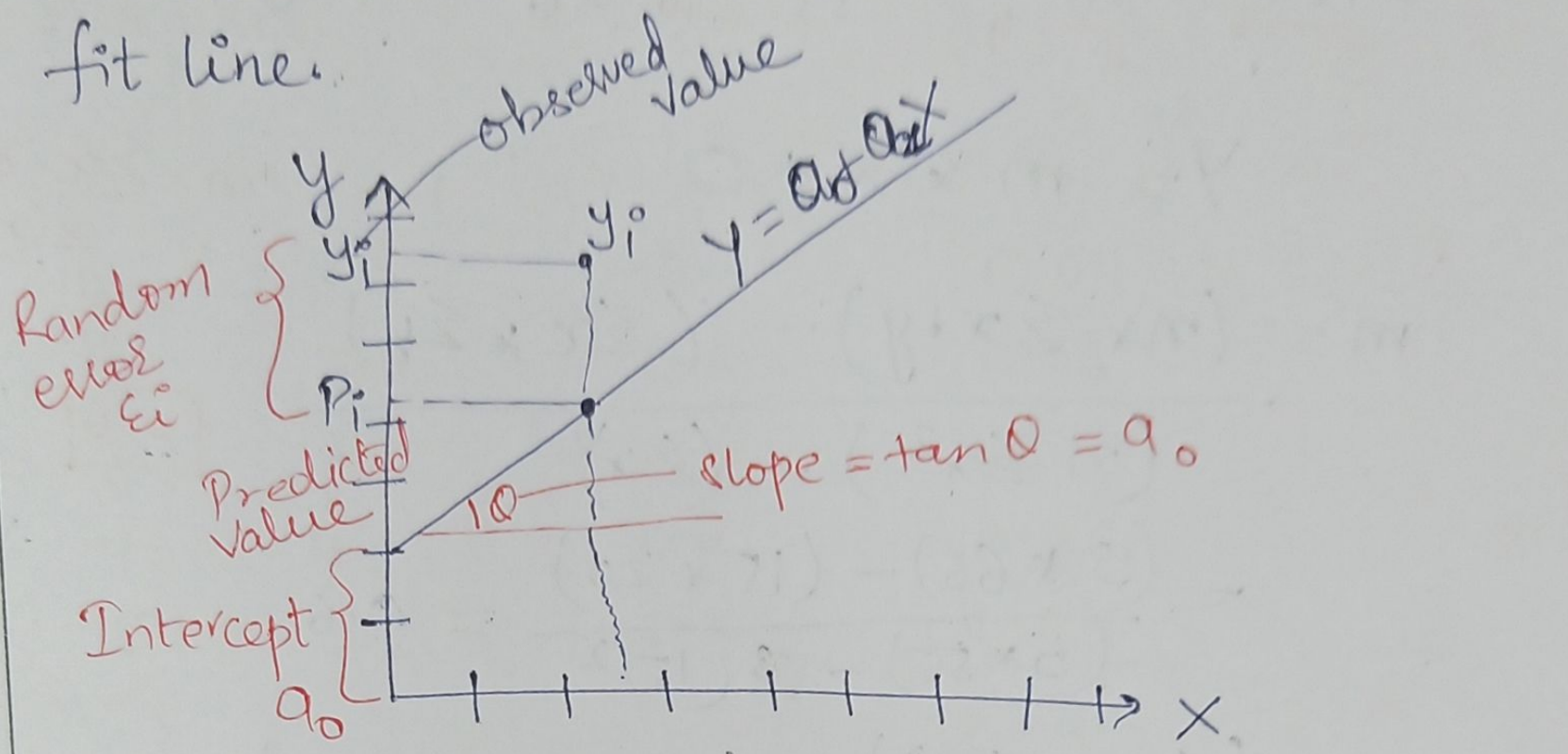
β_0 → intercept.

$\beta_1, \beta_2 \cdots \beta_n$ → slopes

-The goal of the algorithm is to find the best fit line equation that can predict the values based on the independent variables.

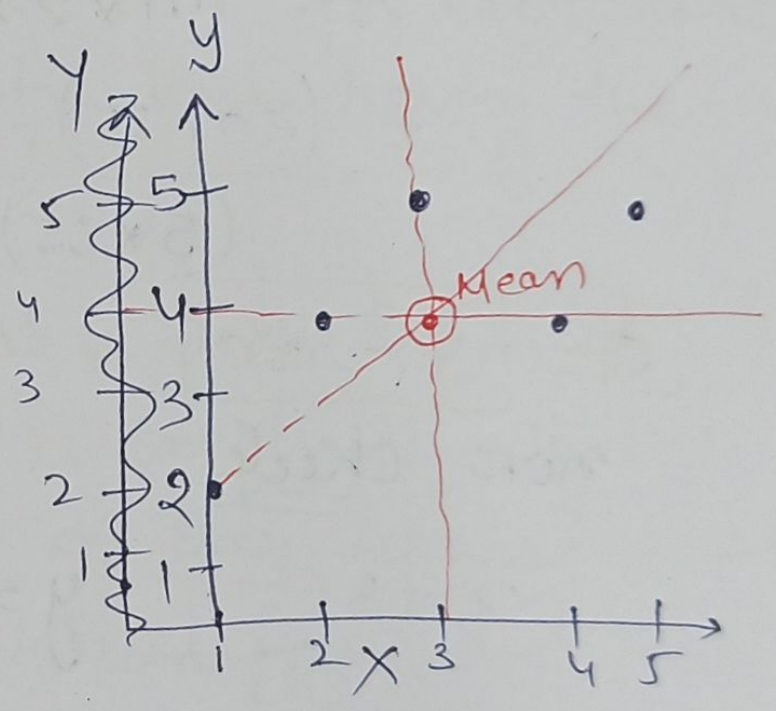
असतो मा सद्गमय
Approved by AICTE, [illegible]
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

- Best-fit line implies that the error b/w the predicted and actual values should be kept to a minimum.
- There will be the least error in the best-fit line.



Let we have

x	Y
1	2
2	4
3	5
4	4
5	5



mean of $x = \frac{5+4+3+2+1}{5} = 3$

$y = \frac{2+4+5+4+5}{5} = 4$

eg: calculated for calculated values m & C

x	y	$x-\bar{x}$	x^2	$(x-\bar{x})^2$	y^2	$y-\bar{y}$	$x\cdot y$	$(x-\bar{x})(y-\bar{y})$
1	2.8	2	-2	1	4	-2	2	4
2	3.4	4	-1	4	1	0	8	0
3	4	5	0	9	0	25	1	15
4	4.6	4	1	16	1	16	0	16
5	5.25	2	2	25	4	25	1	25
Σ	$=\frac{15}{5}=3$	$\frac{20}{5}=4$	$\Sigma 0$	55	Σ to	86	$\Sigma 0$	66

$\Sigma = 6$

$$Y = m * X + C$$

$$m = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

$$m = \frac{(n * \Sigma x\cdot y) - (\Sigma x * \Sigma y)}{(n\cdot \Sigma x^2) - (\Sigma x)^2}$$

$$= \frac{(5 * 66) - (15 * 20)}{(5 * 55) - (15)^2} = 0.6$$

$= \frac{6}{10}$

$= .6$

$25 \times 15 = 75$, $15 \times 15 = 225$

$$C = \frac{(\Sigma y * \Sigma x^2) - (\Sigma x * \Sigma x\cdot y)}{(n * \Sigma x^2) - (\Sigma x)^2}$$

$$= \frac{(20.55) - (15 \times 66)}{(5 * 55) - (15)^2} = \frac{1100 - 990}{275 - 225}$$

$$= \frac{110}{50} = 2.2$$

$\boxed{C} = \boxed{\bar{y} - m\bar{x}} \Rightarrow 4 - .6 * 3$

now check $= 4 - 1.8 = \boxed{2.2}$

$x = 3$ $\quad y = 0.6 \times 3 + 2.2$

$= 4$

Q:-

x	15	23	18	23	24	22	22	19	19	16	24	11	24
y	49	63	58	60	58	61	60	63	60	52	62	30	59

16	23
49	68

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

(3)

- The distance between the actual and predicted values are known as residual or errors.
- The best fit line should have the least sum of squares of these errors also known as e square.

Y	Y_{pred}	$(Y - Y_{pred})$	$(Y - Y_{pred})^2$
2	2.8	-0.8	0.64
4	3.4	0.6	0.36
5	4	1	1
4	4.6	-0.6	0.36
5	5.2	-0.2	0.04
		$\Sigma = 2.4$	

$\bar{Y} - \bar{Y}$ $2 - 4$ $0 - 0$ $1 - 1$ $0 - 0$ $1 - 1$

SST

Σ (5)

- SSE Sum of Squared Error

- The sum of squared errors for this regression line is 2.4. We check this error for each line and conclude the best fit line having the least e square value.

- In Linear Regression the Mean Squared Error (MSE) cost fun^c is employed which calculates the average of the squared errors between the predicted values ypred and the actual value y_i.

$$MSE(J) = \frac{1}{n} \sum_{n}^{i} (y_{pred} - y_i)^2$$

SLS

Minimize the Distance:- There are lots of ways minimize the distance between the line and the points like squared error, Sum of Absolute errors, Root mean Square error

Multiple Linear Regression:-

- In a multiple regression model, two or more independent variables, i.e predictors are involved in the model.

eg:. Price of Property & Area of Property (in sq.m)
dependend predictor variable

However location, floor, no. of years since purchase, amenities etc. are also important predictor variable.

$$\text{Price}_{\text{property}} = f(\text{Areap}, \text{location}, \text{floor}, \text{Ageing}, \text{Amenities})$$

$$\hat{y} = a + b_1X_1 + b_2X_2 + b_3X_3 \ldots + b_nX_n$$

↓
Partial regression coefficient.

- Regression line can be valid only over a limited range of data, If line extended, it may only lead to wrong predictions.

...pura, Jaipur-302017, INDIA ...nagement & Gramothan,
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

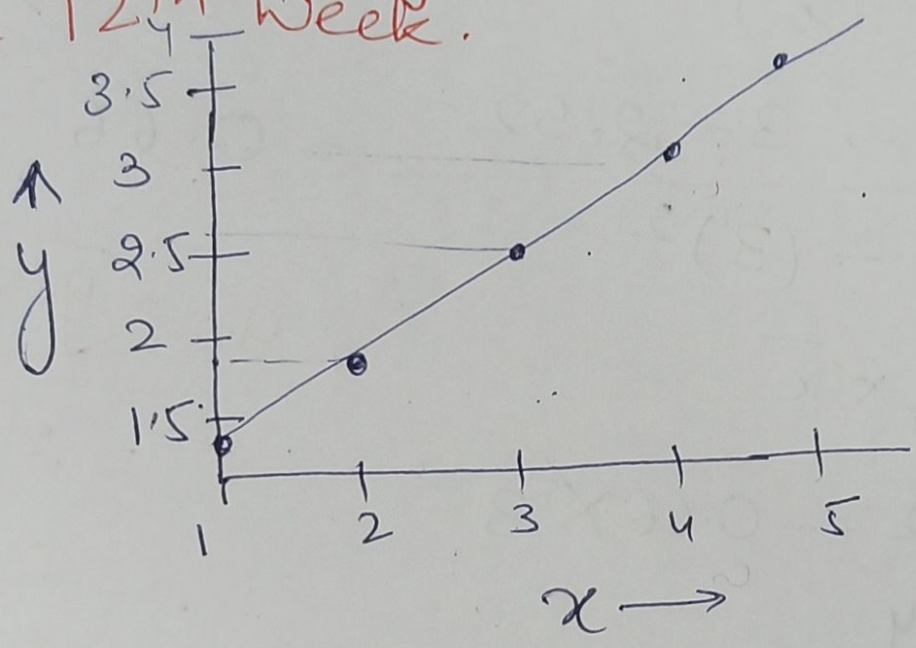
Linear Regression

eg: Q1

x (week)	yi (Sales in thousands)
1	1.2
2	1.8
3	2.6
4	3.2
5	3.8

Apply linear regression and predict sales in 7th & 12th week.

Ans:-



$$y = a_0 + a_1 \cdot x + e$$

$$a_1 = \frac{(\overline{xy}) - (\bar{x})(\bar{y})}{\overline{x^2} - \bar{x}^2}$$

$$a_0 = \bar{y} - a_1 * \bar{x}$$

Q2. Home price in NJ (USA). find price area = 3300 & 5000

Area	Price
2600	550000
3000	565000
3200	610000
3600	680000
4000	725000

$$a_1 = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

$$a_0 = \bar{y} - b\bar{x}$$

$n = 5$

x	y	x^2	$x_i * y_i$	$(x-\bar{x})^2$	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$
1	1.2	1	1.2	4	−2	−1.32	2.64
2	1.8	4	3.6	1	−1	−0.72	0.72
3	2.6	9	7.8	0	0	0.08	0
4	3.2	16	12.8	1	1	0.68	0.68
5	3.8	25	19	4	2	1.28	2.56
Sum 15	12.6	55	44.4	Σ10	Σ0	Σ0	Σ6.6

$\bar{x} = \frac{15}{5} = 3$ $\quad \bar{y} = \frac{12.6}{5} = 2.52$ $\quad \overline{x^2} = 11$ $\quad \overline{xy} = 8.88$

$$a_1 = \frac{\overline{xy} - (\bar{x})(\bar{y})}{\overline{x^2} - (\bar{x})^2}$$

$$= \frac{8.88 - 3 * 2.52}{11 - (3)^2} = 0.66$$

$$a_0 = \bar{y} - a_1 * \bar{x}$$

$$= 2.52 - 0.66 \times 3$$

$$= 0.54$$

Regression equation

$$y = a_0 + a_1 x$$

$$y = 0.54 + 0.66 \cdot x$$

x = 7th week $\quad y = 0.54 + 0.66 * 7 = 5.16$

= 12th week $\quad y = 0.54 + 0.66 \times 12 = 8.46$

$$a_1 = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (x-\bar{x})^2}$$

$$= \frac{6.6}{10} = 0.66$$

$a_0 = \bar{y} - a_1 \cdot \bar{x}$

$= 2.52 - 0.66 \times 3$

$= 0.54$

Let chele

$x = 1$

$y = a_0 + 1 \cdot a_1$

$= 0.54 + 1 \times 0.66$

$= \boxed{1.2}\ y$

actual = ?

SKIT
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Lineal Regression → Hike in Salary / Emp rating (1)

Logistic Regression :→ { It's used for categorical data } { for numeric Lineal Regression used }

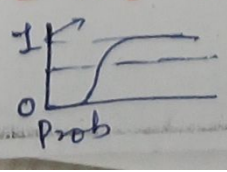
- Logistic regression is a supervised machine learning algorithm used for classification tasks
- where the goal is to predict the prob. that an instance belongs to a given class or not.
- Logistic regression is used for binary classification where we use sigmoid func that takes input as independent variable and produces a probability value between 0 & 1.

e.g. logistic regression predict whether a political candidate will win or lose an election (Win/Lose)

e.g. whether a high school student will be admitted or not to a particular college. (Yes/No)

- Logistic regression predicts the output of a categorical dependent variable.
- Instead of fitting a regression line, we fit an "S" shaped logistic fun".
- Sigmoid func is a mathematical fun used to map the predicted values to probabilities.

? Promotion should given or not based on their performance (prob based)



- It maps real value into another value within a range of 0 and 1.
- The S-form curve is called the Sigmoid func or the logistic func.
- In logistic regression, use the concept of the threshold value, which defines the prob of either 0 or 1.
- Such values the threshold value tends to 1, and a value below the threshold values tends to 0.

Types of Logistic Regression:→

1) Binomial:- In binomial logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or fail etc.
2) Multinomial:- There can be 3 or more possible unordered types of the dependent variable such as 'cat', 'dogs' or 'sheep'.
3) Ordinal:- There can be 3 or more possible ordered types of dependent variables, such as 'low', 'Medium', or 'High'.

Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Assumption of Logistic Regression:-

Logistic regression as understanding these assumptions is important to ensure that we are using appropriate application of the model.

1. Independent observations:- No correlation b/w any input variables.
2. Binary dependent variable:-
   Dependent variable must be binary or dichotomous. for more than two categories Softmax func are used.
3. Linearity relationship b/w independent variable and logs odds: Relationship b/w independent variable and dependent variable is linear.
4. NO Outlier:- No outlier in the dataset.
5. Large sample size:- The sample size is sufficiently large.

Terminologies Involved in Logistic Regression:

1) Independent variables:- Input characteristics / pred
2) Dependent variables:- Target variable / predict val
3) Logistic func:- formula b/n input & output value and tell the prob value b/w 0 & 1
4) Odds:- Ratio of something occuring to something not occuring
5) Log-odds:- It is a logit func, linear combination of the independent variable & the intercept.

$$y = \beta_0 + \beta_1 * x \quad -(2)$$

6) Cofficient:-
7) Intercept:-
8) Maximum likelihood estimation

→ $Odds(Q) = \frac{\text{Prob of an event happening}}{\text{Prob of an event not happening}}$

$$Odds(Q) = \frac{P}{1-P} \quad -(1)$$ take log both side

Now $$\log\left(\frac{P(x)}{1-P(x)}\right) = \beta_0 + \beta_1 * x$$

$$e^{\ln\left(\log \frac{P(x)}{1-P(x)}\right)} = e^{\beta_0 + \beta_1 \cdot x} \qquad e^{\log(x)} = x$$

$$\frac{P(x)}{1-P(x)} = e^{\beta_0 + \beta_1 x}$$

Sigmoid $$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

Lineal Regression	Logistic Regression
1. Solve for regression problem	1. Used to slove classification problem.
2. The response variable are continuous in nature	3. The response variable is categorical in nature
3. It helps estimate the dependent variable when there is a change in the independent variable	3. It helps to calculate the possibilty of a particular event taking place
4. It is a straight line	4. It is an S-curve (Sigmoid)

eg1:- The student dataset has entrance marks based on the historic data of those who are selected or not selected.

$\therefore \beta_0 = 1, \beta_1 = 8$

- Assume marks of $x = 60$, compute the resultant class.

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

$1 + 8 \times 60$
$= 481$

$$p(x) = \frac{1}{1 + e^{-481}} = 0.44$$

$0.44 < 0.5 \rightarrow$ Studend will not be selected.

Q:-

Study hours	Result
2	0
3	0
4	0
5	1
6	1
7	1
8	1

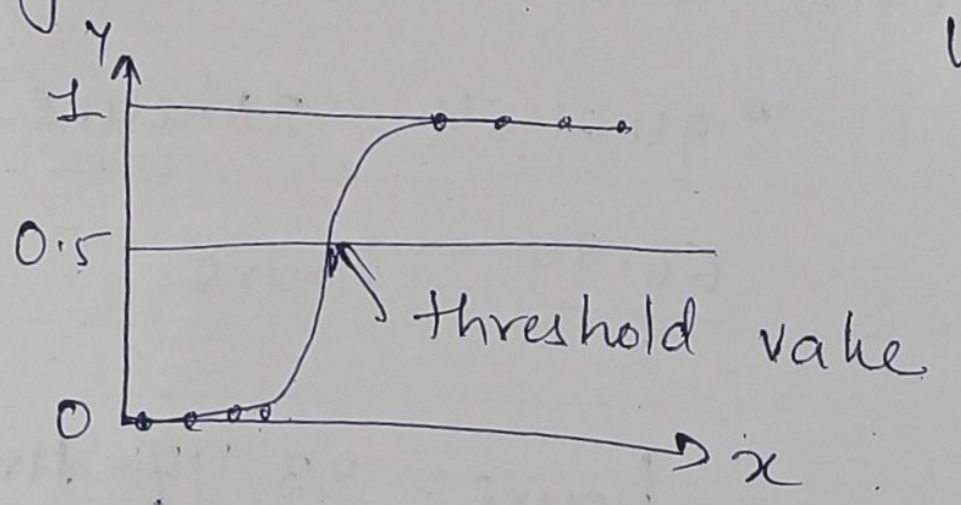
$a_0 = -1.5$
$a_1 = 0.6$
$x = 5$

eg:-

Hours Study	Pass (1) / Fail (0)
29	0
15	0
33	1
28	1
39	1

1. Calculate the prob. of pass for the student who studied 33 hr.
2. At least how many hours students should study that makes he will pass the course with the probability of more than 95%.

Ans:- $\log(\text{odds}) = -64 + 2 * hr = Z$



$$S(x) = \frac{1}{1+e^{-x}} \quad , \quad P = \frac{1}{1+e^{-Z}}$$

Ans1:-

$S(x) =$

$Z = -64 + 2 * 33 = -64 + 66 = 2$

$$P = \frac{1}{1+e^{-2}} = 0.88 \longrightarrow$$ There is 88% Chance that the student will pass the exam.

Ans2:-

$$P = \frac{1}{1+e^{-Z}} = 0.95$$

$$0.95 \times (1+e^{-Z}) = 1$$

$$.95 + .95e^{-Z} = 1$$

$$.95 \cdot e^{-Z} = 1 - .95$$

$$e^{-Z} = \frac{.05}{.95} = 0.0526$$

log both side

$$\ln(e^{-Z}) = \ln(0.0526)$$

$$+Z = +2.94$$

$$Z = 2.94$$

$$\log(\text{odds}) = Z = -64 + 2 * \text{hours}$$

$$2.94 + 64 = 2 * \text{hours}$$

$$\frac{66.94}{2} = \text{hours}$$

$$\text{hours} = 33.47 \text{ Hr}$$

Student study at least 33.47 hourse, he will pass the exam

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

Naive Bayes Classifiers :-

- Naive Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.
- It mainly used in text classification that includes a high-dimensional training dataset.
- It is simple and most effective classification algo which help in building the fast m/c learning models that can make quick predictions.
- It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.
- Naive Bayes algo — example → Spam filtration
  Sentimental Analysis
  classifying articles.

Naive :- means occurance of a certain feature is independent of the occurrence of other feature. Eg shape, color taste → fruit
independent occurance
red, spherical, sweet → apple.

Bayes :- Bayes' Theorem.

→ It's called naive because it makes the assumption that all attributes are independent of each other.

Bayes' Theorem :→

Bayes' Rules / Bayes' law → determine the probability of a hypothesis which prior knowledge.
It depends on the conditional probability.

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Where → Probability of A occuring given that B has already occured.

$P(A|B)$ is a posterior prob:- Prob of hypothesis A on the observed event B

$P(B|A)$ is a likehood prob.- Prob of evidence given that the prob of a hypothesis is true.

$P(A)$ is Prior Prob.-

$P(B)$ - Marginal Prob.

→ It works on conditional Probability

eg: Total 52 cards
→ Choose king of heart
condition
total = 13
king = 1
$P(k/h) = \frac{1}{13}$

Working of Naive Bayes Classifier:-

- Condition is that ur able to play or not according to the weather conditions.
- So following steps will be followed.

1. Convert the dataset into freq tables
2. Generate likelihood table by finding the probabilities of given features
3. Use Bayes theorem to calculate the posterior prob.

Choose king of heart

$P(k/heart)$ → Probability of king given that A/B

B is true

$P(heart/k)$ → Prob of heart given that king is true

$P(A)$ = Prob of event A

$P(B)$ = Prob of event B

$P(H/k) = 1/4$

$P(k) = 1/13$

$P(heart) = 1/4$

$$P(k/Heart) = \frac{P(H/k) \cdot P(k)}{P(S)} = \frac{\frac{1}{4} * \frac{1}{13}}{\frac{1}{4}} = \frac{1}{4 \times 13} * \frac{4}{1}$$

$$= 0.076$$

$\boxed{P(k/h) = \frac{1}{13} = 0.076}$ → Directly find.

	Outlook	Play
1	Rainy	Y
2	Sunny	Y
3	Overcast	Y
4	Overcast	Y
5	Sunny	N
6	Rainy	Y
7	Sunny	Y
8	Overcast	Y
9	Rainy	N
10	Sunny	N
11	Sunny	Y
12	Rainy	N
13	Overcast	Y
14	Overcast	Y

$P(Yes) = 10/14$

$P(No) = 4/14$

Outlook	Yes	No
Rainy	2/10	2/4
Sunny	3/10	2/4
Overcast	5/10	0/4

Apply Baye's theorem:

$$P(Yes/Sunny) = \frac{P(Sunny|Yes)\,P(Yes)}{P(Sunny)}$$

$$= \frac{\frac{3}{10} \times \frac{10}{14}}{5/14}$$

$$= \frac{3}{14} \times \frac{14}{5} = .6$$

$$P(No/Sunny) = \frac{P(Sunny/No)\,P(No)}{P(Sunny)}$$

↑ Sunny → Play.
P(Yes) > P(No)

$$= \frac{\frac{2}{4} \times \frac{4}{14}}{5/14} \Rightarrow \frac{2}{14} \times \frac{14}{5} = .4$$

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

Naive Bayes Classifier

Day	Outlook	Temp	Humidity	Wind	Play/Tennis
1	Sunny	Hot	High	Weak	No
2	S	H	H	Strong	No
3	Overcast	H	H	W	Yes
4	Rain	Mild	H	W	Yes
5	R	Cool	Normal	W	~~No~~ Yes
6	R	C	N	S	~~Yes~~ No
7	O	C	N	S	Yes
8	S	M	H	W	No
9	S	C	N	W	Yes
10	R	M	N	W	Yes
11	S	M	N	S	Yes
12	O	M	H	S	Yes
13	O	H	N	W	Yes
14	R	M	H	S	No

find Outlook = sunny, Temp ⇒ Cool, Play = ?
Humidity = high, Wind ⇒ Strong

- Calculate Prior Probability & Current Probability

Prob(Yes) = 9/14 = .64 → Prior
Prob(No) = 5/14 = .36

Here we have 4 attribute, find conditional prob of 4 attribute.

Current prob

Outlook	Y	N
Sunny	2/9	3/5
Rain	3/9	2/5
Overcast	4/9	0

Temp	Y	N
Hot	2/9	2/5
mild	4/9	2/5
Cool	3/9	1/5

Humidity	Y	N
high	3/9	4/5
normal	6/9	1/5

Windy	Y	N
Strong	3/9	3/5
Weak	6/9	2/5

? outlook = sunny
Temp = Cool
Humidity = high
Wind = Strong

$$V_{NB} = \underset{v_j \in \{yes, no\}}{\operatorname{argmax}} \; P(v_j) \prod P(a_j | v_j)$$

$P(\text{outlook} = \text{sunny} | v_j)$ $P(\text{temp} = \text{Cool} | v_j)$
$P(\text{Humidity} = \text{high} | v_j)$ $P(\text{Wind} = \text{Strong} | v_j)$

$V_{NB}(\text{Yes}) = P(\text{Yes})\, P(\text{Sunny}/\text{Yes})\, P(\text{Cool}/\text{Yes})\, P(\text{high}/\text{Yes})\, P(\text{Strong}/\text{Yes})$

$= .64 * \frac{2}{9} * \frac{3}{9} * \frac{3}{9} * \frac{3}{9}$

$= .0053$

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

$V_{nB}(No) = P(no)\ P(S/N)\ P(C/N)\ P(H/N)\ P(Strong/N)$

$= .36 * \frac{3}{5} * \frac{1}{5} * \frac{4}{5} * \frac{3}{5}$

$= .0206$ → Prob of $V_{nB}(No) >$ Prob $V_{nB}(Yes)$

So No will be the ans.

$V_{NB}(Yes) = \frac{V_{NB}(Yes)}{V_{nB}(Yes) + V_{nB}(No)} = 0.205$

$= \frac{.0053}{.0053 + .0206} = 0.795$

$V_{NB}(No) = \frac{V_{NB}(No)}{V_{nB}(Yes) + V_{nB}(No)}$

$= \frac{.0206}{.0206 + .0053}$

Advantages of Naive Bayes :-

- One of the fast & easy ML algo to predict a class of datasets.
- It can be used for binary as well as Multi-class classifications.
- It performs well in Multi-class predictions as compared to the other algorithms.
- Most popular choice for text classification problems.
- Effective in cases with a large no. of features

Disadvantages of Naive Bayes Classifier

- It assumes that all features are independent or unrelated, so it cannot learn the relationship between features.
- Can be influenced by irrelevant attributes.
- Assumes that features are independent, which may not always hold in real-world data.
- May assign zero prob to unseen events, leading to poor generalizations.

Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

Naive Bayes :-

Example 2

No	Color	legs	Height	Smelly	Species
1	White	3	Short	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	Yes	H

- Using these probability estimate the prob value for new instance

(Color = Green, leg = 2, Height = Tall, Smell = No)

Ans:- $P(M) = \frac{4}{8} = 0.5$ $P(H) = \frac{4}{8} = 0.5$

Color	M	H
White	2/4	3/4
Green	2/4	1/4

legs	M	H
3	3/4	0/4
2	1/4	4/4

Height	M	H
Short	3/4	2/4
Tall	1/4	2/4

Smelly	M	H
Yes	3/4	1/4
No	1/4	3/4

$$P(M/\text{instance}) = P(M) \cdot P(\text{Color}=\text{Green}/M) * P(\text{legs}=2/M) * P(H=T/M) * P(\text{Smell}=no/M$$

$$= 0.5 * \frac{2}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = 0.0117$$

$$P(H/\text{instance}) = P(M) \cdot P(\text{Color}=G/M) * P(\text{legs}=2/M) * P(H=T/M) * P(\text{Smell}=no/H)$$

$$= 0.5 * \frac{1}{4} \cdot \frac{4}{4} \cdot \frac{2}{4} \cdot \frac{3}{4} = 0.047$$

$P(H/\text{New instance}) > P(M/\text{New instance})$

$0.047 > 0.0117$

Hence the new instance belongs to Species H

Ar

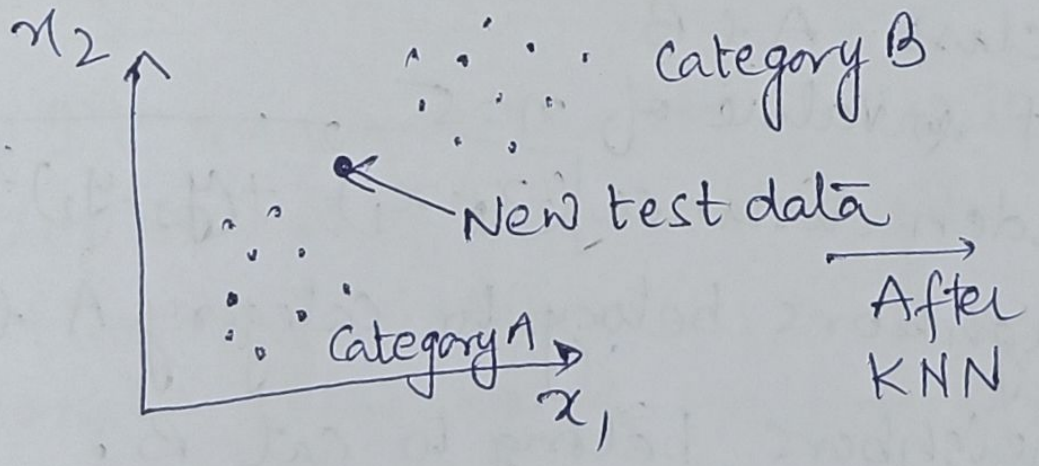
Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

kNN (k- Nearest Neighbor) Classifier :→

- It's a supervised ml method.
- One of the simplest ML algo.
- It assumes the similarity b/w new case/data & available cases and put the new case into the category that is most similar to the available categories.
- KNN algo can be used for regression as well as for classification but mostly used for classification.
- It don't make any assumption/non-parametric algo.
- It is called lazy learner algo because it does not learn from training set.

Eg:- Let we have cat & dog, for testing purpose we find the similarity b/w test data & cat, & dog.



It belong to category 1 because the diff is less b/w category 1 and test data

KNN Algorithm :→

Input :- Training & test data, value of k (i.e. nearest neighbors)

Steps: 1. Do for all test data points
- Calculate the distance (using Euclidean distance) of test data point from the different training data points.

2. Find the closest 'k' training data points, i.e training data points whose distance are least from the test data points.

If k=1
Then assign class label of training data point to test data point.

Else Which class level is more in the training data, assign that to test data point.

End do.

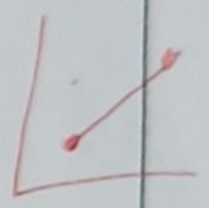
Eg:- We have 2 class A & B
New test point & value of $n=5$
- find the Eucliden distance $\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$
- 3 nearest neighbors belong to category A & 2 nearest neighbors belong to cat B.
Hence this new data point must belong to category A.

असतो मा सद्गमय
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Distance Metrices :-

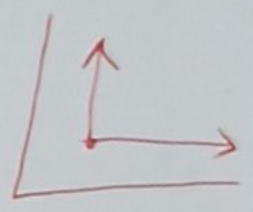
1. Euclidean Distance :- / displacement of shortest dist

$$\text{distance}(x, X_i) = \sqrt{\sum_{j=1}^{d} (x_j - X_{ij})^2}$$



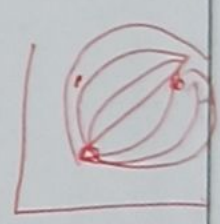
2. Manhattan Distance :-

$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$



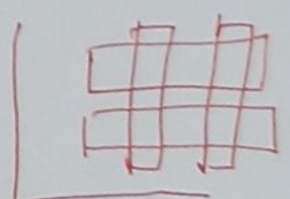
3. Minkowski Distance :-

$$d(x, y) = \left(\sum_{i=1}^{n} (x_i - y_i)^p\right)^{1/p}$$



if $p = 2$ → Euclidean distance
if $p = 1$ Manhattan distance

(4) Hamming Distance :-



- It measures the ~~ob~~ similarity b/w two strings of the same length.
- Hamming distance b/w 2 strings of the same length is the no. of positions at which the corresponding characters are different

1. $X(1,2,3)$ $Y(4,5,6)$

$$Dis(X,Y) = \sqrt{(4-1)^2 + (5-2)^2 + (6-3)^2}$$

$$= \sqrt{3^2 + 3^2 + 3^2}$$

$$= \sqrt{9+9+9} = \sqrt{27} = 5.196$$

2. $X(1,2,3)$, $Y(4,5,6)$

$$Dis(X,Y) = |4-1| + |5-2| + |6-3|$$

$$= 3 + 3 + 3 = 9$$

3. $X(1,2,3)$, $Y(4,5,6)$ p = order of norm

$$dis(X,Y) = \sqrt[3]{(1-4)^3 + (5-2)^3 + (6-3)^3}$$

$$= \sqrt[3]{3^3 + 3^3 + 3^3}$$

$$= \sqrt[3]{27+27+27}$$

$$= \sqrt[3]{81}$$

$$= 4.32$$

4. eg: Karolin / kathrin Distance = 3

1011101 / 1001001 Distance = 2

2173896 / 2233796 D = ?

Karolin / kerstin D = ?

(3)

How to select k value in KNN algo:-

1. No way to determine value of K, K=5 value best / or odd value of k.
2. At k=1 or k=2 noisy and lead to outliers in the model.
3. Large value of k are good, but it may find some difficulties.
4. Time time classes are given in input data

eg:-

Name	Aptitude	Commⁿ	Class	k=3
Ken	2	5	Speaker	
Ben	2	6	S	
Pen	3	5.5	S	
Jen	4	7	S	
Bobby	5	3	Intel 1.500	Intel
Rai	6	2	I	
Gen	~~6~~	4	I 1.18	Intel
Perl	~~8~~ 7	2.5	I	
Gov	~~8~~ 8	3	I	
Sus	~~8~~ 6	5.5	Leader 1.414	Leader
Bob	6	7	L	
Gorg	7	6	L	
Den	8	6	L	
Pap	9	7	L	
Josh	5	4.5	???	→ Intel.

Advantages of KNN:-

1. Easy to implement
2. It can be more effective if the training da is large. / Recommender system design.
3. Robust to the noisy training data
4. Very fast or almost no time required for the training phase.

Disadvantages of KNN:-

1. Does not learn anything in real sense. Classificatin is done completely on the basis of training data. This makes this algo both time-consuming & resour exhausting.
2. Curse of dimensionality:- Algo faces a hard time classifying the data points properly when the dimensionality is too high.

Application of KNN

1. Recommender System
2. Searching documents / contents similar to docume or It's a part of Information retrieval.

...pur-302017, INDIA
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Eg:-

	P_1	P_2	Class
i=1	7	7	false
2	7	4	false
3	3	4	True
4	1	4	True

? Perform kNN

X (P1 = 3, P_2 = 7)

k = 3

1) $D(x,1) = \sqrt{(3-7)^2 + (7-7)^2} = 4$ — N_3 → false

2) $D(x,2) = \sqrt{(3-7)^2 + (7-4)^2} = 5$

3) $D(x,3) = \sqrt{(3-3)^2 + (7-4)^2} = 3$ — N_1 - True

4) $D(x,4) = \sqrt{(3-1)^2 + (7-4)^2} = 3.6$ — N_2 - True

Two trues, One false

→ So It belongs to **True**

Eg:-

Maths	CS	Result
4	3	fail
6	7	Pass
7	8	Pass
5	5	fail
8	8	Pass

Q:- X (Maths = 6 , CS = 8 k = 3

① $\sqrt{(6-4)^2 + (8-3)^2} = \sqrt{29} = 5.38$ Class

② $\sqrt{(6-6) + (8-7)^2} = 1$ —— ① Pass

③ $\sqrt{(6-7)^2 + (8-8)^2} = 1$ —— ② Pass

④ $\sqrt{(6-5)^2 + (8-5)^2} = \sqrt{10} \doteq 3.16$

⑤ $\sqrt{(6-8)^2 + (8-8)^2} = \sqrt{2^2} = 2$ —— ③ Pass

So X → Pass

Q:-

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Decisin Tree:-

- Decision tree is a supervised learning technique that can be used for both classification and regression.

- It is a tree-structured classifier, where
  - → internal nodes represents features of dataset
  - → branches represent – decision rules
  - → each leaf - represents outcome.

-- A decision tree simply ask a question, based on Answer (Yes/No.) to split it into subtrees.

- There are specialized terms associated with DT

① Root nodes:-
Original feature from where tree branches.

② Internal nodes:-(Decision Nodes)
After visiting these node go to other nodes.

③ Leaf nodes(Terminal)
Decision nodes or answer.

④ Branches :-
Links b/w nodes

⑤ Splitting

⑥ Parent node

⑦ Child Node

⑧ Decision Criterion

⑨ Pruning :-

How decision Trees Work?

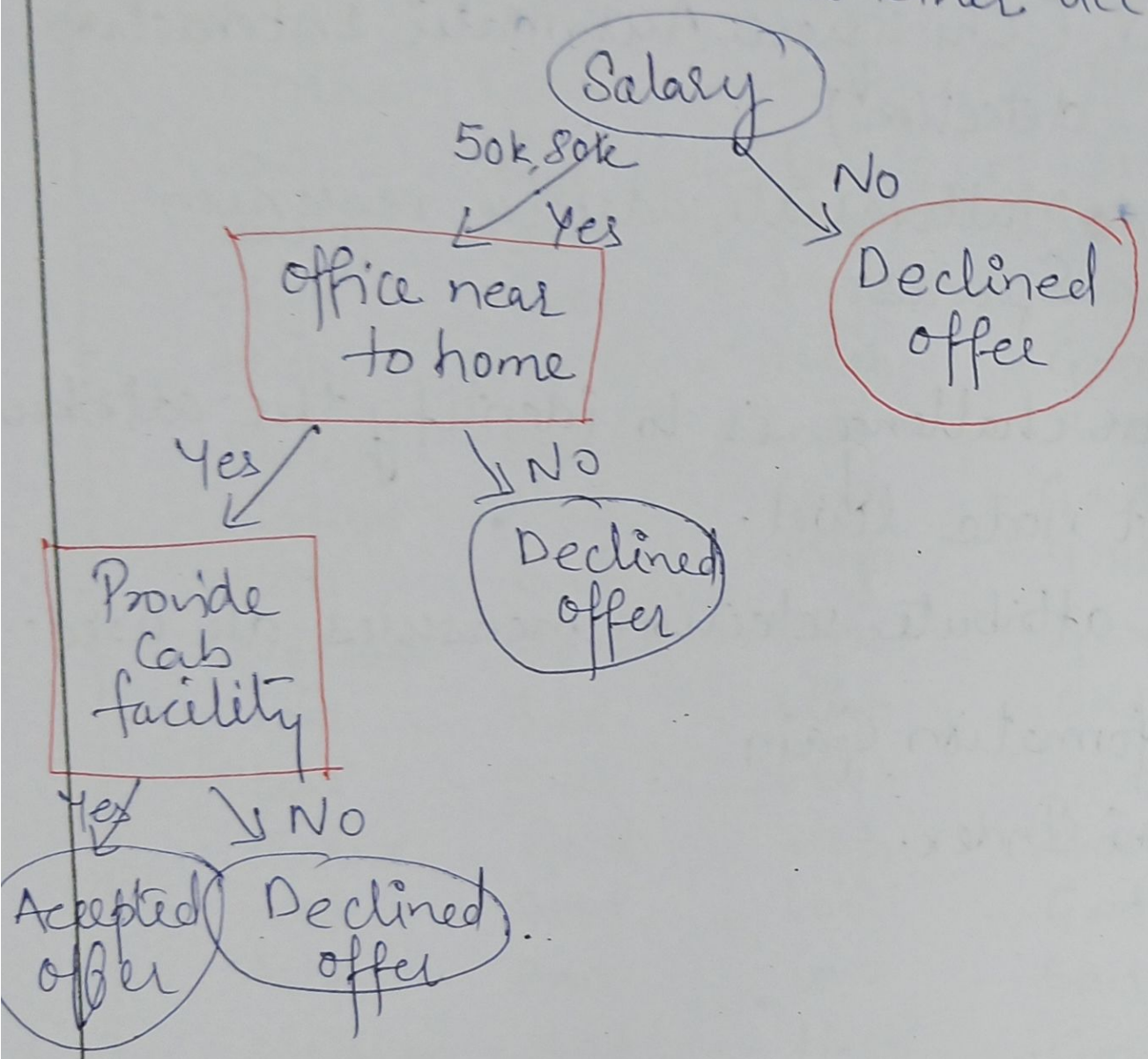
It use recursively partitioning the data based on the values of different attributes. It select best attribute to split the data at each internal node based on Info. Gain & Gini impurity.

Steps used in decision Tree:-

1. Begin the tree with the root node (S) which contain the complete dataset.
2. Find the best attribute using Attribute Selection Measure (ASM).
3. Divide the S into subsets that contains possible values for the best attributes.
4. Generate the decision tree node, which contain the best attribute.
5. Recursively make new DT using subset of dataset till not find final node as leaf node.

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Eg:- Candidate have job offer → Wants to decide whether accept offer a not.



- featured values are prefered to be categorical.
If the values are continuous then they are discretized prior to building the model.

- Decision Tree follow Sum of Product (SOP) rep
- The SOP is also known as Disjunctive Normal

Algorithms used in DT

(1) ID3

(2) C4.5

(3) CART (Classificatin And Regression Tree)

(4) CHAID (Chi-square Automatic Interaction detection)

(5) MARS → Multivariate adaptive regression Splines

- In DT, major challenge is to identify the attribute for the root node level.

- Two popular attribute selection measures are used:.

(1) Information Gain

(2) Gini Index.

Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

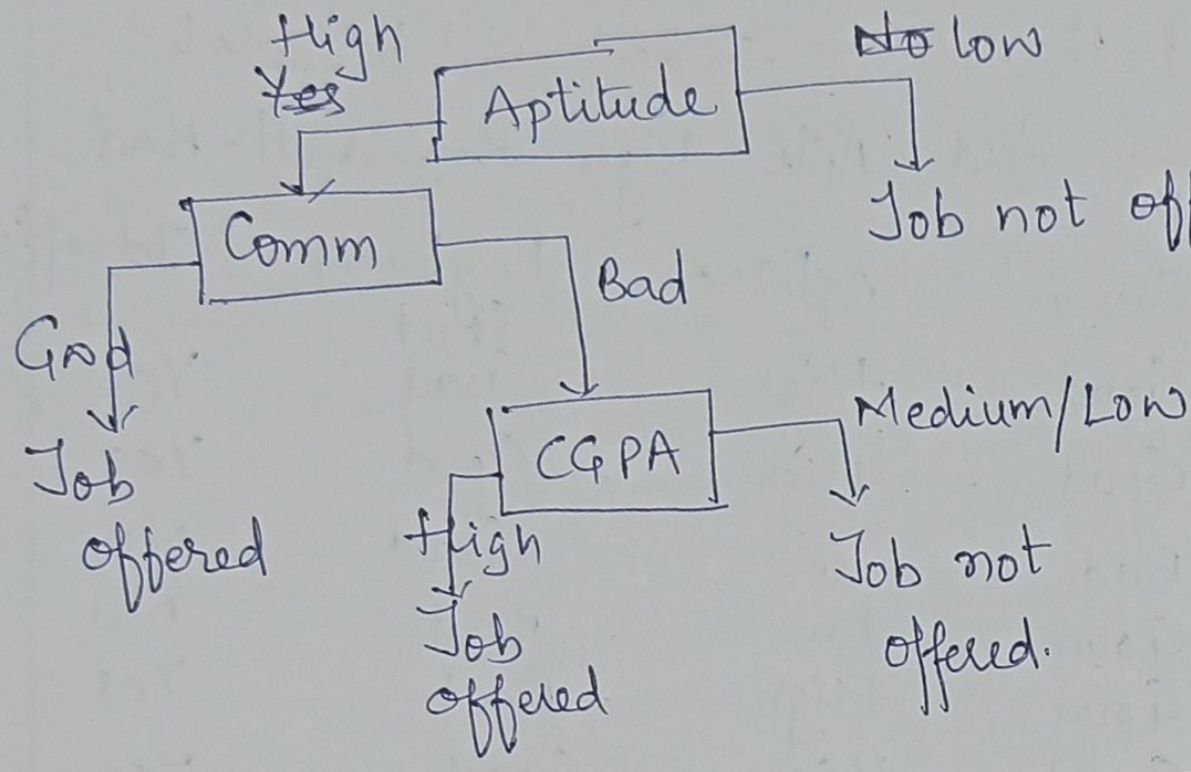
Q. Company is coming for placement & the table is given. Find the following student will be selected or not.

CGPA = High, Comm = Bad, Aptt = High, Prog. Skill - Bad

CGPA	Comm^n	Aptitude	Prgramig Skill	Job offred
High	Good	High	Good	Yes
Medium	Good	High	Good	Yes
Low	Bad	low	Good	No
Low	Good	low	Bad	NO
High	Good	High	Bad	Yes
High	Good	High	Good	Yes
Medium	Bad	low	Bad	No
Medium	Bad	low	Good	No
High	Bad	High	Good	Yes
Medium	Good	High	Good	Yes
Low	Bad	High	Bad	No
Low	Bad	High	Bad	No
Medium	Good	High	Bad	Yes
Low	Good	low	Good	No
High	Bad	low	Bad	NO
Medium	Bad	High	Good	NO
High	Bad	low	Bad	NO
Medium	Good	High	Bad	Yes

$n = 18$

If we see Aptitude = High, Job offer = Yes
else Aptitude = low, Comm^n = Good — CGPA = High
Comm = Bad — No, Job offer = No



Q:- Aptitude → high → Comm^n → Bad → CGPA → High → Job offered

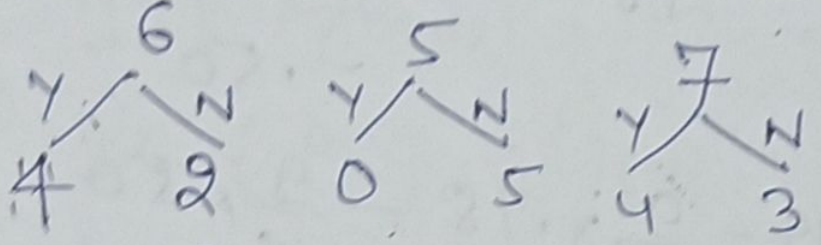
$$\text{Entropy}(S) = \sum_{i=1}^{C} -P_i \log_2 P_i$$

Job offered = Yes $8/18 = 0.44$
No $10/18 = 0.56$

$$\text{Entropy}(S) = -0.44 \log_2 0.44 - 0.56 \log_2 0.56$$
$$= 0.99$$

Entropy (CGPA)

High Low Medium



$$\text{Entropy(High)} = -\frac{4}{6}\log_2\frac{4}{6} - \frac{2}{6}\log_2\frac{2}{6}$$

$$= +.67 \times .58 - 0.33 \times 1.59$$

$$= .39 + 0.33 \times 1.59 .53$$

$$= 0.92$$

$$\text{Entropy(low)} = .0$$

$$\text{Entropy(Medium)} = -\frac{4}{7}\log\frac{4}{7} - \frac{3}{7}\log\frac{3}{7}$$

$$= 0.57 \times .811 + 0.43 \times 1.22$$

$$= 0.46 + 0.52$$

$$= 0.99$$

$$\text{InfoGain(CGPA)} = 0.99 - \frac{6}{18} \times .92 - \frac{7}{18} \times .99 + \frac{5}{18} \times 0$$

$$= 0.99 - 0.307 - 0.385$$

$$= 0.30$$

Comm²

9

Good — Bad

9 — Y / N

Y / N — Y / N

7 — 2 — 1 — 8

$$\text{Entropy(Good)} = -\frac{7}{9}\log_2\frac{7}{9} - \frac{2}{9}\log_2\frac{2}{9}$$

$$= .78$$

$$\text{Entropy(Bad)} = -\frac{1}{9}\log\frac{1}{9} - \frac{8}{9}\log_2\frac{8}{9}$$

$$= .50$$

$$\text{Info Gain}(\text{Comm}^2) = .99 - \frac{9}{18} \times .78 - \frac{9}{18} \times .50$$

$$= 0.36$$

Aptitude

11 — 7

High — low

Y / N — Y / N

8 — 3 — 0 — 7

$$\text{Entropy(low)} = 0$$

$$\text{Entropy(High)} = -\frac{8}{11}\log_2\frac{8}{11} - \frac{3}{11} \times \log_2\frac{3}{11}$$

$$= 0.85$$

$$\text{Info(Gain(A))} = .99 - \frac{11}{18} \times .85$$

$$= 0.47$$

Programing Skill

9 — 9

Good — Bad

Y / N — 3 / 6

4 — 4 — Y — N

$$\text{Entropy(Bad)} = .92$$

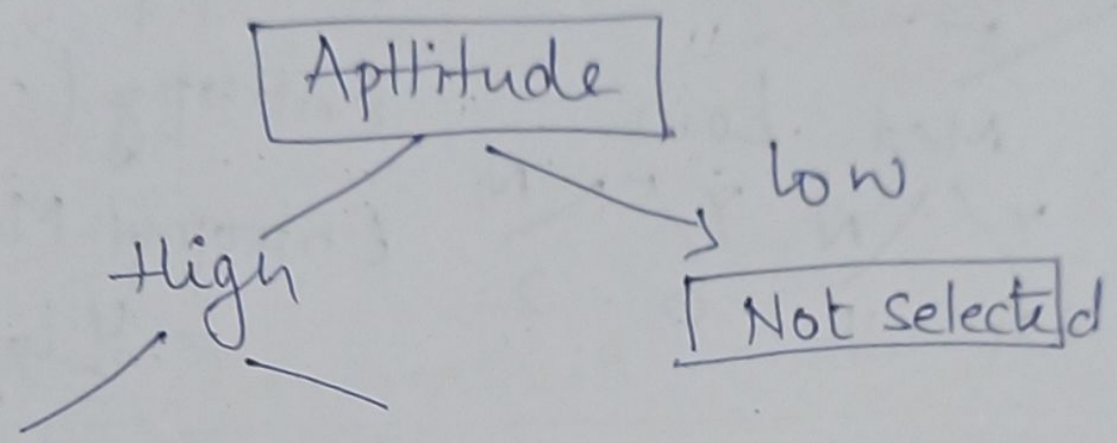
$$\text{Entropy(Good)} = .99$$

$$\text{Info Gain(PS)} = .99 - .92 \times \frac{9}{18} - .99 \times \frac{9}{18}$$

$$= 0.04$$

Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

IG (CGPA) = .30
IG (Comm^2) = .36
IG (Aptitude) = .47
IG (PS) = .04



CGPA	Comm^2	Progm Skill	Job
High	Good	Good	Yes
Med	Good	Good	Yes
High	Good	Bad	~~No~~ Yes
High	Good	Good	Yes
High	Bad	Good	Yes
Med	Good	Good	Yes
low	Bad	~~High~~ Bad	No
low	Bad	Bad	No
Med	Good	Bad	Yes
Med	Bad	Good	No
Med	Good	Bad	Yes

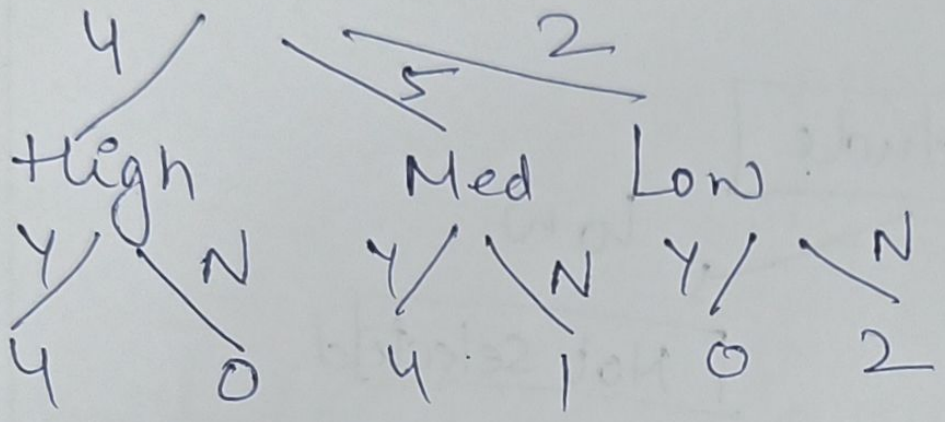
Total Case = 11

Yes = 8 , No = 3

$$\text{Entropy}(S) = -\frac{8}{11}\log_2\frac{8}{11} - \frac{3}{11}\log_2\frac{8}{11}$$

$$= .85$$

CGPA



Entropy (Low) = 0

Entropy (High) = 0

Entropy (Med)

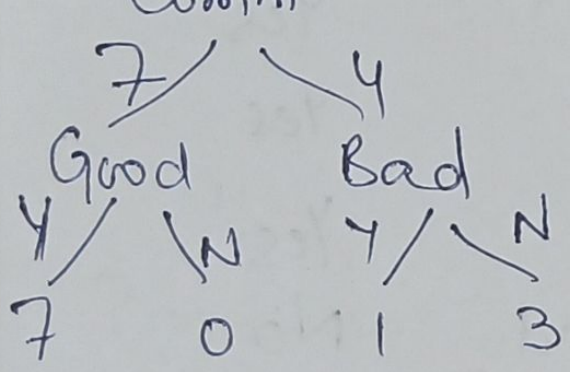
$$= -\frac{4}{5}\log_2\frac{4}{5} - \frac{1}{5}\log_2\frac{1}{5}$$

$$= .72$$

$$IG(CGPA) = .85 - \frac{5}{11} \times .72$$

$$= \boxed{.052}$$

Comm



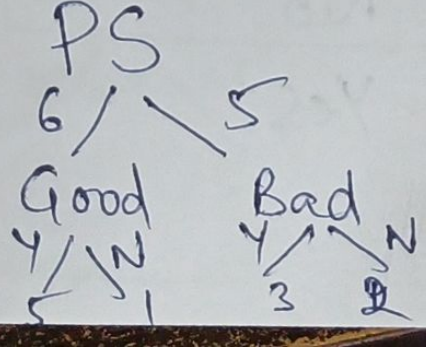
Entropy (Good) = 0

$$\text{Entropy(Bad)} = -\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4}$$

$$= .81$$

$$IG(CGPA) = .85 - \frac{4}{11} \times .81 = \boxed{0.55}$$

PS

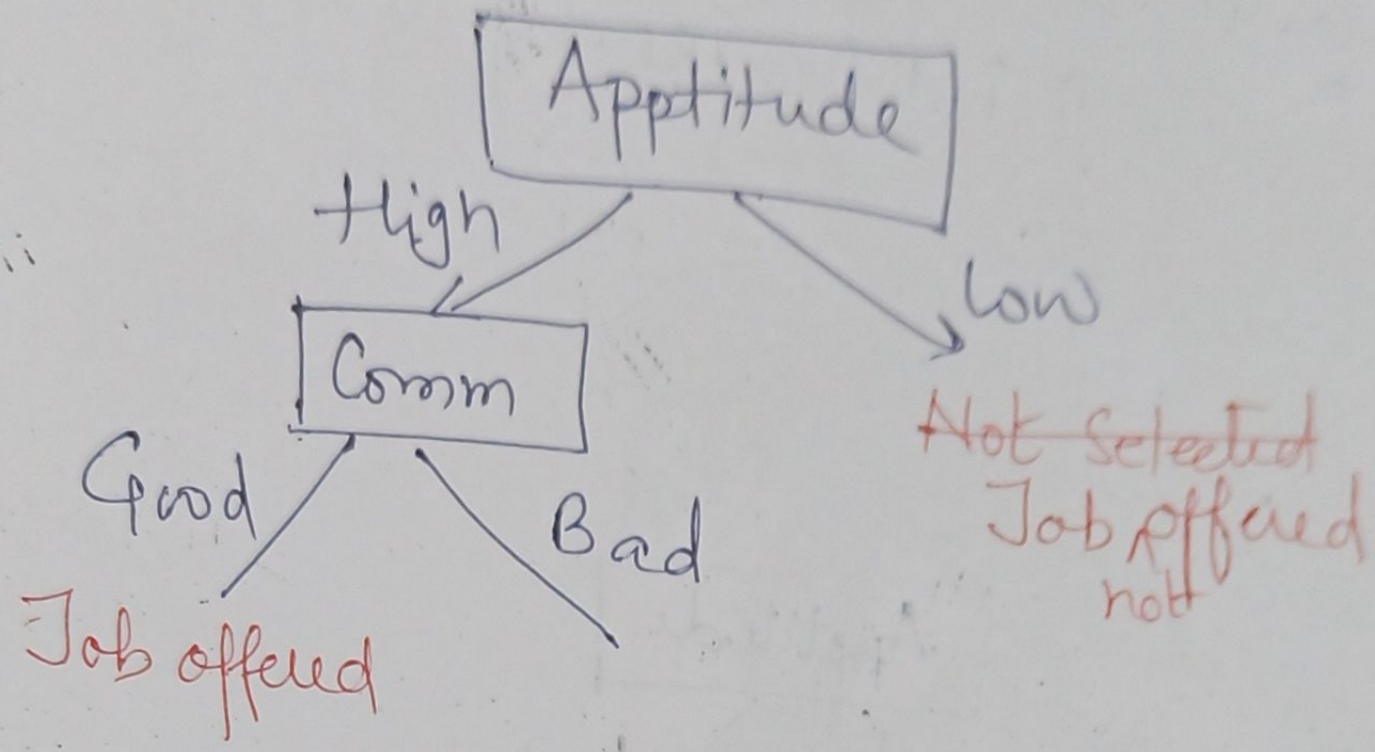


Entropy (Good) = 0.65

Entropy (Bad) = 0.97

$$IG = .85 - \frac{6}{11} \times .65 - \frac{5}{11} \times .97 = \boxed{0.05}$$

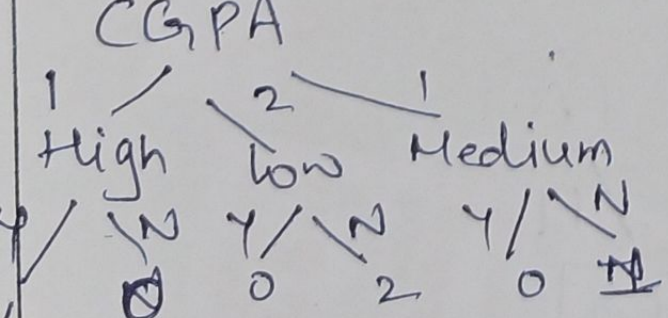
by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in



CGPA	ProgramSkill	Job
High	Good	Yes
Low	Bad	No
Low	Bad	No
Medium	Good	No

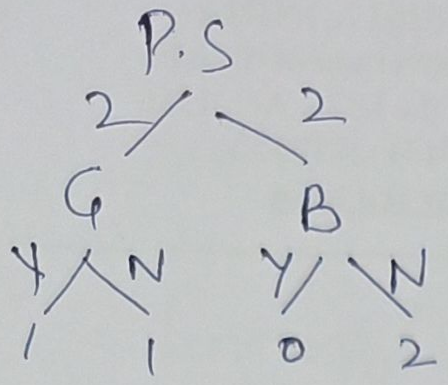
Yes = 1
No = 3

$$\text{Entropy} = -\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4}$$
$$= 0.81$$



Entropy (H) = 0
Entropy (Low) = 0
Entropy (Med) = 0

IG(CGPP) = 0.81

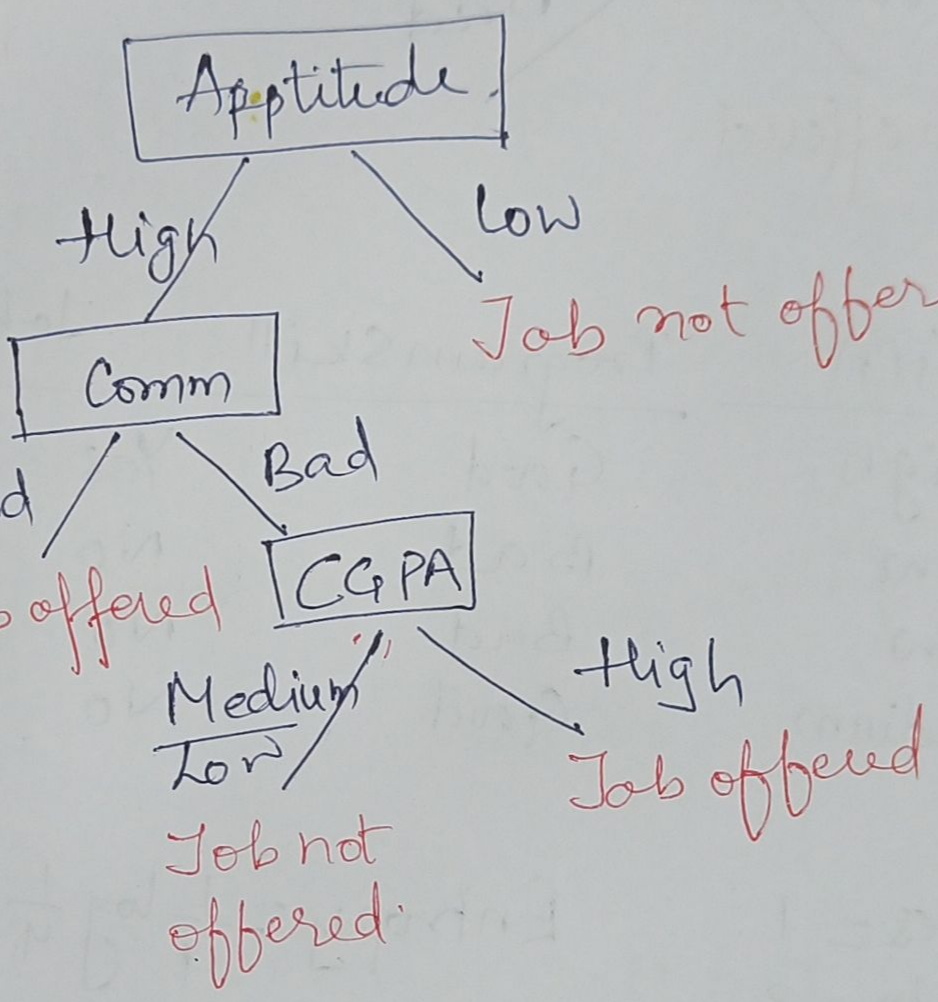


Entropy(Bad) = 0

Entropy(Good) = 1

$$IG(PS) = .81 - \frac{2}{4} \times 1 \quad \{.81 - .5\}$$

$$= .31$$



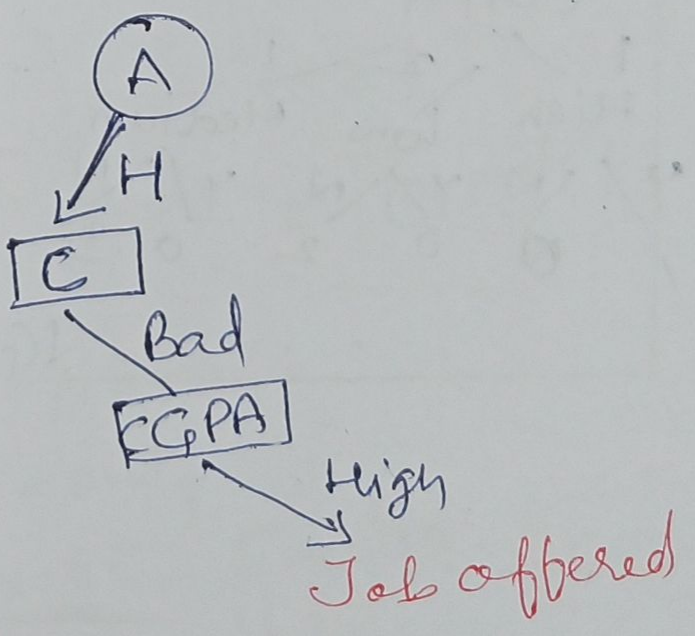
Predict for

CGPA = High

Comm - Bad

Aptitude - High

P.S → Bad



Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Gini Index:→

It measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen.

-Gini index is a proportion of the impurity or inequality of a circulation, regularly utilized as an impurity in decision tree algo.

$$Gini = 1 - \sum_{i=1}^{n} (P_i)^2$$

{ It values varies from 0 to 0.5 }

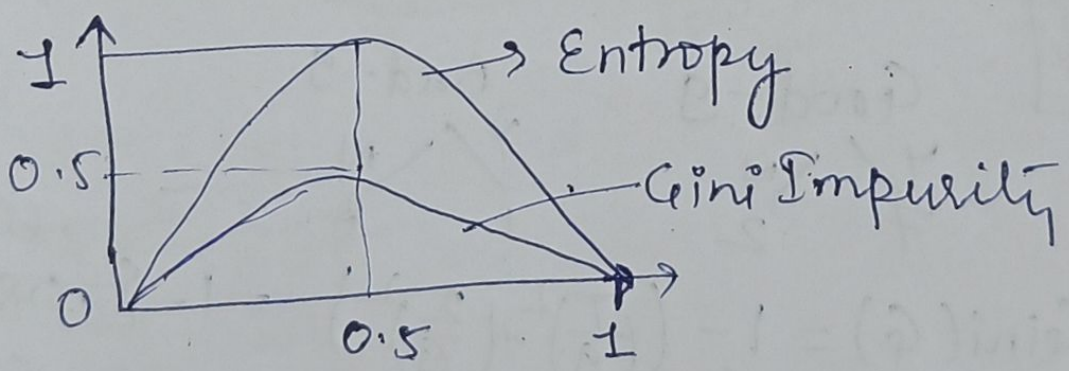
Entropy / Gini Impurity:-

$$\text{Entropy}(S) = \sum_{i=1}^{n} -P_i \log_2 P_i$$

$$Gini(S) = 1 - \sum_{j=1}^{n} P_j^2$$

$-.6 \log_2 .6$
$-.4 \log_2 .4$
$+.6 \times +.737$
$+.4 \times 1.322$
$= .44 + .53$
$= .96$

Let
$X = 6$
$N = 4$
$1 - \{(.6)^2 + .4^2\}$
$1 - .36 + .16$
$1 - .52$
$= .48$



-Gini impurity is more efficient than entropy, in terms of computing power.

Gini Index:-

CGPA →

High = 6

Y	N
4	2

Low = 5

Y	N
0	5

Medium = 7

Y	N
4	3

$$Gini(H) = 1 - \left(\left(\frac{4}{6}\right)^2 + \left(\frac{2}{6}\right)^2\right)$$

$$= 1 - (.44 + .11)$$

$$= 1 - .55 = .0.45$$

$$Gini(L) = 1 - \left(\left(\frac{5}{5}\right)^2 + \left(\frac{0}{5}\right)^2\right) = 0$$

$$Gini(M) = 1 + \left(\left(\frac{4}{7}\right)^2 + \left(\frac{3}{7}\right)^2\right)$$

$$1 - \left((.57)^2 + (.42)^2\right) = 1 - (.33 + .06)$$

$$= .61$$

$$\text{Weighted Sum} = \frac{6}{18} \times .45 + 0 + \frac{7}{18} \times .61 = .33$$

$$= .1485 + .237$$

$$= \boxed{.39}$$

.034

Comm 2

Good-9

Y	N
7	2

Bad-9

Y	N
1	8

$$Gini(G) = 1 - \left(\left(\frac{7}{9}\right)^2 + \left(\frac{2}{9}\right)^2\right) = 1 - \{.086 + .025\} = .89$$

.604 + .04 .345

.012 + .79

$$Gini(B) = 1 - \left\{\left(\frac{1}{9}\right)^2 + \left(\frac{8}{9}\right)^2\right\} = 1 - \{.012 + .099\} = .89$$

.1975

.345 .1975

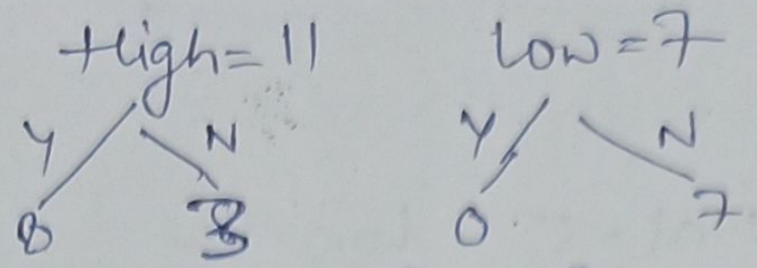
$$\text{Weighted Sum} = \frac{9}{18} \times .89 + \frac{9}{18} \times .89 = .89$$

.275

Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

Apptt

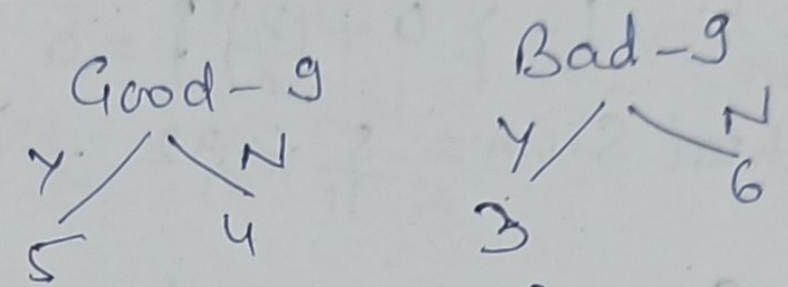


$$Gini(H) = 1 - \left\{\left(\frac{8}{11}\right)^2 + \left(\frac{3}{11}\right)^2\right\} = 1 - \{.53 + .074\} = 0.39$$

$$Gini(L) = 1 - \left\{\left(\frac{0}{7}\right)^2 + \left(\frac{7}{7}\right)^2\right\} = 1 - 1 = 0$$

$$\text{Weighted Sum} = \frac{11}{18} \times .39 + \frac{7}{18} \times 0 = \boxed{.23}$$

Program Skills :-

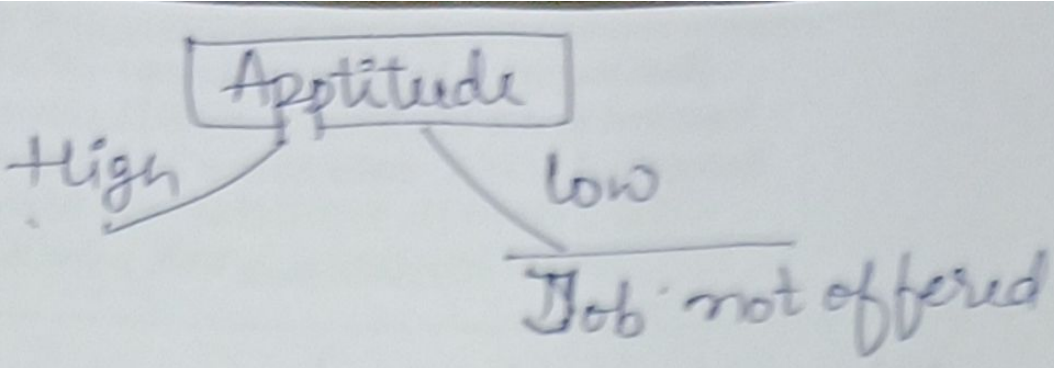


$$Gini(G) = 1 - \left\{\left(\frac{5}{9}\right)^2 + \left(\frac{4}{9}\right)^2\right\} = 1 - \{.06 + .05\} = \boxed{.89}\ .49$$

$$Gini(B) = 1 - \left\{\left(\frac{3}{9}\right)^2 + \left(\frac{6}{9}\right)^2\right\} = 1 - \{.04 + .07\} = .89\ .44$$

$$\text{Weighted Sum} = \frac{9}{18} \times .89\ .49 + \frac{9}{18} \times .89 = .89\ .465\ .44$$

→ CGPA → .39 .34
→ Comm² → .89 .275
→ Apptt → .23 .24
→ Progskills → .89 .465



Remaing Cases:→ 11

CGPA

High = 4 Med = 5 low = 2

High		Med		low	
Y	N	Y	N	Y	N
4	0	4	1	0	2

$Gini(H) = 1 - \{(\frac{4}{4})^2 + 0\} = 0$

$Gini(M) = 1 - \{(\frac{4}{5})^2 + (\frac{1}{5})^2\} = 1 - \{.64 + .04\} = .32$

$Gini(L) = 1 - \{0 + (\frac{2}{2})^2\} = 0$

$Weighted\ Sum = \frac{5}{11} \times .32 = \boxed{.145}$

Comm^2

Good - 7 Bad - 4

Good		Bad	
Y	N	Y	N
7	0	1	3

$Gini(G) = 1 - \{(\frac{7}{7})^2 + 0\} = 0$

$Gini(B) = 1 - \{(\frac{1}{4})^2 + (\frac{3}{4})^2\} = 1 - \{.06 + .56\} = .38$

$Weighted\ Sum = \frac{4}{11} \times .38 + 0 = \boxed{.1038}$ ← minimum

PS

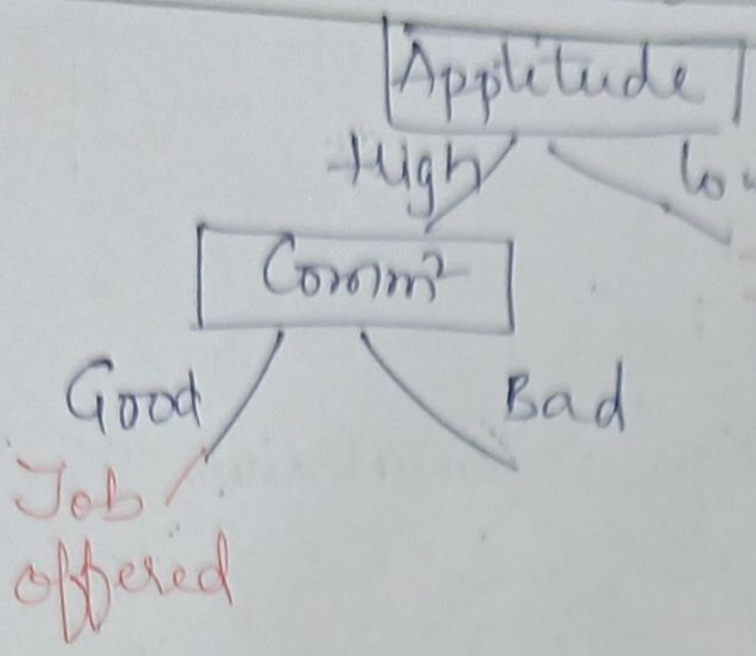
Good - 6 Bad - 5

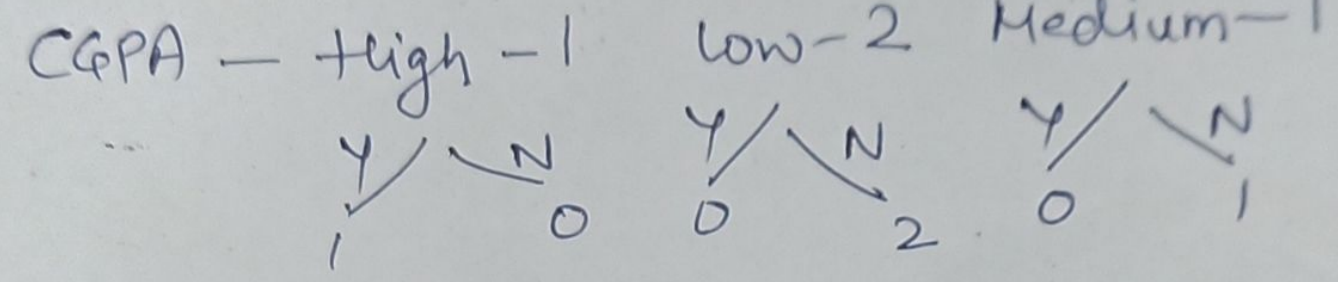
Good		Bad	
Y	N	Y	N
5	1	3	2

$Gini(G) = 1 - \{(\frac{5}{6})^2 + (\frac{1}{6})^2\} = 1 - \{.139 + .028\} = .83$

$Gini(B) = 1 - \{(\frac{3}{5})^2 + (\frac{2}{5})^2\} = 1 - \{.36 + .16\} = .48$

$Weighted\ Sum = \frac{6}{11} \times .83 + \frac{5}{11} \times .48 = .45 + .22 = \boxed{.67}$

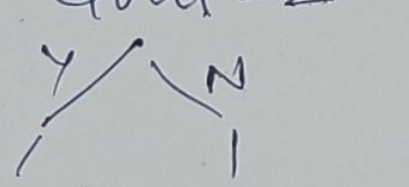




Gini (H) = 0
Gini (L) = 0
Gini (M) = 0

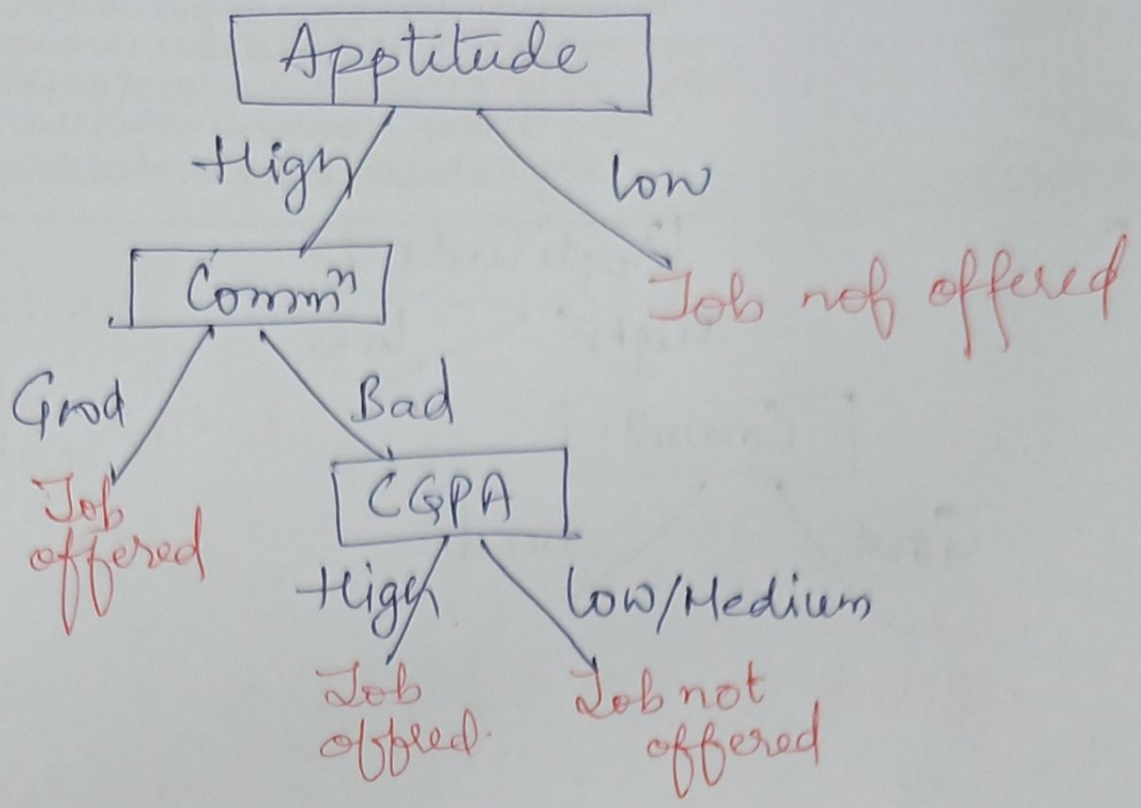
Weighted Sum = 0

~~CGPA~~
Prog Skill



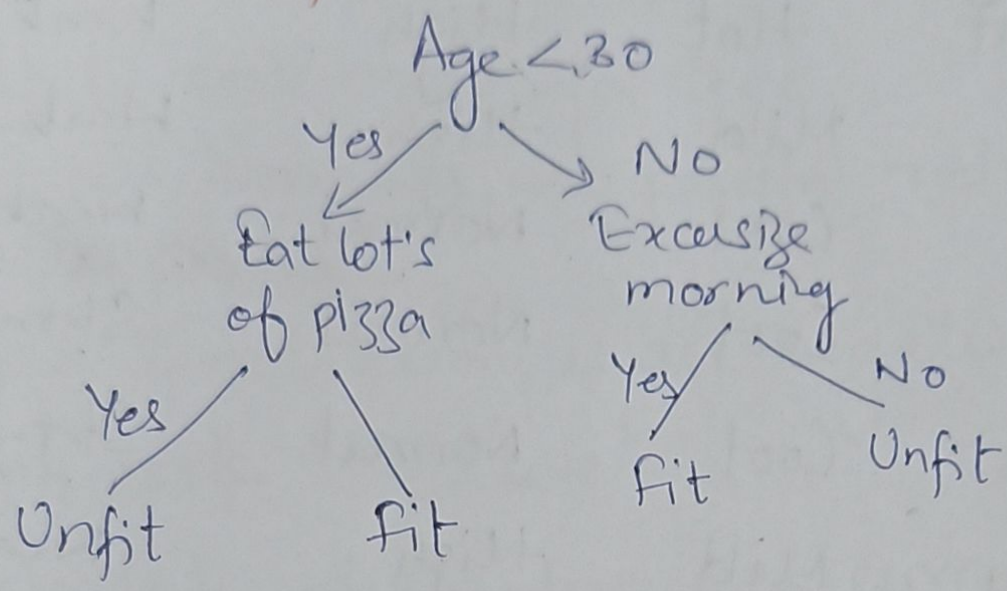
Gini (G) = ~~0.5~~ $1-\{(\frac{1}{2})^2+(\frac{1}{2})^2\} = 0.5$
Gini (B) = 0

Weighted Sum = $\frac{2}{4} \times 0.5 = .25$


Apptitude
High
low
Comm^n
Job not offered
Good
Bad
Job offered
CGPA
High
low/Medium
Job offered
Job not offered

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Eg: Person is fit or not



ID3 (Iterative Dichotomiser 3)

(1) Entropy = $H(S) = \sum_{x \in X} p(x) \log_2 p(x)$

measure the amount of uncertainty or randomness in data

Example → if toss coin, prob of head $P(H) = 0.5$

if both side head, prob(1) → so entropy = 0 no randomness.

(2) Information Gain

$$IG(S, A) = H(S) - H(S, A)$$

$$IG(S, A) = H(S) - \sum_{i=0}^{n} P(x) * H(x)$$

↓ entropy of entire set

- It denote effective change in entropy after deciding on a particular attribute A.

Day	Outlook	Temp	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Strong	No
6	Rain	Cool	Normal	Strong	Yes
7	Overcast	Cool	Normal	Weak	No
8	Sunny	Mild	High	Weak	Yes
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	~~Mild~~ Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

असतो मा सद्गमय
Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

ID3 Algorithm :→

1. Create root node for tree
2. If all examples are positive, return leaf node 'positive'
3. Else if all examples are negative, return leaf node 'negative'
4. Calculate the entropy of current state H(S)
5. For each attribute, calculate the entropy with respect to the attribute 'x' denoted by H(S, x)
6. Select the attribute which has maximum value of IG(S, x)
7. Remove the attribute that offers highest IG from the set of attributes
8. Repeat until we run out of all attributes, or the tree has all leaf nodes.

Yes = 9 $\quad$ No = 5 $\rightarrow$ Total = 14

$P(Y) = 9/14 \quad P(N) = 5/14$

① Entropy (S) $= \sum p_x \log_2 \frac{1}{p(x)}$

$= \frac{9}{14} \log_2 \frac{1}{9/14} + \frac{5}{14} \log_2 \frac{1}{5/14}$

$= .64 \times \log_2 1.56 + 0.357 \cdot \log_2 2.80$

~~0.193~~ .6415 $\quad$ 1.485 $\quad$ .447

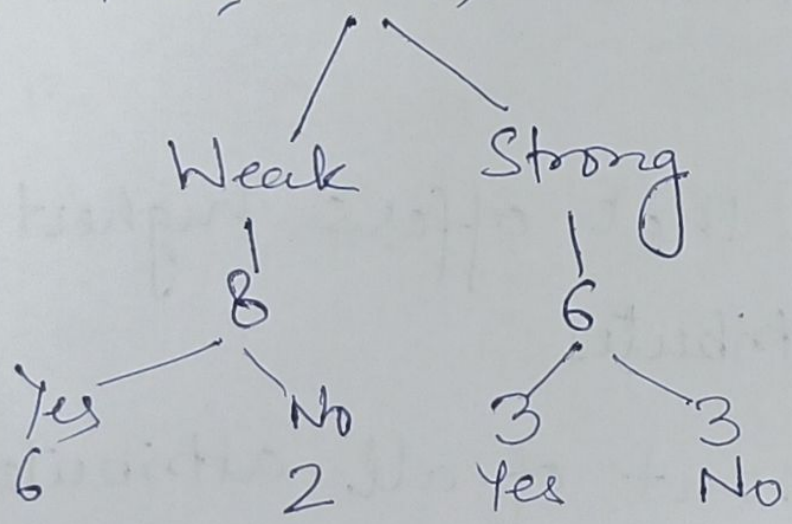
~~.123~~ .4096 + .530

$= \boxed{.940} \longrightarrow$ Distributed.

If entropy = 0 → All members belong to same Class

If entropy = 1 - half belongs to one class & half belongs to second class

② $IG(S, Wind) = H(S) - \sum_{i=0}^{n} P(x) * H(X)$



$P(Weak) = 8/14$

$P(Strong) = 6/14$

$Entropy(Weak) = -\left(\frac{6}{8}\right) \log_2 \frac{6}{8} - \frac{2}{8} \log_2 \frac{2}{8}$

$= .811$

$Entropy(Strong) = 1$

$$\text{Info Gain}(S, \text{Wind}) = 0.940 - \left(\frac{8}{14}\right)(.811) - \left(\frac{6}{14}\right)(1)$$

$$\boxed{\text{Wind} = 0.048}$$

Outlook

Sunny Overcast Rain

5 4 5

2 3 4 0 3 2

$P(\text{Sunny}) = 5/14$ $P(\text{Overcast}) = 4/14$ $P(\text{Rain}) = \frac{5}{14}$

$$\text{Entropy}(\text{Sunny}) = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5}$$

$= .4 \times 1.322 + .6 \times .737$

$= .5288 + .4422 = 0.971 = 1.5142$

Entropy (Overcast) → 0 ↑ same

$$\text{Entropy}(\text{Rain}) = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5}$$

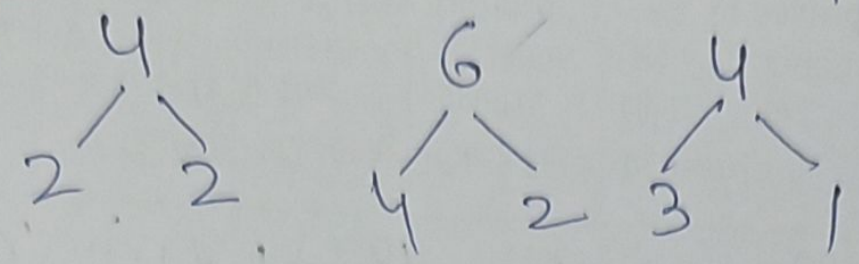
$= 0.971$

$$\text{InfoGain}(S, \text{Outlook}) = .940 - \frac{5}{14} \times .971 - \frac{5}{14} .971$$

$= .940 - .694$

$$\boxed{\text{Outlook} = .246}$$

Temp → Hot Mild Cool



$E(Hot) = 1$

$E(Mild) = -\frac{4}{6}\log_2\frac{4}{6} - \frac{2}{6}\log_2\frac{2}{6} \Rightarrow .67 \times .5995 + .33$
$.402 + .528 = 1.599$

$E(Cool) = -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4} = .929$

$= .75 \times .415 + .25 \times 2$

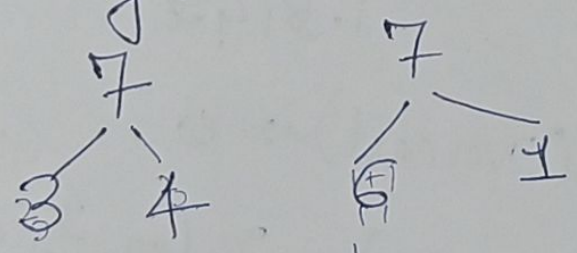
$= .311 + .5$

$= .811$

$IG(Temp) = .940 - \frac{4}{14} \times 1 - \frac{6}{14} \times .929 - \frac{4}{14} \times .811$

$= .029$

Humidity High Normal



$E(High) = -\frac{3}{7}\cdot\log_2\frac{3}{7} - \frac{4}{7}\cdot\log_2\frac{4}{7}$

$= .429 \times 1.221 + .571 * .8084 =$

$= .524 + .462 = .985$

$E(Normal) = -\frac{6}{7} \times \log_2\frac{6}{7} - \frac{1}{7} \times \log_2\frac{1}{7}$

$= .857 \times .223 + .143 \times -2.806$

$= .191 + .401$

$= .592$

$IG(Humidity) = .940 - \frac{7}{14} \times .985 - \frac{7}{14} \times .592 = .1515$

$.4925$ $.296$

IG(Outlook) = .246 ← Highest
IG(Wind) = .048
IG(Temp) = 0.029
IG(Humidity) = 0.151

Outlook

Sunny | Overcast | Rain

Sunny → 2 Yes, 3 No

Entropy(Sunny) = $-\frac{3}{5}\log\frac{3}{5} - \frac{2}{5}\log\frac{2}{5}$

= .96 .97

Temp + Wind

Hot	High	Weak	No
Hot	High	Strong	No
Mild	High	~~Strong~~ Weak	No
Cool	Normal	Weak	Yes
Mild	Normal	Strong	Yes

Temp → Hot → 2 No, 0 Yes → 0
Mild → 1 No, 1 Yes → 1
Cool → 1

IG(Temp) = 0.96 − 0.52
= .044 / .057
= 0.96 − $\frac{2}{5}$ × 1(.4)
= .57

Humidity → High → 3 No, 0 Yes → 0
Normal → 2 Yes, 0 No → 0

IG(Humidity) = .971 − $\frac{3}{5}$ × 0 − $\frac{2}{5}$ × 0

= .971 − 0 − 0

= .971

Humidity

High → Yes: 0, No: 3
Normal → (2) Yes, No: 0

Entropy(High) = 0

Entropy(Normal) = 0

IG(Humidity) = .96

Weak → Y: 1, N: 2
Strong → Y: 1, N: 1 → 1

Entropy = $-\frac{1}{3}\log_2\frac{1}{3}$

$-\frac{2}{3}\log_2\frac{2}{3}$

$IG(Wind) = .971 - \frac{3}{5} \times .924 - \frac{2}{5} \times 1$

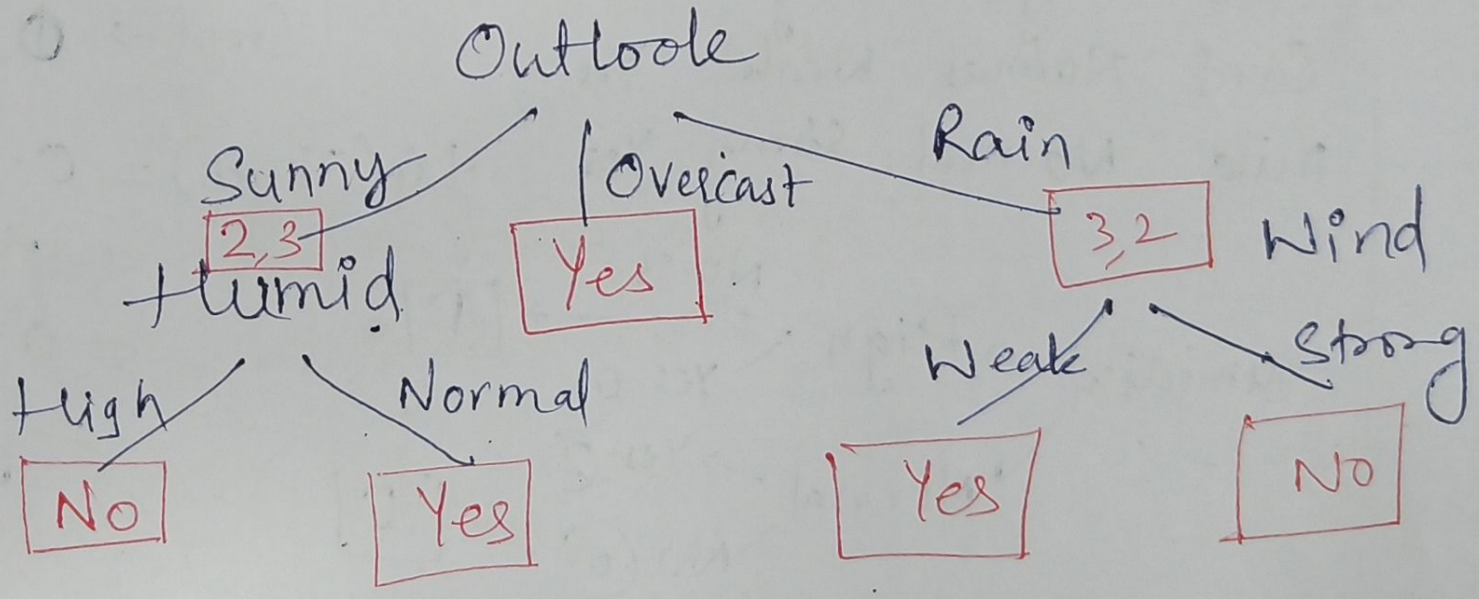
$= 0.019$

$= .33 \times 1.5995 + .66 \times$

$= .528 + .396 \cdot 599$

$= .92367$

IG(Wind) = 0.019
IG(Humid) = .96
IG(Temp) = 0.57



Rain

Temp	Humidity	Wind	Play
Mild	High	Weak	Yes
Cool	Normal	Weak	Yes
Cool	Normal	Strong	No
Mild	Normal	Weak	Yes
Mild	High	Strong	No

$$E(Rain) = -\frac{2}{5}\log\frac{2}{5} - \frac{3}{5}\times\log_2\frac{3}{5}$$

$$= .971$$

Temp — Mild: Y 2, N 1 — Entropy $= -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}$

$= .66\times1.598 + .33\times .599$

$= .92367$

Temp — Cool: Y 1, N 1 → Entropy(Cool) = 1

$$IG(Temp) = .971 - \frac{2}{5}\times1 - \frac{3}{5}\times .92367$$

$$= .971 - 0.4 - .554$$

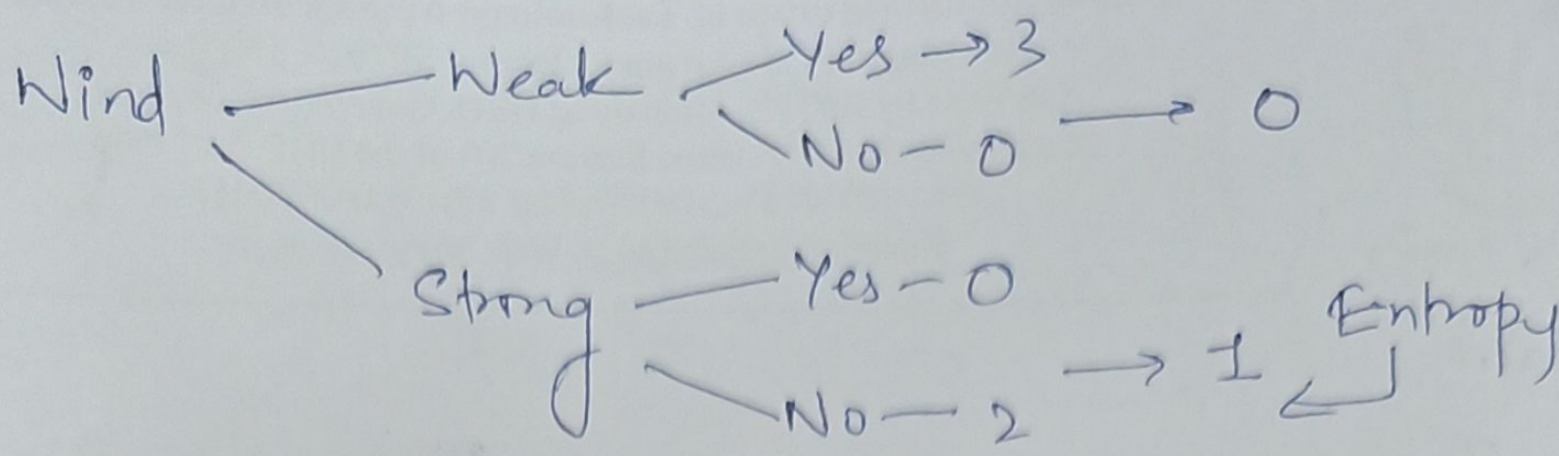
$$\boxed{= .017}$$

Humidity — High: Y 1, N 1 → 1

Humidity — Normal: Y 2, N 1 → .92367

IG(Humidity) = .017

$= .017$



$$IG(Wind) = .971 \times \frac{2}{5} \times 1$$

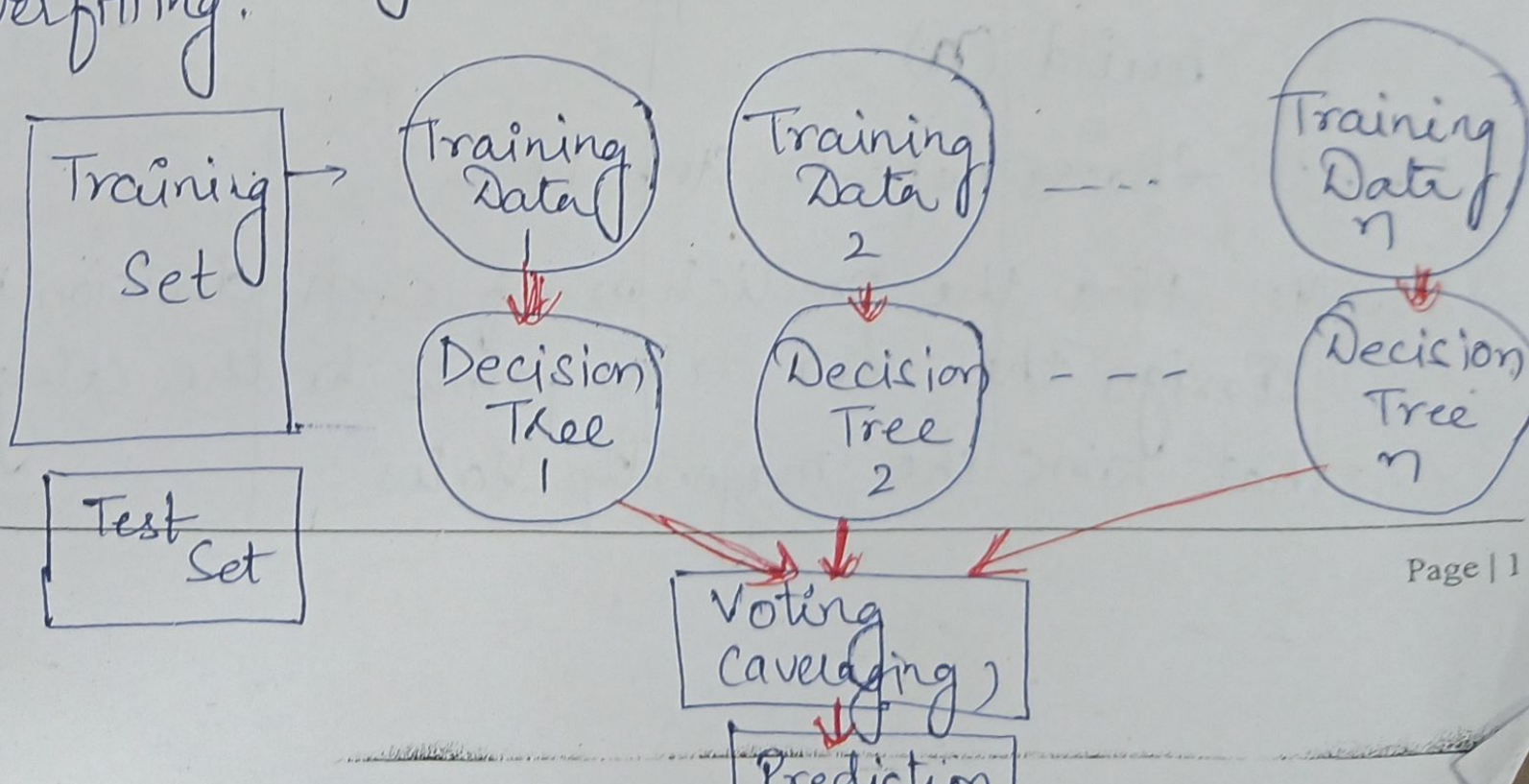
$$= .971 - .4$$

$$IG(Wind) = .571$$

SKIT
असतो मा सद्गमय
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Random forest Algorithm :→

- It's popular machine learning algo that belong to supervised learning technique.
- A random forest is an ensemble learning method where multiple decision tree are constructed and then they are merged to get a more accurate prediction.
- Random forest became popular because of its ease of use and fexibility in handling both classification and regression problems.
- ensemble learning :- procer of combining multiple classifier to solve a complex problem and to improve the performance of the model.
- The greater no. of trees in the forest leads to higher accuracy and prevents the problem of overfitting.



Assumption for Random forest :-

- Random forest is combination of multiple decision tree { some gives correct prediction and some gives wrong prediction }
- few assumptions are used in Random forest
  1) Actual values should be there in feature variable of dataset, so classifier predict accurate result.
  2) Each tree must have low correlation

Working of Random forest algorithm:→

Random forest works in two-phase

1) Create random forest by combining N decision tree,
2) Make prediction of each tree

Lets
N=10 — a
10-D — each
30 → K

Step 1:- Build decision tree with selected data points (K)

Step 2:- Choose the no. for decision tree you want to build (n)

Step 3: ~~Choose~~ Repeat Step 1 & 2

Step 4: find the prediction of each decision tree, and assign the new data points to the category that wins the majority votes.

Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

Example:- You have 100 fruits given, make the 3 DT with 30, 30, 30 data points

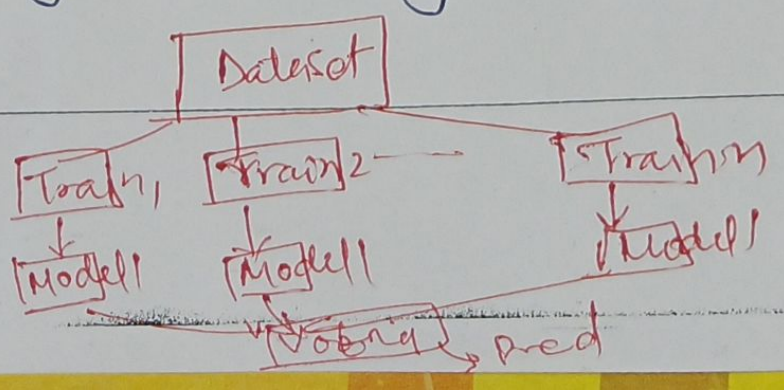
- Let DT-1 predict - Apple
- " DT-2 predict - Apple
- DT-3 predict - Banana

> → final prediction → Apple

Types of Ensemble learning:→

1) Bagging:- (Bootstrap Aggregating)
It involves training multiple models on random subsets of the training data. Prediction from the individual models are then combined, typically by averaging.

2) Boosting:→ It involves training a sequence of models, where each subsequence model focuses on the errors made by the previous model. The predictions are combined using a weighted voting scheme.



Application of Random forest: →

1) Banking → {identification of loan}
2) Medicine → {disease trends & risk of disease}
3) Land Use → {similar areas identification}
4) Marketing → {Marketing trend}

Advantages of Random forest :-

- Capable for both classification and Regression
- Handle large dataset with high dimensionality
- Enhances the accuracy of the model and prevents the overfitting issues.

Disadvantages of Random forest :-

- Not more suitable for Regression task.
- More expensive than decision tree

Random forest Work: →

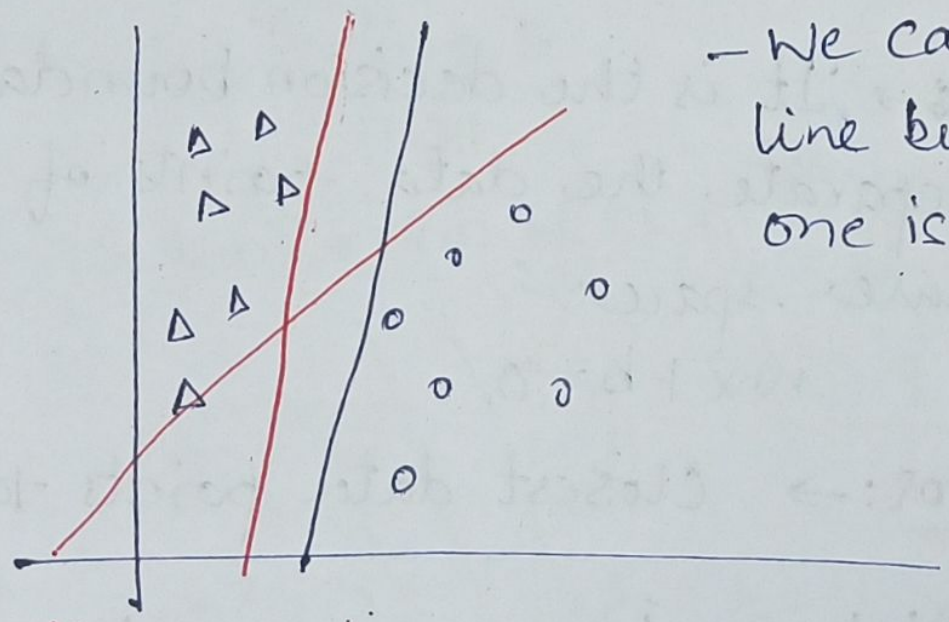
Shri Ramswaroop Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

Support Vector Machine (SVM):-

- Support Vector Machine or SVM is one of the most popular Supervised learning algo, which is used for classification as well as Regression problems.
- It is used for Classification problems in Machine learning.
- The goal of the SVM algo is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future.
- This best decision boundary is called a hyperplane.

eg..

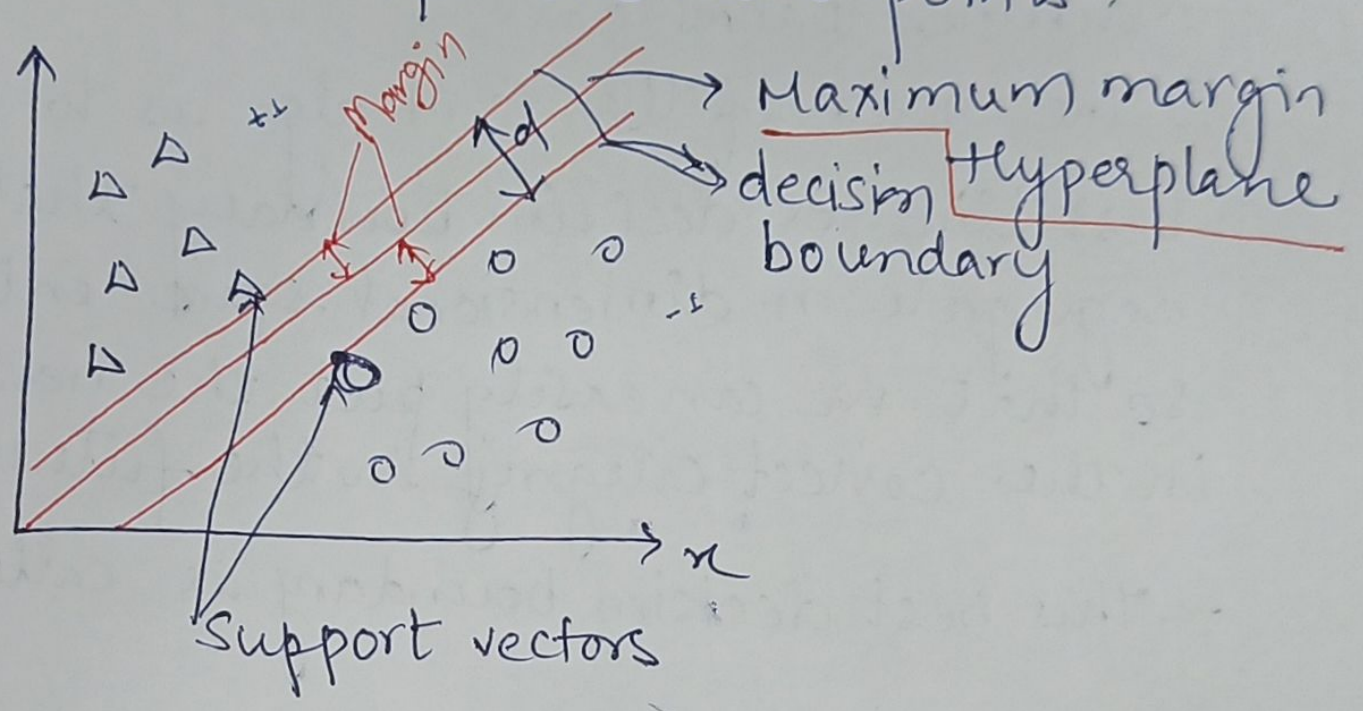


- We can draw many line but find which one is the best

- for n-dimension feature space, hyperplan is a flat subspace of dimension (N-1)

→ 2 features - 1 straight line
→ 3 feature - 2-D hyperplane

- The dimension of the hyperplane depend on the feature present in the dataset, which means if there are 2 features (x, y axis), then hyperplan will be a straight line.
- If more than 2 features, hyperplane will be a 2-dimension plan.
- Hyperplane has a maximum margin, means maximum distance b/w the data points.



Hyperplanes → It is the decision boundary that is used to separate the data points of diff. classes in a feature space:-

$$wx + b = 0$$

Support vector:→ Closest data points to the hyperplane

Margin:→ Distance b/w. support vector and hyperplane.

Kernel:- Mathematical func, used to map the original i/p data points into high-dimensional feature space,

Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Mathematical of SVM:-

Let have two classes -1 & +1. We have training dataset consisting of input feature vectors X and their corresponding class labels Y.

Liner hyperplane

$$w^T x + b = 0$$

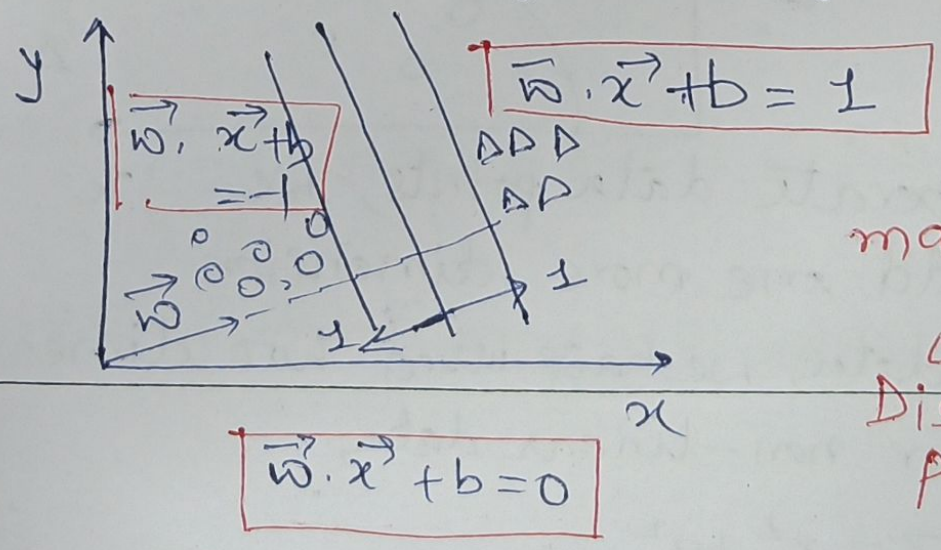
w - normal vector to the hyperplane (direction perpendicular to the hyperplane)

b → offset or distance of the hyperplan from the origin along the normal vector w.

-Distance between a data point x_i and the decision boundary can be calculated as:

$$d_i = \frac{w^T x_i + b}{\|w\|}$$

$\|w\|$ - Euclidean norm of the weight vector w.

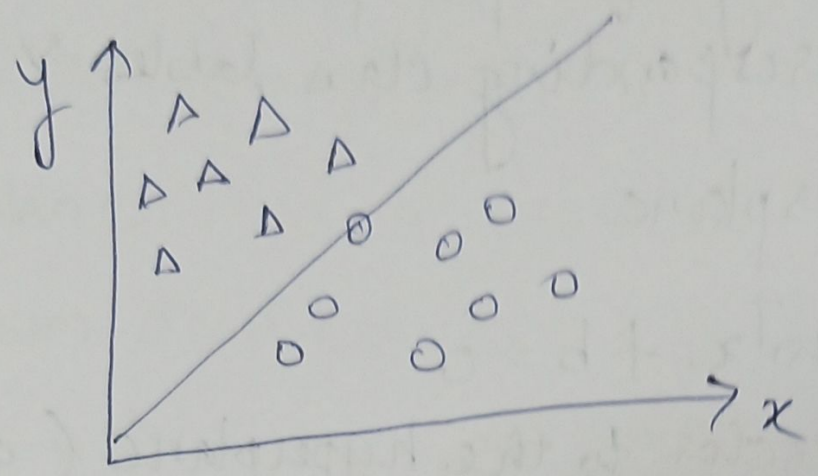


$\max \frac{2}{\|w\|}$

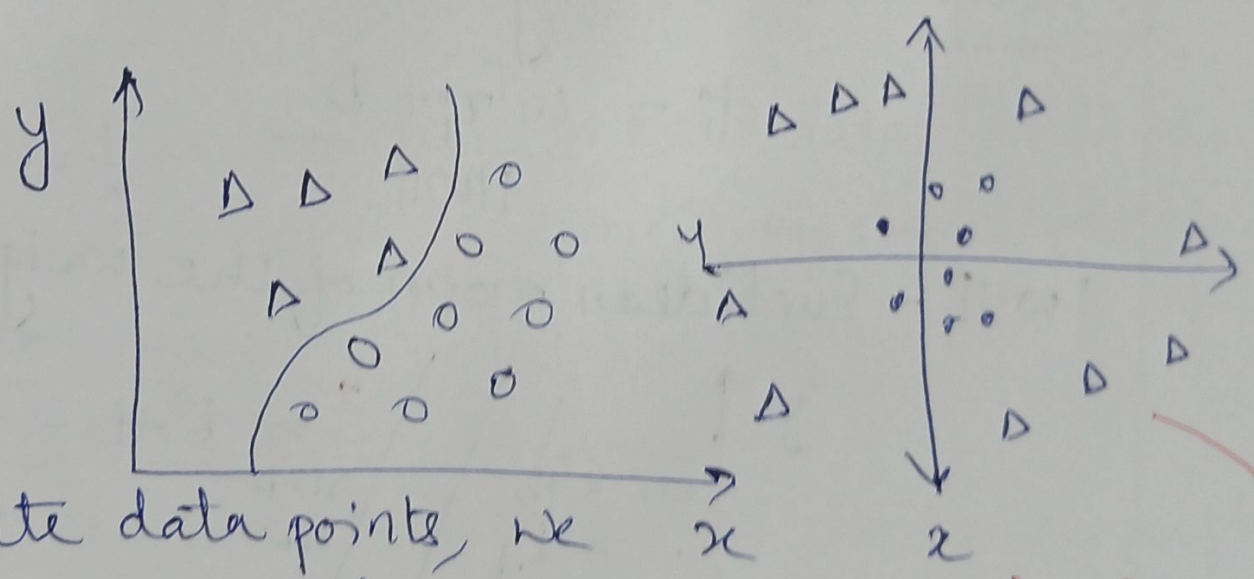
↙ Distance of planes

Types of Support Vector Machine:→

1) Linear SVM:→ Dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classier is called linear SVM classifier.

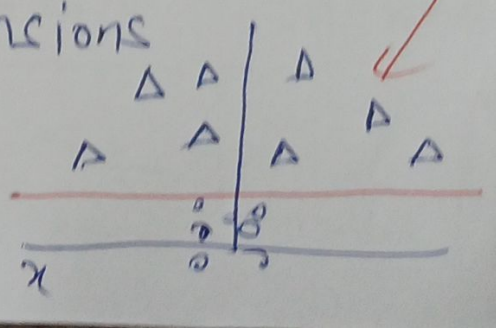


2) Non-linear SVM:→ Dataset can't be classified by using a straight line, then such data is termed as non-linear data and classifier used is called non-linear SVM



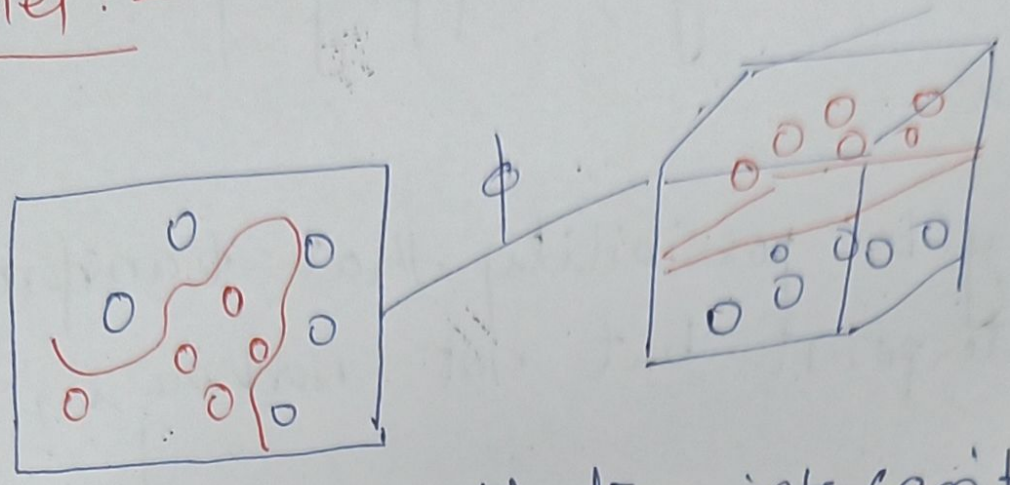
- So to seperate data points, we need to add one more dimension
- for linear data, we have used two dimensions x & y, so for non-linear data,

$$z = x^2 + y^2$$



(3)

Kernel :-



- As we can see that points can't be classified using linear line.
- for that we have to convert the data from one space to another space or known as feature space.

eg: $\phi : R^2 \rightarrow R^3$ → transform from 2D space to 3D space

$$\phi(x, y) = (x^2, \sqrt{2}xy, y^2)$$

Let points are (2,3) in 2D space, map in 3D

$x = 2, \quad y = 3$

3D Space: $(2^2, \sqrt{2} * 2 * 3, 3^2) = (4, 6\sqrt{2}, 9)$

- This mapping func play an important role,

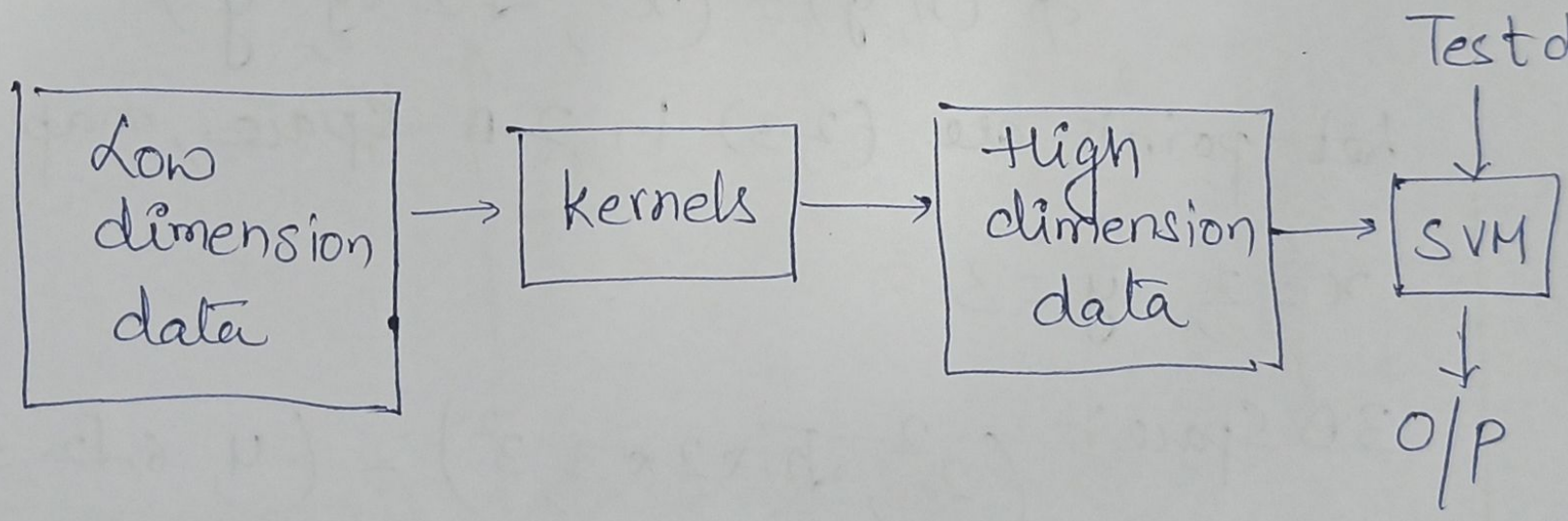
→ Disadvantage of transformatin ~~fun~~ is th
there is no generalized thumb rule for mapp
and if data is large, mapping procers takes huge
amt of time.

→ There may be possibility that transform is working
on D_1 datapoint but not working on D_2 datapoin

- Kernel are used to compute the value without transforming the data.

Kernel :-

- It's a set of func used to transform data from lower dimension to higher dimension and to manipulate data using dot product at higher dimensions.



$$k(x, y) = \phi(x) \cdot \phi(y)$$

✓ Replace mapping func with dot product

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

Types of kernel :

- Linear kernel $k(\vec{x}_i, \vec{x}_j) = \vec{x}_i \cdot \vec{x}_j$
- Polynomial kernel $k(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j + 1)^d$
- Exponential kernel
- Homogeneous kernel
- Inhomogeneous kernel
- Gaussian kernel (RBF) → $k(\vec{x}_i, \vec{x}_j) = e^{\frac{-\vec{x}_i - \vec{x}_j^2}{2\sigma^2}}$
- Hypebolic or Sigmoid kernel $k(\vec{x}_i, \vec{x}_j) = \tanh(k \cdot \vec{x}_i \cdot \vec{x}_j - \delta)$
- Radial basis function kernel

eg: Points (1,2) and (3,4)

→ mapping fun $\phi = (x^2, y^2, \sqrt{2}xy)$

$(1,2) = (1, 4, 2\sqrt{2})$

$(3,4) = (9, 16, 12\sqrt{2})$

$\phi(1,2) \cdot \phi(3,4) = 121$

In kernel (Polynomial)

$\phi(x,y) = (x^T \cdot y)^2$

$((1,2)^T \cdot (3,4))^2 =$

$\left(\begin{pmatrix}1\\2\end{pmatrix} \cdot (3,4)\right)^2 = (3+8)^2 = (11)^2 = 121$

Strength of SVM :-

- SVM can be used for both classification and regression.
- It is robust, ie. not much impacted by data with noise or outliers
- The prediction results using this model are very promising.

Weakness of SVM :→

- SVM is applicable only for binary classification, i.e when there are only two classes in the problem domain.
- SVM is very complex -
- It's slow for a large dataset, { large no. of features or instances }
- It's quite memory-intensive.

Application of SVM

→ Most suitable for binary classification.
→ Mostly used for bioinformatics (cancer / genetic disorders)
→ face recognition

SKIT
असतो मा सद्गमय
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

FP Growth Algorithm :→

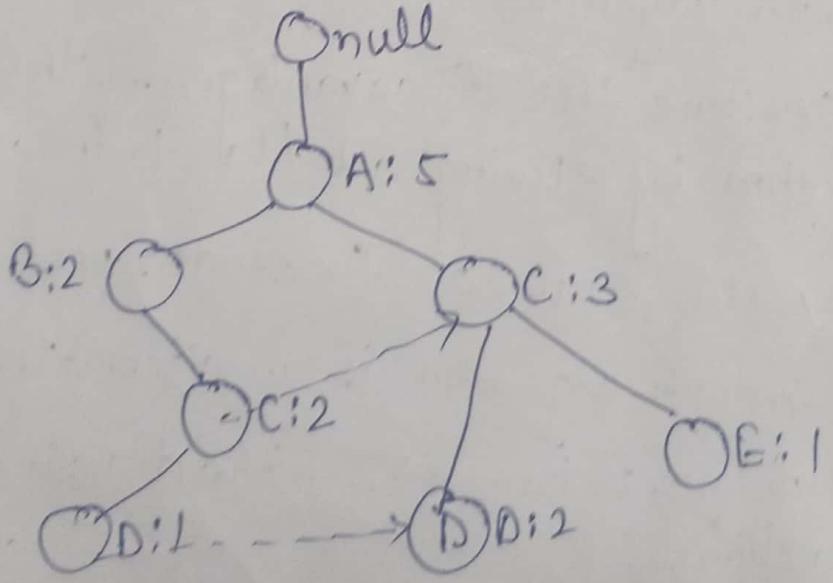
When we use apriori algorithm, we face following drawbacks in the algorithm.

1) At each step, we have to calculate candidate sets.
2) for creating the candidate sets, the algo has to repeatedly scan the dataset.

- Due to these two drawbacks, the apriori algorithm is slow.
- To overcome these drawback, new association-rule mining algo was developed.
- This new algo is known as frequent Pattern Growth algorithm.
- It overcome the disadvantages of the Apriori algorithm by storing all the transaction in a Trie Data structure. (extended prefix-tree structure)
- fp-Growth algorithm is an alternative way to find frequent item sets without using candidate generations.
- It uses divide & conquer strategy

- frequent-pattern tree (FP-tree) is a compact data structure that stores quantitative information about freq. pattern in a database.
- Each transaction is read and then mapped onto a path in FP-tree.
- following steps are used in FP-tree with item-prefix subtree and freq-item-header table

1) Root labelled as "null"
2) item-prefix-subtree consist of 3 fields
   (a) item-name
   (b) Count
   (c) node-link
3) freq-item-header consist of 2 fields
   (a) Item-name
   (b) Head of node



- Worst case for FP-Growth is, when every transaction has a unique item set.

SKIT
असतो मा सद्गमय
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Ministry of HRD, Government of India
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

FP-Growth Algorithm: →

i/p :- Transaction db with min-support
o/p → fp-tree

1. Scan DB to find occurance of itemsets. (1-itemset) freq
2. Create fp-tree with null as root node.
3. find the freq of the item, max-count item is taken at the top, and arrange in descending order.
4. Next transaction in the database is examined. Items ordered in deseending order, If transaction already in tree, it share a common prefix to root.
5. New node increase the count by 1.
6. Mine the fp tree from the lowest node and create conditional fp tree.
7. freq-Pattern are generated from conditional fp tree.

Eg:- min-supp = 50% min-confidence = 60%

TID	List of items
T_1	I1, I2, I3
T_2	I2, I3, I4
T_3	I4, I5
T_4	I1, I2, I4
T_5	I1, I2, I3, I5
T_6	I1, I2, I3, I4

Support = $\frac{50}{100} \times 6 = 3$

1) Count items

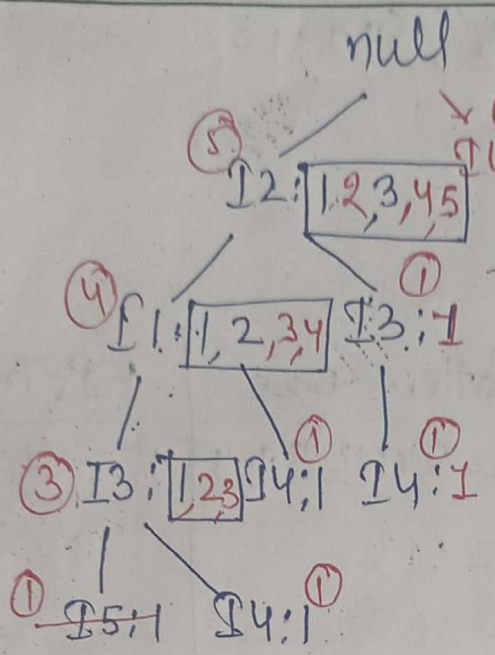
I1 – 4 ✓
I2 – 5 ✓
I3 – 4 ✓
I4 – 4 ✓
I5 – 2 ✗ Discard < 3

2) Sort in descending order

I2 – 5
I1 – 4
I3 – 4
I4 – 4

3) Build FP-Tree

null



Scan T1 I1, I2, I3

Scan T2 : I2, I3, I4

Scan T3, I4, I5

Scan T4

I1, I2, I4 {I2, I1, I4}

Scan T5 : I1, I2, I3, I5

{I2, I1, I3, I5}

Scan T6 : I1, I2, I3, I4

{I2, I1, I3, I4}

1. Scan T1 : I1, I2, I3, Arrange in Decending order I2, I1, I3

now make them child with count = I2: 1, I1; 1 & I3: 1

2. Scan T2 : I2, I3, I4 Arrange it, already arrange

add in existing branch → count I2: 2, make new branch using I3: 1 & I4: 1.

3. Scan T3 : I4, I5 already arranged.

make new branch I4: 1, I5: 1

4. Scan T4 : I1, I2, I4 → arrange I2, I1, I4

add in existing I2 = 3, I1 = 2, make new branch I4: 1

5. Scan T5 : I1, I2, I3, I5 arrange I2, I1, I3, I5

add in existing I2 = 4, I1 = 3, I3 = 2, new I5 = 1

6. Scan T6 : I1, I2, I3, I4, arrange I2, I1, I3, I4

addin existing I2 = 5, I1 = 4, I3 = 3, new I4 = 1

item	Conditional Pattern Base
I4	{I2, I1, I3:1} {I2, I1:1} {I2, I3:1}
I3	{I2, I1:3} {I2:1}
I1	{I2:4}
I2	

Now create FP-Tree

Item	Conditional Pattern Base	FP-Tree
I4	{I2, I1, I3:1} {I2, I1:1} {I2, I3:1}	I2:3
I3	{I2, I1:3} {I2:1}	I2:4, I1:3
I1	I2:4	I2:4

- from conditional FP-tree, we will calculate the freq pattern

Item	FP Generated
I4	{I2, I4:3}
I3	{I3, I2:4} {I1, I3:3}
I1	{I1, I2:4} {I1, I2, I3:3}

SKIT
असतो मा सद्गमय
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Eg:-

T1	E, K, M, N, O, Y
T2	D, E, K, N, O, Y
T3	A, E, K, M
T4	C, K, M, U, Y
T5	C, E, I, K, O, O

A - 1 ×
B -
C - 2 ×
D - 1 ×
E - 4 ✓
I - 1 ×
K - 5 ✓
M - 3
N - 2 ×
O - 4 ✓
U - 1 ×
Y - 3

Given min_Support = 3

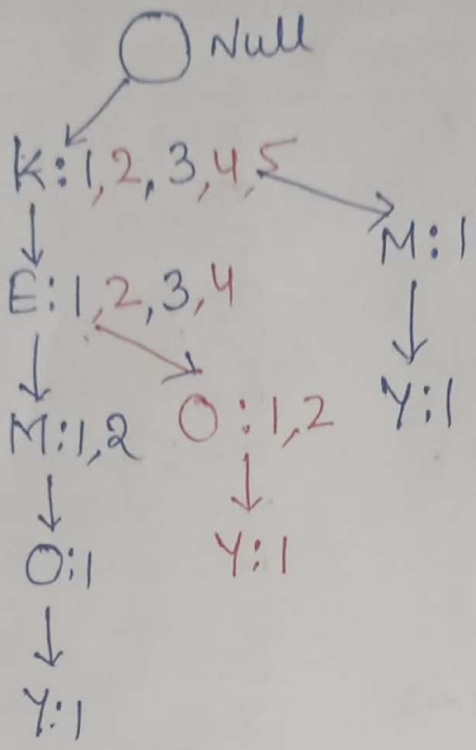
1) Arrange in Decending order

K = 5
E = 4
O = 4
M = 3
Y = 3

Prepare ordered set for Transactions

T_1	K, E, O, M, Y
T2	K, E, O, Y
T3	K, E, M
T4	K, M, Y
T5	K, E, O

① Insert T1 {K, E, M, O Y}



② Insert T2 { K, E, O Y }

③ Insert T3 { K, E, M }

④ Insert T4 { K, M, Y }

⑤ Insert T5 { K, E, O }

Now, Compute Conditional Pattern Base

item	Conditional Pattern Base
Y	{K, E, M, O:1} {K, E, O:1} {K, M:1}
O	{K, E, M:1} {K, E:2}
M	{K, E:2} {K:1}
E	{K:4}
K	

- Now for each item, Conditional frequent Pattern Tree is built,
- for this common elements in all path is counted.

Item	Conditional Pattern Base	Conditiond FP-Tree
Y	{k E, M, O : 1} {k, E, O : 1} {k, M : 1}	k = 3
O	{k E, M : 1} {k, E : 2}	k, E : 3
M	{k, E : 2} {k : 1}	k = 3
E	{k : 4}	k = 4
k		

- from the conditional fp tree, the freq Pattern rules are generated by pairing the items of Conditional fp tree set

Items	fp Generated
Y	k, Y : 3
O	{k, O : 3} {E, O : 3} {k, E, O = 3}
M	{k, M : 3}
E	{k, E : 4}
k	

Advantages of FP Growth

1. It scan the DB only twice when compared to Apriori
2. Pairing is time consuming, which we are not doing in this
3. efficient and scalable for mining both long & short freq. patterns.

Disadvantages of FP-Growth

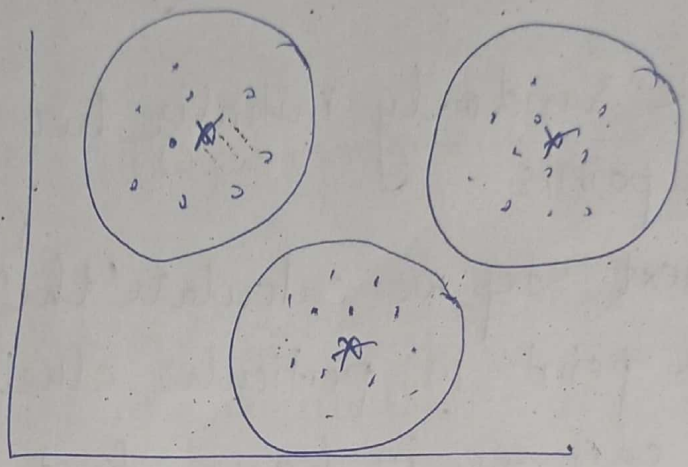
1. FP tree making is complex
2. Expensive method.
3. Algo may not fit in shared memory when DB is large.

Apriori & FP Growth Difference

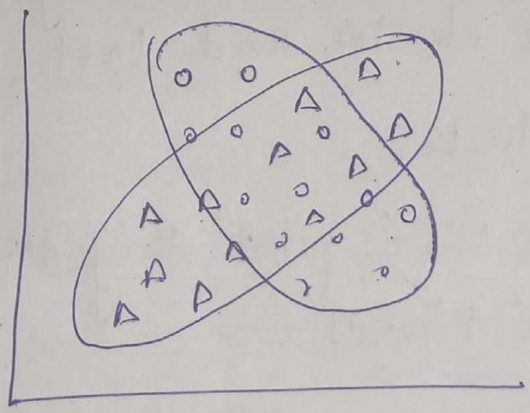
Apriori	FP-Growth
- It use pairing for single, double & triple itemset	- It use FP-Tree for making frequent patterns.
- It uses candidate generation where freq. subsets are extended one item at a time	- FP-growth generates a conditional FP-Tree for every item in the data
- Scan the DB in each step, time consuming for data.	-
-	

SKIT
असतो मा सद्गमय
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

In clustering we use k-means, where we have separate clustees like

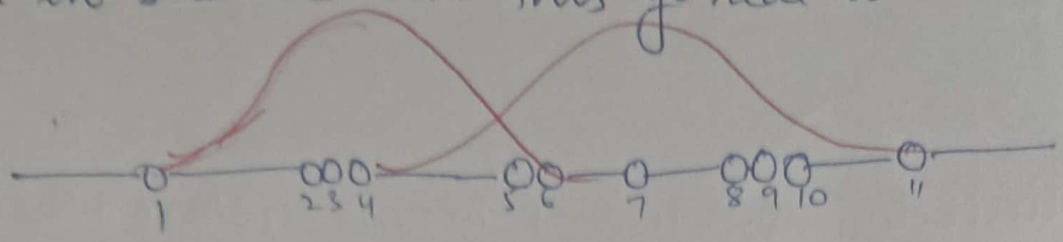


In this we have centroid and there is no overlapping in the clusters. now Suppose we have this type of data points



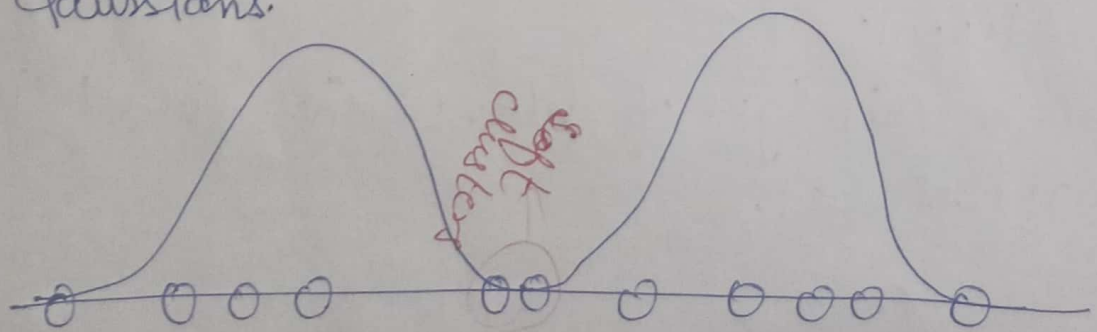
If we apply k-means on this they can't be seprated and their centroid will be overlapped.
So for this type of datasets, Gaussian Mixture Model is useed.

Let in 1D we have this go real no.



on this real no. we can apply Gaussian Model.

1. firstly we randomly initialize two Gaussians on these data points
2. In the next step we calculate the Probability of belonging the each data point in perticular cluster.
   → So we can see in figure - first point is near to the 1st GMM and last points is near to the 2nd GMM, so we can say first point belongs to 1st cluster and last point belongs to the 2nd cluster.
   - What about 5 & 6 point, they are in between 1st GMM & 2nd GMM
3. Considering the responsibility {prob} calculated in previous step, figure out more appropriate Gaussians.



4. Repeat step 3 & 4 untill each point when changes are not significant.

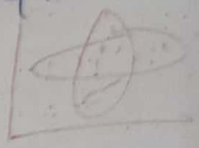
SKIT
असतो मा सद्गमय
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Gaussian Mixture Model (GMM)

K-mean is clustering method where distance is calculated b/w ekch data points using Euclidean

But in GMM it use probability distribution function.

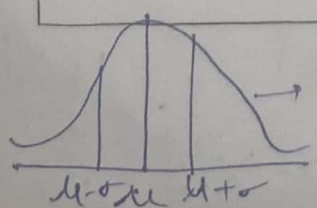
k-mean is hard clustering method & GMM is soft clustering method.



- Suppose we have dataset then using PDF is used for GMM
- GMM is a probabilistic model for representing normally distributed subpopulations within an overall population.
- In real life dataset can be modeled by Gaussian Distribution (Univariate or multivariate)
- OR GMM tried to mixture of several Gaussian Distributions.
- In one dimension the PDF of Gaussian distributed is given by

$$G(X|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

mean, variance σ^2



~~Gaussian Mixture Model~~

For multivariate → covariance matrix (d x d)

$$G(X|\mu,\Sigma) = \frac{1}{\sqrt{2\pi|\Sigma|}} \exp\left(-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)\right)$$

μ ↓ d dimensional vector mean

Let have k clusters → so calculate μ & Σ for k cluster.

Probability cluster for k distributions

$$p(X) = \sum_{k=1}^{K} \pi_k G(X|\mu_k,\Sigma_k)$$

π_k → mixing cofficient for k^{th} distribution

for estimating the parameters by the maximum log-likehood method compute $p(X|\mu,\Sigma,\pi)$

$$\ln p(X|\mu,\Sigma,\pi) = \sum_{i=1}^{N} p(X_i)$$

$$= \sum_{i=1}^{N} \ln \sum_{k=1}^{K} \pi_k G(X_i|\mu_k,\Sigma_k)$$

now define random variable $r^k(X)$

$$r_k(X) = p(k|X)$$

Use Bays theorem

$$r_k(X) = \frac{P(X|k)\cdot p(k)}{\sum_{k=1}^{K} P(k)\cdot p(X|k)} = \frac{P(X|k)\cdot \pi_k}{\sum_{k=1}^{K} P(X|k)\pi_k}$$

SKIT
असतो मा सद्गमय
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Expectation Maximization (EM) Algo.

- In real-world application of ML, it is very common that there are many relevant features available for learning but only a small subset of them are observable.
- The Expectation-Maximization algo can be used for the latent variables (variables that are not directly observable and are actually inferred from the values of the other observed variables)

EM algo:→

It is a iterative way to find maximum-likelihood estimates for model parameters when the data is incomplete or has some missing data points or has some hidden variable.

- EM choose some random points for missing data points and estimate new set of data.
- These new values are recursively used for the estimation of better values, until the values get fixed.

- In EM, E-step & M-steps are used.

Estimation Step (E-Step):-

- first we initialize model parameters like the mean μ_k, variance Σ_k and mixing coefficient π_k.
- Calculate posterior prob. for each data point belonging to each centroid using current parameters. These prob. represented by latent variables Y_k
- Lastely estimate value of latent variable Y_k based on current parameters. value.

Maximization step:→ (M-step)

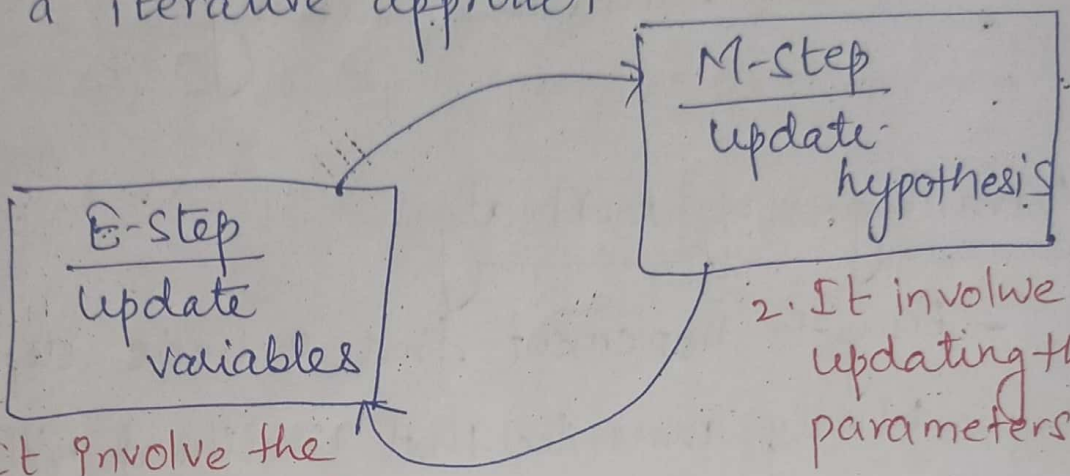
- update parameta values μ_k, Σ_k & π_k using Y_k
- update μ_k by weighted average of data points using latent variable prob.
- Update Σ_k (cov. matrix) by weighted avg. of squared diff b/w data points & mean; using latent variable prob.
- Update π_k by avg. of latent variable prob. of each components.

Repeat Estep & M-step until the log-likelihood converging.

SKIT
असतो मा सद्गमय
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

EM Algorithm:→

-It's a iterative approach



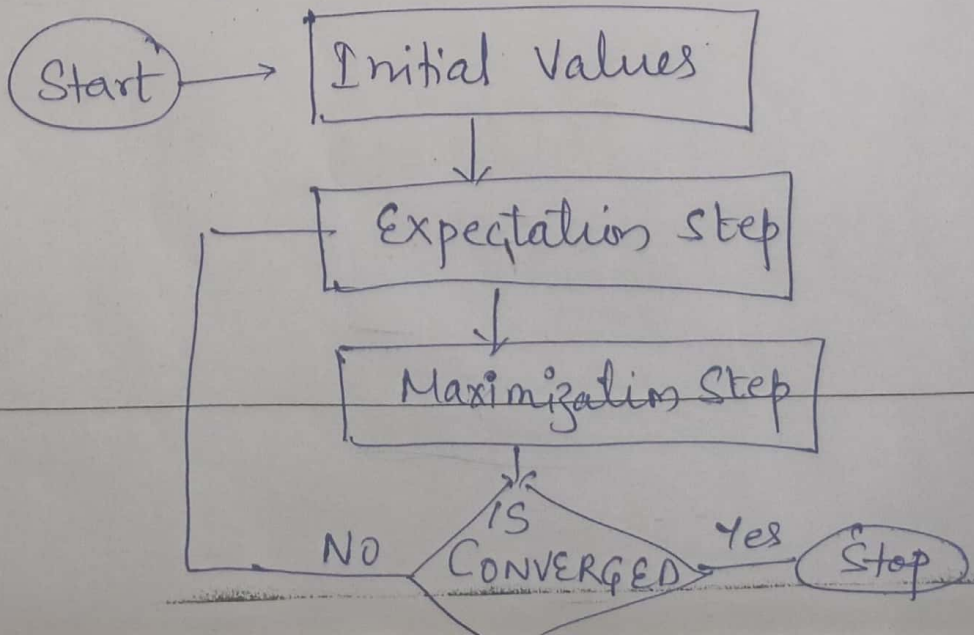
2. It involve updating the parameters.

1. It involve the estimation(guess) of missing values

Convergence:→

Convergence is defined as the specific situation in probability based on intuition.

Steps in EM algo.



~~Gaussian Mixture ...~~

Application of EM Algo:-

- Make data clusters.
- Used in computer vision & NLP
- Used in GMM
- Medical & healthcare industry

Advantages of EM algo:-

- Easy to implement first 2 basic steps
- Mostly guaranteed that likelihood will enhance after each iteration

Disadvantages of EM algo:→

- Convergence of EM algo is very slow.
- It can make convergence for local optima
- It takes forward and backward prob

SKIT
असतो मा सद्गमय
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

EM Algo Example

- Suppose we have two coins C1 & C2
- bias of C_1 is θ_1 (prob. of getting head in C_1)
- bias of C_2 is θ_2 (prob. of getting head in C_2)

H T T T H H T H T H	H = 5	T = 5
H H H H T H H H H H	H = 9	T = 1
H T H H H H H T H H	H = 8	T = 2
H T. H T. T T. H H T T	H = 4	T = 6
T H H H T H H H T H	H = 7	T = 3

Let
$\theta_A = .60$
$\theta_B = .50$

Coin A Head = .60
Coin B Head = .50

Step 1

Step 2: E-step

$P(H/A) = \binom{n}{x} \theta_A^{x} (1-\theta_A)^{n-x} = \binom{5}{5} (.6)^5 (.5)^5$

$P(H/B) = \binom{n}{x} \theta_B^{x} (1-\theta_B)^{n-x} = \binom{5}{5} (.5)^5 (.5)^5$

.01008 × 4

$= 0.012$
$= .0048$

$P(A/H) = \frac{0.012}{0.012 + .0048} = .45$

$P(B/H) = .55$

	A	B	A	B
			H=2.2, 2.2=T	H=2.8, 2.8=T --
(5,5)	.45	.55	7.2, 0.8	1.8, 0.2
	.80	.20	5.9 1.5	2.1 0.5
	.73	.27	1.4 2.1	2.6 3.9
	.35	.65	4.5 1.9	2.5 1.1
	.65	.35		
			21.3, 8.6	11.7, 8.4

Step3 M-step $\theta_A = \frac{21.3}{21.3+8.6} = .71$

$\theta_B = \frac{11.7}{11.7+8.4} = .58$

previous $\theta_A = .60$

$\theta_B = .58$

So Repeat same step

$\theta_A = .80$

$\theta_B = .52$ stop here

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Unit 4 : Semi Supervised learning

Reinforcement learning

Markov Decision Process (MDP), Bellman equation, Policy evaluation using monte carlo, Policy iteration and value iteration, Q-learning, State-Action-Reward-State-Action (SARSA), Model-based Reinforcement learning.

Markov Decision Process (MDP) :-

Reinforcement learning

- Reinforcement learning is a type of ML
- Reward and feedback is required for the agent to learn its behaviour.

Definition :→

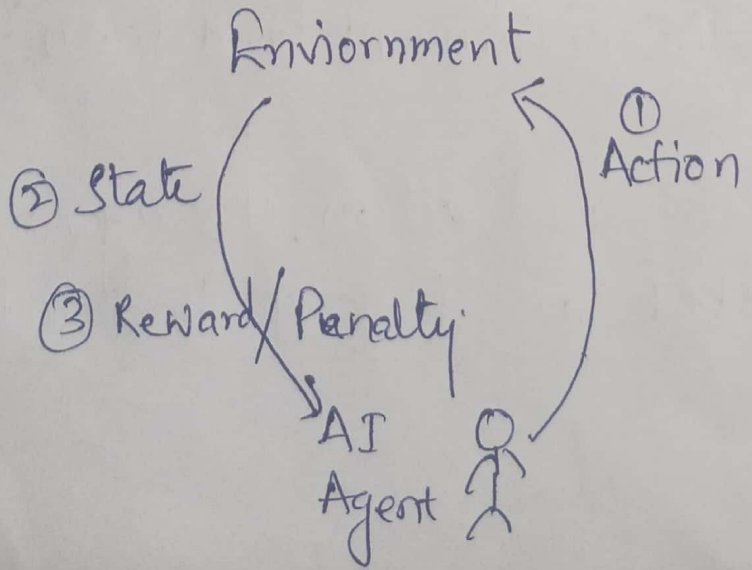
Reinforcement learning is a type of ML method where an intelligent agent (computer program) interacts with the environment and learn to act within that.

- In this for each good action, the agent gets positive feedback, and for each bad action

the agent gets negative feedback or pena

- In RL agent learn by itself (feedback), no labeled data is there
- E.g Where RL used Game-playing, Robotics.
- RL is a core part of AI and all AI agents works on the concept of reinforcement learning.
- We pre-program the agent.
- eg:- Suppose AI agent present in maze enviornment. Goal is to find diamond, and it interact with the enviornment by performing some action.
- Agent will follow these 3 things

1) take action
2) Change state / Remain same state
3) Get feedback.



Terms Used in RL :→

1. Agent → Enlity explore the enviornment
2. Enviornment → Situation in which agent present or surrounded.
3. Action :- Move by agent in Enviornment.
4. State :- Situation returned by enviornment.
5. Reward : feedback returned
6. Policy : Strategy applied by agent for next action.
7. Value :- long-term returned
8. Q-value :- Similar to value but take additional parameter + current action(a).

Key features of RL :-

1) Agent is not instructed abt enviorn. and what action need to be taken.
2) Hit & trial based process.
3) Agent may get delayed reward
4) Agent need to explore in such a way to get the max positive reward.
5) Agent take action, change state and take feedback.

Approach to Implement RL :-

Three ways

1) Value-based

- It find the optimal value-fun^{n}.
- Max value at a state under any Policy.
- Agent expects the long-term return at any state(s) under policy (π)

2) Policy-based

- find the optimal policy for the max future rewards without using the value fun^{n}.
- It use 2 type of Policy

(1) Deterministic

- same action is produced by the policy(π) at any state

(2) Stochastic

Prob. determines the produced action.

3) Model-ba

- A virtual mo is created for the environment.
- Agent explore that environment to learn it.
- No particular soln^{n}/algo for this approach because the model representation is diff for each environment.

Elements of RL :→

Main Elements are ① Policy ② Reward Signal ③ Value func ④ Model of Enviornment.

① **Policy** :→ It means a way how an agent behaves at a given time.

- It map (state of enviornment, action taken on those state)
- deterministic policy: $a = \pi(s)$
- Stochastic policy: $\pi(a|s) = P[A_t = a | S_t = s]$

② **Reward Signal** :-

- At each state, the enviornment sends an immediate signal to the learning agent.
- Reward Signal depends on good and bad action taken by the agent.
- Main objective of agent is to maximize the total no. of rewards for good action.
- Reward signal can change the policy.
- If an action selected by the agent leads to low reward, then the policy may change to select other action in the future.

③ Value func:- It gives info. about how good situation and action are.

- reward indicates the immediate signal for each good & bad action
- Whereas value func specifies the good state and action for the future.
- Value func → depend on Reward
  if no reward — no value.

④ Model:-

- It tell the behavior of the environment.
- Model Inference about how the environment will behave.
- If state and action are given → then model can predict the next state and reward.

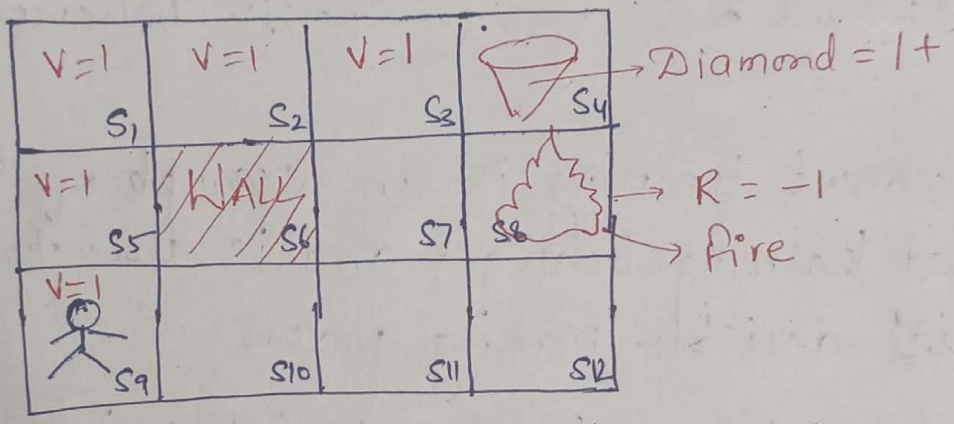
- Model is used for planning. Approach solved with the help of model are termed as model-based approach.
- Approach without using a model is called a model-free approach.

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

How does RL Work?

for RL we need two value

1. Environment { Room, Maze, football }
2. Agent { AI robot }



→ Agent is standing - S9 / S6 - Wall / S8 - fire / S4 - Diamond

→ Agent can't cross the block because it is wall.

→ If reach S4 → get +1 reward / S8 - get -1 reward

→ He can move up, down, left, right.

→ He will move S9 → S5 → S1 → S2 → S3, S0 he will get +1 → reward point

- To memorize the steps, it assign 1 value to each previous step.

So V=1 put at S9, S5, S1, S2, S3 in figure

- Agent will be confused whether to move after S1 → up or down.

Types of Reinforcement Learning:-

2 Type of RL (1) Positive RL (2) Negative RL

PRL is a recurrence of behaviour due to positive rewards
- Reward increase strength & freq of a specific behaviour
(Add something to increase the tendency that expected behaviour would occur again)

NRL, negative reward are used as a deterrent to weaken the behaviour
→ -ve Reward decrease strength and freq of specific behaviour

- Reinforcement learning is an example of semi-supervised learning technique and is used to model sequential decision-making process
- RL is used for gaming & Robot learning, chess

Advantages of PRL

1. Maximize Performance
2. Sustain change for a long period of time
3. Too much reinforcement can lead to an overload of states which can diminish the results.

Advantages of NRL

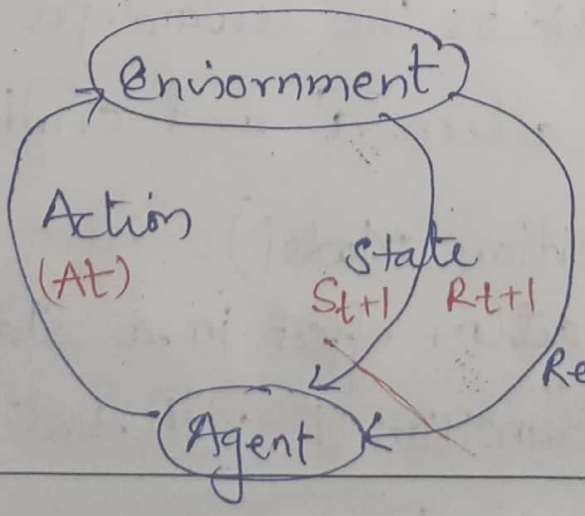
1. Increase Behaviour
2. Provide defiance to a minimum standard of performance
3. It only provides enough to meet up the min behaviour.

eg of RL:- 1) Self driving cars
2) Industry automation
3) finance
4) NLP
5) Healthcare

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Markov Decision Process (MDP)

- A Markov Decision Process (MDP) is defined as a stochastic decision-making process that uses a mathematical framework to model the decision-making of a dynamic system in scenario where the result are either random or controlled by a decision maker,
- MDP is used to formalize the reinforcement learning problems. If the enviorement is completely observable, then its dynamic can be modeled as a Markov Process.
- In MDP, agent constantly interacts with the enviornment and perform actions; at each action, the enviornment respond and generate a new state



The MDP framework has the following key

1) State(S)
2) Action (A)
3) Reward (R_a)
4) Policy (P_a) $\pi: S \to A$ {

1) It is a set of tokens that represent every state that the agent can be in.

2) Action(A) is set of all possible actions. A(s) defines the set of actions that can be taken being in state(s).

3) A Policy (P_a) is a solun to the MDP. A policy is a mapping from s to a.

4) Reward (R_s) indicates the reward for simple being in the state S.

R(S,a) indicates the reward for being in a state S and taking an action 'a'.

R(S,a,S') indicates the reward for being in a state S, taking an action 'a' and ending up in state S'.

Model: - (Transition Model)

It gives an action effect in a state

T(S,a,S') → transition T with state S ⟶ action a ⟶ Give state S'

may be same/or may not be

P(S'|S,a) → Prob. of reaching a state S' if action a is taken in state S.

eg: State: S

Model: $T(S, a, S') \sim P(S'|S, a)$

Action : $A(S)$, A

Reward: $R(S)$, $R(S, a)$, $R(S, a, S')$

Policy : $\pi(S) \rightarrow a$

π^*

MDP uses Markov Property :-

Markov Property :-

If the agent is present in current state - S_1,
- perform action a_1 &
- move to state S_2,
- Then the state transition from $S_1 \rightarrow S_2$ only depends on the current state and future action.
- It do not depend on past action, reward or states.
- We represent agent using Markov State.
- State S_t is Markov state if it follow the given condition

$$P[S_{t+1} | S_t] = P[S_{t+1} | S_1, S_2, S_3 \ldots S_t]$$

prob. of next step — previous state — present state

- Markov state follow Markov property which that "future is independent of the past and can only be applied with the present".

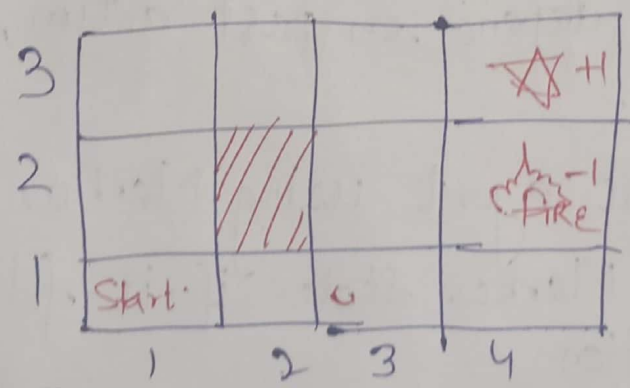
- RL work on fully obseved enviornments where the agent can observe the enviorment and act for new state → This whole procers is known as Markov Decision procers.

eg: Chess game → players only focus on the current state and do not need to remember past actions or states.

Markov Procers :→

MP is a memoryless procers with a seq. of random states S1, S2, ... St that uses the Markov Property. Markov procers is also known as Markov Chain, which is a tuple (S, P) on state S and policy-fen P.

eg:-



- 3x4 Grid
- start state on geid (1, 1)
- Purpose to reach at (4, 3)
- Agent can move up, down, Left, right
- Avoid fire geid (4,2)
- Avoid (2,2) Wall.

Aim:- find shortest sequence.

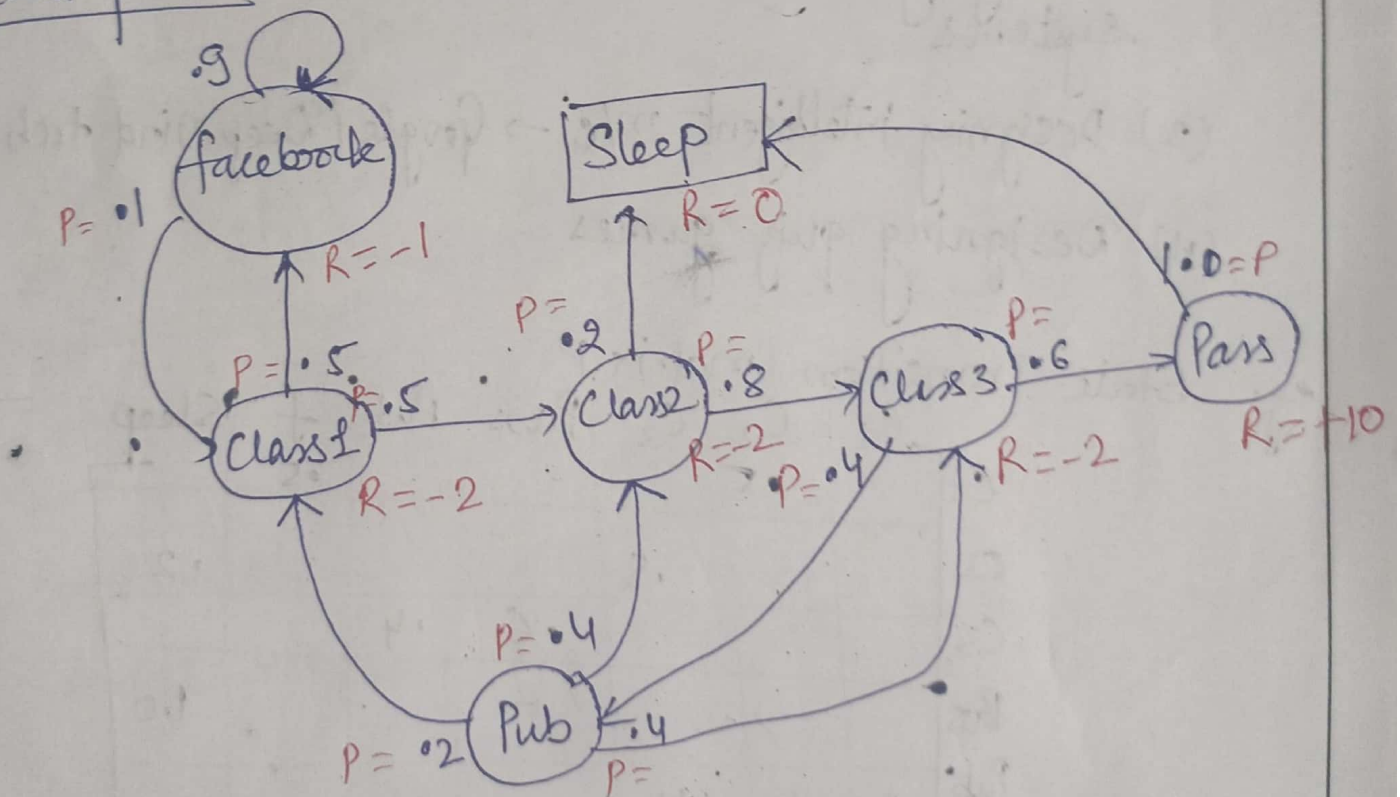
Solu 1:- R, R, U, U, R .1+.1+.8+.8+1 = 2.8

Solu 2 : U, U, R, R, R .8+.8+.1+.1+1 = 2.8

Prob: of UP = 0.8, left = 0.1, Right = 0.1

So, two steps are there to reach till the final destination

Example 2:- Student Markov chain



Before solving this we should know about discount factor(γ) & policy.

$$V(s) = E\left[\sum_{t=0}^{\infty} \gamma^t r_t \mid S_t\right]$$

↓ value identifies the reward fun r_t at each specific State

↓ or sum of discounted future rewards

MDP-based sequential decision making is used to (1) Travelling salesman Problem (TSP)

(2) Managing maintenance and repair of dynamic systems

(3) Designing intelligent m/c → Google / Deepmind tech.

(4) Designing quiz games

eg:. State Transition matrix P

	C1	C2	C3	Pass	Pub	f	Sleep
C1		.5				.5	
C2			.8				.2
C3				.6	.4		
Pass							1.0
Pub	.2	.4	.4				
f	.1					.9	
Sleep							1

- Sum of each row = 1

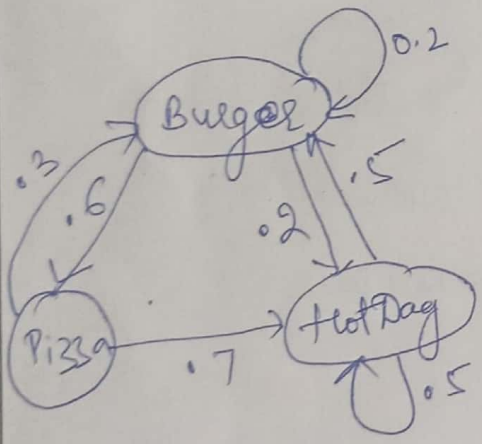
⇒ Suppose We are at class 1

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Starting with S1=C1

- C1-C2-C3-Pass-Sleep
- C1 → FB-FB-C1, C2-Sleep
- C1, C2, C3, Pub C2 C3 Pass Sleep
- E1, FB, FB, C1 C2 C3 Pub C1 FB FB FB
- C1, C2, Sleep

C1, C2, C3, Pub, C2, Sleep



If today Burger day there may be a propability that tomorrow they serve
Pizza → with .6 prob
Burger → .2
HotDog → .2

Random Walk

Burger → Pizza →

$$A = \begin{array}{c} \\ B \\ P \\ H \end{array} \begin{array}{c} \begin{array}{ccc} B & P & H \end{array} \\ \begin{bmatrix} .2 & .6 & .2 \\ .3 & 0 & .7 \\ .5 & 0 & .5 \end{bmatrix} \end{array}$$

Let take $\pi = [0 \ 1 \ 0]$ next is :

↙ Row vector

Prob distribution of states

$$\pi_0 A = [0 \ 1 \ 0] \begin{bmatrix} .2 & .6 & .2 \\ .3 & 0 & .7 \\ .5 & 0 & .5 \end{bmatrix}$$

$$= [.3 \ 0 \ .7] = \pi_1$$

$\pi_1 A = [.3 \; 0.7] \begin{bmatrix} A \end{bmatrix} = [.41 \; .18 \; .41]$ — π_2

$\pi_2 A = [.41 \; .18 \; .41] [A] = [.31 \; .25 \; .11]$

$\boxed{\pi A = \pi}$ → linear algebra $Av = \lambda v$, $\lambda = 1$, $v \cdot A = v$

Prob Distribution → $\pi[1] + \pi[2] + \pi[3] = 1$ should be

When we will solve these equation

$$\pi = \left[\frac{23}{71} \quad \frac{15}{71} \quad \frac{31}{71}\right] = [.35 \; .21 \; .43]$$

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Bellman Equation RL :-

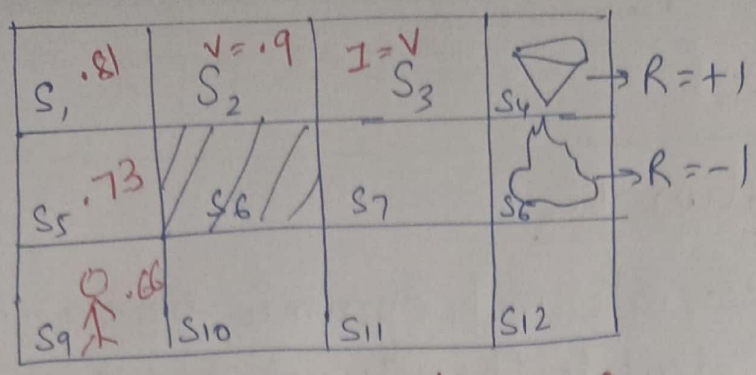
- This algo is associate with dynamic programming and used to calculate the value of a decision problem at a certain point by including the values of previous states.
- Key elements used in Bellman eq:
  1) Action performed by the agent is referred as 'a'
  2) State occured by performing action 's'
  3) Reward/feedback "R"
  4) discount factor Gamma 'γ'

Bellman eq:

$$V(s) = \max [R(s,a) + \gamma V(s')]$$

- $V(s)$: Value calculated at particular state
- $R(s,a)$: Reward of (s, a)
- γ : Discount γ [0 - 1]
- $V(s')$: value in the previous state

- In this equation, we are taking the max of the complete values because the agent tries to find the optimal solution always.



Agent = S_9 Readh/Goal = S_4

So - we have to readh at S_3 → find its value

We have $V(S3) = \max[R(S,a) + \gamma V(s')]$

Let $\gamma = .9$ $V(s') = 0$ → because ~~it is a intial state~~ there is no further state to move

$V(S3) = \max[1 + .9 \times 0]$

$= 1$

$V(S2) = \max[\text{Reward} = 0 + .9 \times 1]$

↓ no Reward

$= \max[0, +.9] = \boxed{.9} \rightarrow S_2$

$V(S_1) = \max[\text{Reward} = 0 + .9 \times .9]$

$= \max[0 + .81] = \boxed{.81} \rightarrow S_1$

$V(S_5) = \max[0 + .81 \times .9] = .73$

$V(S_9) = \max[0 + .73 \times .9] = 0.66$

$V[S_{10}] = \max[0 + .66 \times .9] = .59$

$[V S_{11}] = \max[0 + .9 \times .59] = .53$

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

$S_{12} = .48$

S_7 = previous value at S_{11} can't move left & Right

$S_7 = \max[0 + .9 \times 1] = .9$

Now for S_{11} see S_7 also

Calculate S_{11} by using S_7

$S_{11} = .81 / .53 \rightarrow .81$

$S_{12} = .73 / .48 \rightarrow .73$

$S_{10} = .73 / .59 \rightarrow .73$

$S_9 = .73 \times .9 = .657 = .66 / .66 = .66$

Let a Agent at S1 → it will check the next state value = .9 > .73 so it will choose S_2

– If it present at S_9 → went to $S_{10} \rightarrow S_{11}$

$S_7 = .9$

$S_{10} \leftarrow S_{11} \rightarrow S_{12}$

.73 .73

So it will move up.

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Policy Evaluation Using Monte Carlo :→

Monte Carlo policy evaluation is a technique within the field of reinforcement learning that estimates the effectiveness of a policy - a strategy for making decisions in an environment.

- Bit like learning the rules of a game by playing it many times, rather than studying its manual.

- Monte-Carlo to evaluate policy π only.

Means We have policy π and accordingly calculate
state-value fun^c ($V_\pi(s)$)
Action-value fun^c ($q_\pi(s)$)

- Monte carlo policy evaluation repeatedly simulates episodes, tracking the total rewards that follow each state and then calculating the average.

- So State value is nothing more than the discounted cumulative reward by starting from state s following a policy π to the end state at the last timestep T:

$$G_t = R_{t+1} + \gamma R_{t+2} + \cdots + \gamma^{T-1} R_T$$

$$G_i = \sum_{k=0}^{T} \gamma^k R_{t+k+1}$$

γ - Discount factor
R - Reward at each st

An example of an episode for a video game of an episode for a video game could be the sequence of all state, the frames, from start to game over.

- Let we are at s_i → initial state
  ↓ follow some policy π & reach at end state s_e.

at this process many times
episode we will get G_t
of these G_t → represents the real value for the initial state s_i.

- In Dynamic & Bellman eq → we calculate the exact value of the state s at timestep t by considering future reward and values of future states.
- But in monte carlo - we obtain the estimate of the true value func by running through many episodes.

$$V(s) = \frac{1}{N(s)} \sum_{i=1}^{N(s)} G_i$$

G_i → return value from ith episode

$N(s)$ → no. of time S is visited

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

There are two different approaches to evaluate the policy by using Monte-Carlo.

1) First Visit
2) Every Visit.

→ We have $N(s)$ → counter for each ~~vis~~ state s

$S(s)$ → sum of the different returns from each episode for each state.

Every visit Monte-Carlo Policy Evaluation

Iterate K episodes.

- It is possible that the same state is reached multiple times in the same episode. for all those visits the counter is increased.

$$N(s) = N(s) + 1$$

Also update $S(s)$ value by adding the return G_t

$$S(s) = S(s) + G_t$$

If all episodes

$$V(s) = \frac{S(s)}{N(s)}$$

First Visit Monte-Carlo Policy Evaluation

- In this we only update N(s) & C(s) only on the first visit of state s. Calculate every value like N(s), S(s) & V(s)

Problems of MonteCarlo:-

1) Work only for episodes not for continuous or infinite problems.
2) Need to complete an episode first before we can update all values.

* Policy Iteration & Value Iteration *

Bellman expectation eq. for state value under policy π.

$$V_\pi(s) = \sum_a \pi(a|s) \sum p(s', r | s, a)[R(s,a) + \gamma \cdot V_\pi(s')]$$

- $V_\pi(s)$: Value of state
- $\pi(a|s)$: policy of taking action a in state s
- $p(s', r | s, a)$: prob. of moving to next state s'

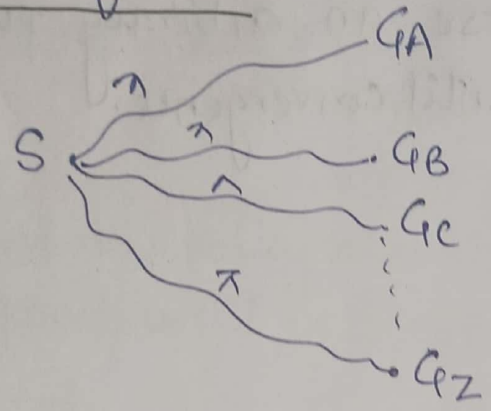
Recursive relation b/w current state's value & successive state value

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

In this equation adding all values so eq is updated.

$$V_{k+1}(s) = \sum_a \pi(a|s) \sum p(s', r|s, a)\left[R(s,a) + \gamma . V_k(s')\right]$$

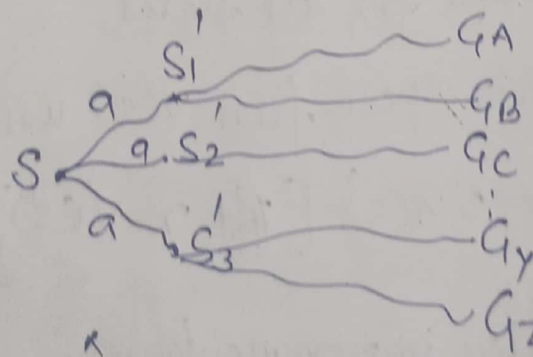
① Value of state



Total 26 discount episods

$$\therefore V_\pi(s) = \frac{1}{26} \sum (G_A + G_B + \cdots G_Z)$$

② Value of actions



$$Q_\pi(s,a) = E_\pi[G_t | S_t = S, A_t = a]$$

$$\hat{Q}_\pi(s,a) = \frac{1}{26} \sum (G_A + G_B + \cdots G_Z)$$

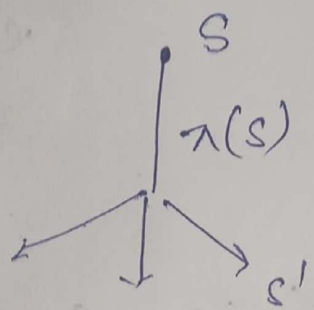
Policy Iteration:-

Policy Iteration consist of 2 steps

(1) Policy ~~impter~~ evaluation
Compute state value for all states
$\downarrow V_\pi(s)$

(2) Policy improvement
Improve the policy.
π_{new}

In this initialy we choose an arbitrary policy π and repeat 2 steps until convergence.



Steps:-

1) Set π as evaluate policy
2) Set arbitrary value for all states
3) Calculate the state space fun $V(s)$ using $\pi(s)$ policy

$$V(s) = \sum_{s'r'} p(s', r | s, \pi(s)) [R + \gamma \cdot V(s')]$$

(4) Calculate the policy improvement

$$\pi(s) \leftarrow \arg\max_a \sum_{s'r} P(s', r | s, a) [R + \gamma \cdot V(s')]$$

5) Repeat step 3 & 4 until convergence of the state values.

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

In other words, we don't care about the initial policy π_0 being optimal or not

1) Set random policy π_0

↓ evaluate it

2) find state value using π → V_{π_0}

↓ improve

3) find new policy and repeat until good (optimal) policy → π_1

↓ Evaluate it

V_{π_1}

↓ improve

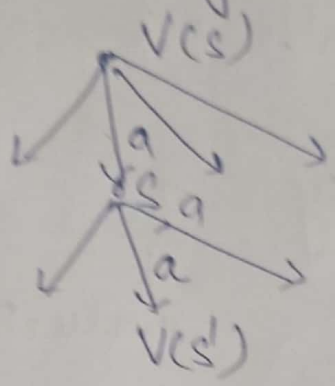
⋮

π_* (converges)

- Since a finite MDP has a finite no. of policies, the defined process is finite. In the end, converging an optimal policy π_* and an optimal value fun^c V_* is guaranteed.

Value Iteration :-

- In this compute the optimal state value func^c by iteratively updating the estimate V(s).



- We start with a random value fun V(s)

$$V(s) = \max_a \sum_{s',r} p(s', r \mid s, a)\,[R + \gamma V(s')]$$

- We look ahead one step and go over all possible actions at each iteration to find the max. value and update it.
- finally the value iteration algo is guaranteed to converge to the optimal values.

Policy Iteration	Value Iteration
1. Starts with random policy	1. Starts with a random value func^c.
2. Algo is more complex	2. Algo is simple
3. Guaranteed to converge	3. Guaranteed to converge
4. Cheaper to compute	4. More expensive to compute
5. Require few iteration to converge	5. Require more iteration to converge
6. Faster	6. Slower

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Model-Based Reinforcement Learning :→

- Let you are playing chess
- You made a move, then your opponent makes a move.
- After few move you realize that your 1st move was not corred.
- So next time, we think about the opponent next move and create a space tree in our mind and made a best move.
- Now if we replace ourself by AI-agent then you get Model-based Reinforcement Learning.

Model . MDP (S, A, P, R)

In model M_i [S, A, P_i, R_i]

- So model M_i going from state s → s' after performing action A, is subject to the probability $P_i(s'|s,A)$ have Reward $R^1(r'|s,A)$

- Model is anything the agent can use to predict how the environment will respond to its actions, concretely the state transition $T(s'|s,a)$ or reward fun $r(s,a)$
- We have 2 type of models

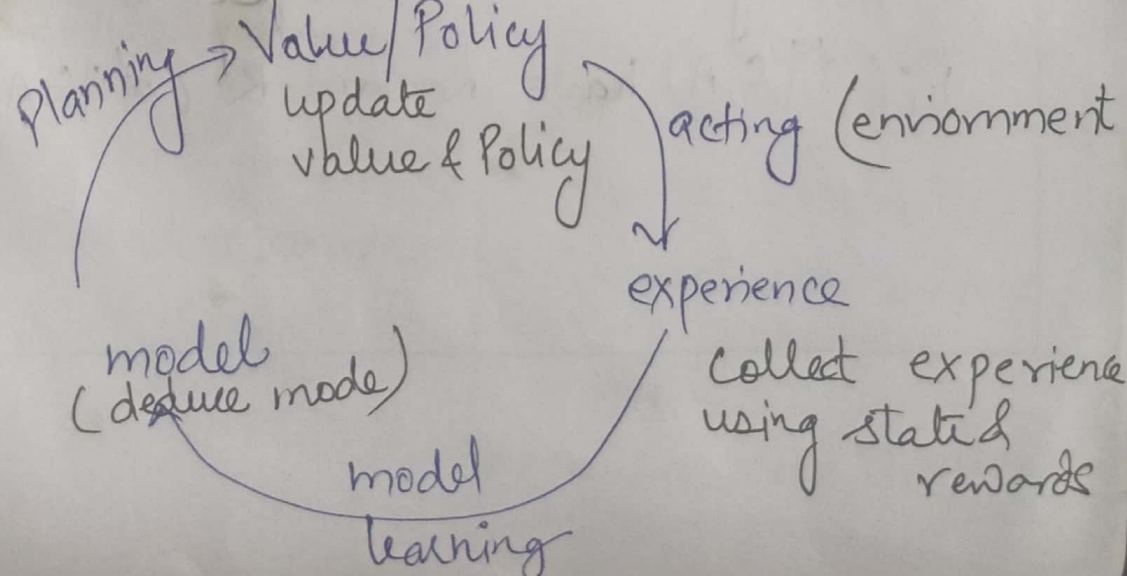
1) Model-free RL

- We perform sampling and simulation to estimate rewards.
- They perform actions either in the real world or in computer and reward — +ve / -ve

2) Model Based RL

- If we define a cost fun^c ourselves, we can calculate the optimal actions using the model directly.
- It reduce no. of iteration with the real environment during learning phase.

Main loop of Model-Based RL



Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

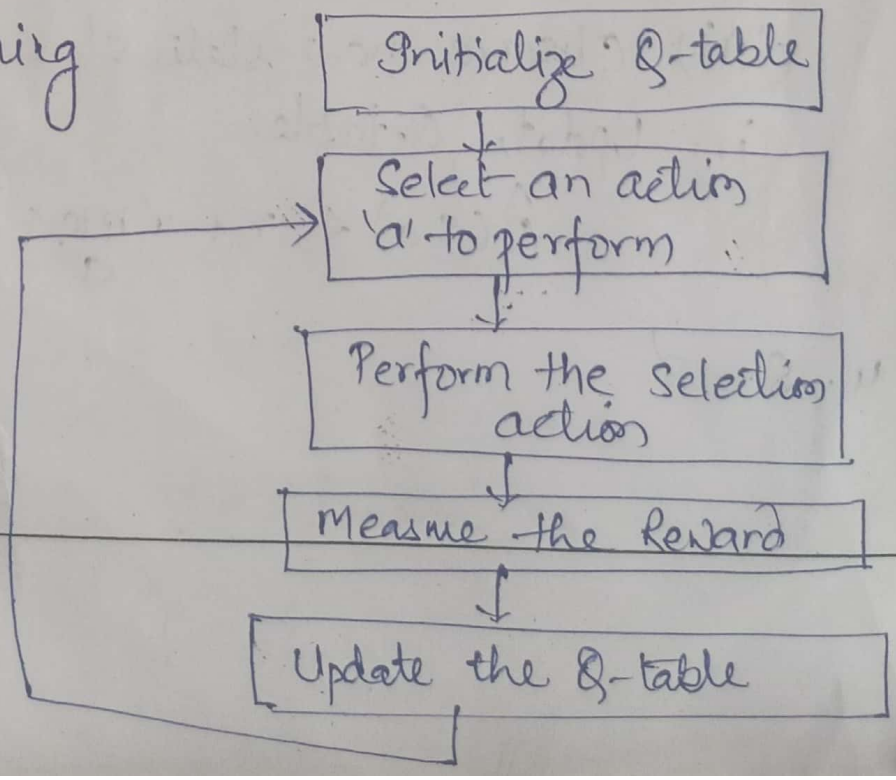
Reinforcement Learning Algorithms :-

- RL algo are basically used in AI application and gaming applicatio
- Main algorithms are
  1) Q-Learning
  2) SARSA (State Action Reward state Action)

Q-Learning :→ (Popular model-free RL Algo based on Bellma

- It is an ~~off~~ policy RL algo, which is used for the eqs temporal difference learning. The Temporal difference learning methods are the way of computing temporally sucessive predictions.
- It learn value fun^c Q(S, a) which means how good to take action 'a' at a particular state 's'.

- Flow of Q-Learning



- The main objective of Q-learning is to le the policy which can inform the agent th what actions should be taken for maxim the reward under what circumstances.

- Goal of agent in Q-learning is to maximize th value of Q.

- Value of Q-learning can be derived from the Bellman Eq:

$$V(s) = \max\left[R(s,a) + \gamma \cdot \sum_{s_t} P(s,a,s') V(\right.$$

Q learning Algo: →

1. for each s, a initialize the Q table $\hat{Q}(s,a)$ to 0
2. Observe current state s
3. Do while
   (i) Select action a & execute
   (ii) Recieve immediate award r
   (iii) Observe new state s'
   (iv) Update Q-table

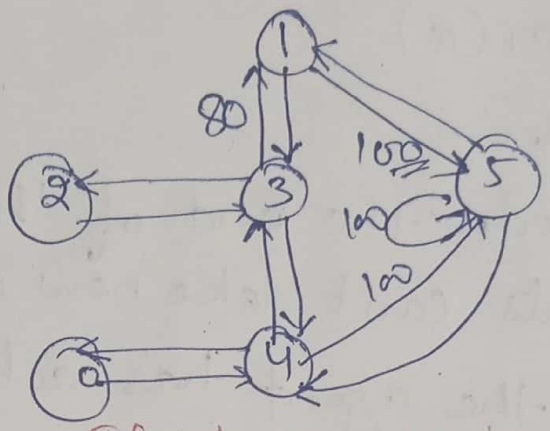
$$\hat{Q}(s,a) \leftarrow r + \gamma \max_{a'} \hat{Q}'(s', a')$$

4. $s = s'$

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

eg:- Let $r = .8$

$S = \{0, 1, 2, 3, 4, 5\}$



If we are moving 0 to 4 action = 4

If we are moving 4 to 0 action = 0

Let we are at

① $S = 3$

Can perfom action $\{1, 2, 4\}$

{means move to 1, 2, 4}

Let move to State = 1

$r = 0$

$\hat{Q}(3,1) = 0 + .8 \max \underline{\hat{Q}(1, \{3, 5\})}$

if move to 3 Reward = 0

if move to 5 Reward = 100

$= 0 + .8 \times \max\{0, 100\}$

$= 0 + .8 \times 100$

$= 80$

② $S = 1$. can perfom action $\{3, 5\}$

Move to state = 5 $r = 100$

move 1 & 4 and reward = 0

$Q(1,5) = 100 + .8 \max \hat{Q}(0,0)$

$= 100 + .8 \times 0$

$= 100$

③ S = Reach the End state so we will stop here.

Key Terminologies in Q-learning →

1. States (s)
2. Action (a)
3. Reward.-
4. Episodes: the end of the stage, where agents can't take new action. It happens when the agent has achieved the goal or failed.
5. $Q(S_{t+1}, a)$ → expected optimal Q-value of doing the action in a particular state
6. $Q(S_t, A_t)$ → current estimation $Q(S_{t+1}, a)$
7. Q-Table
   - Agent maintain the Q-table of sets of states and actions
8. Temporal Difference (TD): used to estimate the expected value of $Q(S_{t+1}, a)$ by using current state and action and previous state & action.

Q-function

$$Q(S, a) = R(S, a) + \gamma \sum_{s'} P(S, a, s') V(s')$$

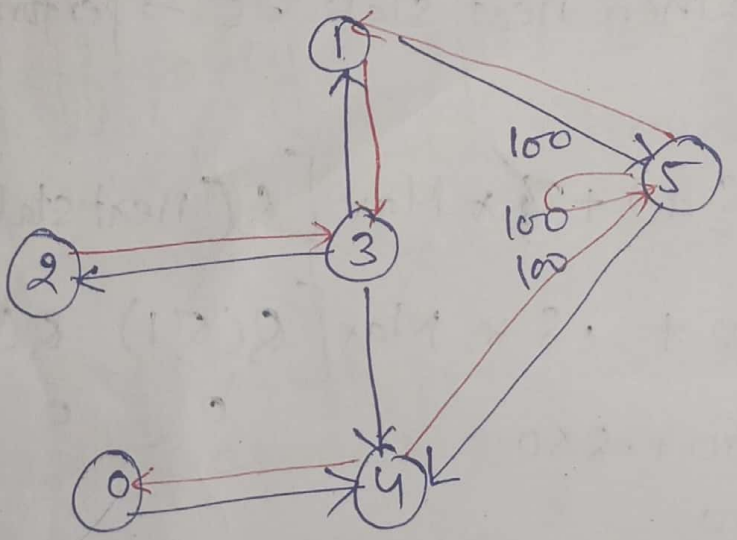
$$V(s) = \max [Q(S, a)]$$

$$Q(S, a) = R(S, a) + \gamma \cdot \sum_{s'} P(S, a, s') \max Q(S', a')$$

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

- Q represent quality of action at each state
- So instead of writing value, we are using pair of state & action Q(s_i, a_i)
- A Q table or matrix is created while performing the Q-learning. It contain state & action pair and intialize the values to zero.

Eg :→



Reward Matrix R → Action

State	0	1	2	3	4	5
0	-1	-1	-1	-1	0	-1
1	-	-	-	0	-	100
2	-	-	-	0	-	-
3	-	0	0	-	0	-
4	0	-	-	0	-	100
5						

3 5

(1, 5) 5

= 1, 5, 4

(3, 1) +

(2, 1, 5)

(1, 3, 5)

$$Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 80 & 0 \\ 0 & 0 & 0 & 64 & 0 & 100 \\ 0 & 0 & 0 & 64 & 0 & 0 \\ 0 & 80 & 50 & 0 & 80 & 0 \\ 64 & 0 & 64 & 64 & 0 & 100 \\ 0 & 80 & 0 & 0 & 80 & 100 \end{bmatrix} \end{array}$$

Let $\gamma = 0.8$

R=0 ↑ ↑ R=100

State = 1 → perform two action (3, 5)

Select Action = 5, next state = 5, Reward = 100

↓

When next state = 5 → perform action

1, 4, 5

$$Q(S,a) = R(S,a) + \gamma * \text{Max}\left[Q(\text{next state}, \text{all actions})\right]$$

$$Q(1,5) = 100 + .8 * \text{Max}\left[Q(5,1), Q(5,4), Q(5,5)\right]$$

$$= 100 + .8 \times 0$$

$$= 100$$

One episod finish

let 3 is the initial state

3 → (1, 2, 4) → next state → 1

state 1 → (3, 5)

$$Q(3,1) = R(3,1) + \gamma \cdot \text{Max}\left(Q(1,3), Q(1,5)\right)$$

$$= 0 + .8 \times \text{Max}(0, 100)$$

$$= 80$$

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

② SARSA (State Action Reward State Action)

- It's a slight variation of the popular Q-learning algo.
- In RL, it use 2 policy
  - ① On Policy → Agent learn the value fun" according to the current policy
  - ② off-policy → Agent learn the value fun" according to the action derived from another policy.
- Q-learning is off policy technique and uses the greedy approach to learn Q-value.
- SARS used on-policy technique and uses the current policy to learn the Q-value.

Q-L → $Q(S_t, a_t) = Q(S_t, a_t) + \alpha\left(r_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, a_t)\right)$

SARSA :-

$$Q(S_t, a_t) = Q(S_t, a_t) + \alpha\left(r_{t+1} + \gamma Q(S_{t+1}, a_{t+1}) - Q(S_t, a_t)\right)$$

- SARSA doesn't need a map of maze.
- Used for sequential decision-making problems.
- Q-L & SARSA belong to the Temporal difference (TD) learning methods

- SARSA

1) Update Q-value based on Current state, Action taken, reward received, next state and following action chosen using the policy.
2) On-policy algo
3) Lead to slower convergence if exploration is not managed correctly

Q-L

1) Update Q-value based on current state, action taken, reward received and max. Q-value of the following state over all possible actions.
2) Off-policy algo.
3) Converge to the optimal policy more efficiently in many cases.

Key features of SARSA

1) State (s)
2) Action (a)
3) Reward (R)
4. Next state (s')
5. Q-Table
6. Episodes
7. TD

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Unit 5:- Recommended System

1) Collobrative filtering
2) Content- Based filtering
3) Artificial neural n/w.
4) Preceptron, Multilayer n/w, Back Propogation,
5) Introduction to Deep learning:

Recommendation System:→

A machine learning algorithm known as a recommendation system combines info. about users and products to forecast a user's potential interests.

OR

A recommendation system is an AI algorithm, usually associated with machine learning that uses Big Data to suggest or recommend additional products to consumers.

- These are based on various criteria

  like past purchases,
  Search history
  demographic Information

→ R.S are used by various large name companies like Google, Instagram, Spotify, Amazon, Netflix.

- Recommender systems are highly useful as they help users discover products and services they might not found their own.

eg: Netflix, Amazon, Spotify

- RS are often seen as a 'black box'

Types of Recommendation Systems

There are vast no. of recommender algorithms and techniques, broad categories are

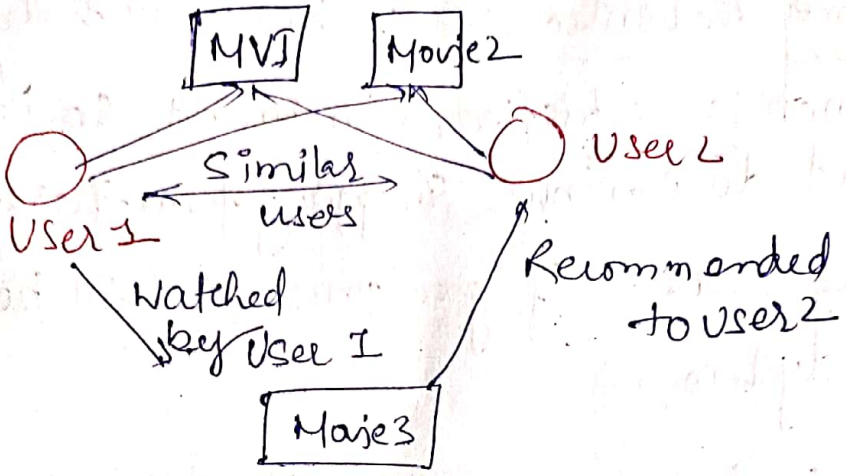
(1) Collobrative filtering
(2) Content filtering
(3) Context filtering

1) Collobrative filtering :-

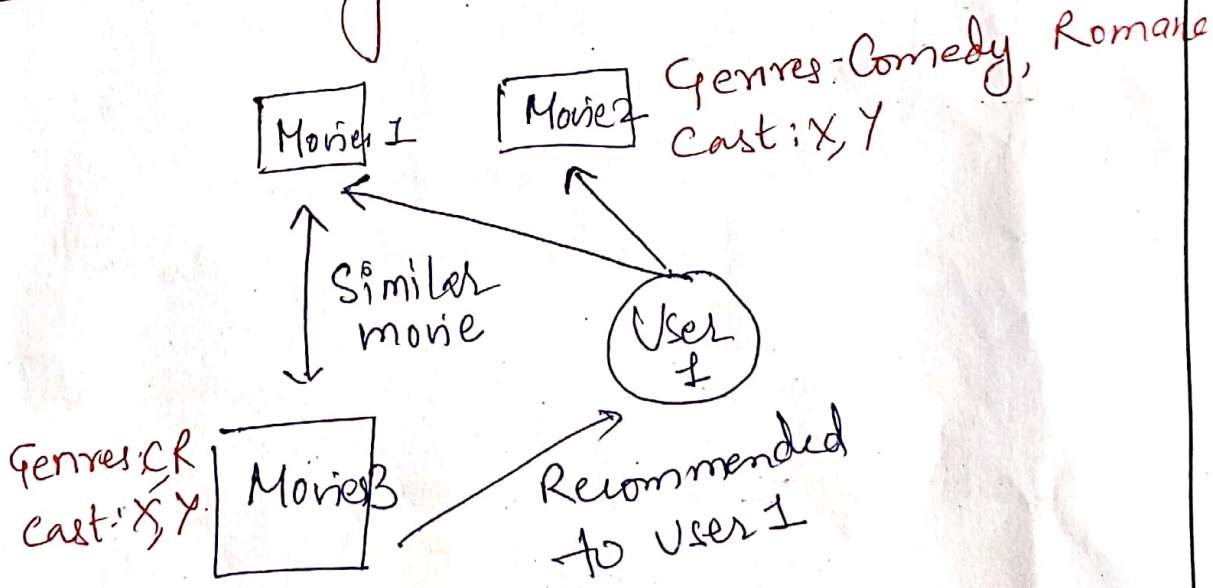
Recommender System build a model from a user's past behavior, such as item purchased previously or rating given to those items and similar decision by other users.

- Some pepople have made similar decision & purchase in the past, like a movie choice, then there is a high probability they will agree on additional future selection.

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in



(2) Content filtering :→



- It uses the attributes or features of an item to recommend other items similar to the user's preference.
- It work on similarity of item and user features, given info. about a user and items they have interacted with.

3) ~~Content filtering:~~ →

Benefits of Recommendation System:→

1. Improving Retention:- If brand is truly and customer is safisfied from brand, so he will be loyal and do continue Shopping from there.
2. Increasing Sales:- If god products will be recommended by system;
3. Helping to form customer habits & trends:-
4. Speeding up the space of work.
5. Boosting Cart value.

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

~~Policy Evaluation Using Monte Carlo~~

Collaborative filtering :→

- It is a ML technique used to identify relationship between piece of data.
- This technique is frequently used in recommender systems to identify similarities b/w user data and items.

User A User B

like

Product A

Product B like by User B

So. Probability that User A will like Product B.

Methodology :→

1. RS keep track of product user like & their characteristics.
2. Product feature in numerical form (value)
3. Two way to identify ~~eg~~ user like product or no
   (i) Ask Rating (numeric Rating)
   (ii) System assume by product purchase.

Eg:-

	Product 1	2	3	4
User 1		1	1	✓ Recom
User 2	1			
User 3		✓ Recom	1	1
User 4		✓ Recom	✓ Recom	1

3) Find the Si...

eg2

Customer	M_1	M_2	M_3	M_4	M_5 * Ignore	
	5 ✓	3 ✓	4 ✓	4 ✓	??	(4) = 5+3+4+4/4 = 16/4 = 4
2	3	1	2	3	3	(2.25) = 2.25
3	4	3	4	3	5	(3.5)
4	3	3	1	5	4	(3)
5	1	5	5	2	1	(3.25)

1. Calculate Row avg (Ignore missing colum from Dataset)

Rating avg of Customer 1 = $\frac{5+3+4+4}{4} = 4$

Customa 2 = $\frac{3+1+2+3}{4} = 1$

r_3 = 4+3+4+3/4 = 3.5

r_4 = 3+3+1+5/4 = 3

r_5 = 1+5+5+2/4 = 3.25

2. Now the rating will be 1, 2, 3, 4, 5

But if we see the last column → no body gives 2 rating so, we will eliminate (2)

Remains — 1, 3, 4, 5

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

3) find the similarity b/w Customer 1 & remaining

$$Sim(a,b) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| \times |\vec{b}|}$$

Range of sim is +1 to -1

Value near +1 are more similar

$$Sim(C_i, C_j) = \frac{\sum (r_{ip} - r_{iavg})(r_{jp} - r_{javg})}{\sqrt{\sum (r_{ip} - r_{iavg})^2} \cdot \sqrt{\sum (r_{jp} - r_{javg})^2}}$$

$$Sim(C_1, C_2) = \frac{(5-4)(3-2.25) + (3-4)(1-2.25) + (4-4)(2-2.25) + (4-4)(3-2.25)}{\sqrt{1^2 + (-1)^2 + (0)^2 + (0)^2} \cdot \sqrt{(.75)^2 + (1.25)^2 + (-.25)^2 + (.75)^2}}$$

$$= \frac{(1)(.75) + (-1)(-1.25) + 0 + 0}{\sqrt{1+1} \cdot \sqrt{\quad}}$$

$$= 0.85 \quad \text{max}$$

$$Sim(C_1, C_3) = 0.7$$

$$Sim(C_1, C_4) = 0$$

$$Sim(C_1, C_5) = 0.79$$

Similarly b/w (C_1 & C_2)

So Rating will be 3

There are two common type of approaches in collaborative filtering

1. Memory Based
2. model Based

① Memory Based :-

- Neighbourhood Collaborative filtering.
- This can be further split into user based collabarative f and item based cf.

User Based

Users are going to yield strong and similar recommendations

Item Based

Item based on the Similarity b/w items calculated using user rating of those items.

② Model-Based Approaches:-

- It is associated with features of dataset that are parameterized as inputs of the model to find optimizated solution.
- It include decision tree, rule based approaches.

Advantages :-

- It is simple
- It capture suitable characters only.

Disadvantage :-

- Not friendly

eg:→ Youtube, Coursera

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

② Content Based Systems / Filtering

- It generate recommendation based on the users preferences and profile. → features
- Content based models focus on the rating provided by the target uses themselves.
- It require following data sources

① Item level :- Require attributes of the item.

② User level data s :- User feedback based on the item you are providing recommendations.

Advantage :→

- Good if rated by more users.
- Easily scalable due to low amt of data

Disadvantage :→

4) It greatly depends greatly on previous known user
1) Never recommend uninteractive item interest
2) Ineffective for new users.
3) Model requires a fair amt of domain knowledge.

Example :

1) Amazon product feed
② Spotify music recommendations

- It is a ML technique that uses similarities in features to make decision.
- This technique often used in recommender systems which are algorithms designed to advertise or recommend things to users based on knowledge accumulated about the users.

Methodology :→

1. Comparing user's interests to product features
2. Most overlapping feature with user interests are what's recommended.

In content based filtering two methods are used.

(1) Users can be given a list of features out of which they can choose whatever they identify with the most.

(2) Algo can keep track of the products the user has chosen before and add those features to the user's data.

Eg:-

	f1	f2	f3	f4
Product1	1	—	1	1
P2	—	1	4	—
P3	3	—	—	1
User Interest	2	—	1	1

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

User Interest Level = $\sum_{i=1}^{q} P_i U_i$

$P_i \rightarrow$ product feature value

$U_i \rightarrow$ user interest value in column i

~~feature~~
User Interest level Product 1 = $1 \cdot 2 + 1 \cdot 1 + 1 \cdot 1 = 5$

Product 2 = $4 \times 1 = 4$

Product 3 = $3 \times 2 + 1 \times 1 = 7$

So, user is more interested in product 3.

Example 2 :-

User like Movie1: Batman V Superman (Adventure, Super hero)

Movie2: Guardian of the Galaxy (Comedy, Adventure, Super Hero, Sci-fi)

Movie3: Captain America Civil War (Comedy, Superhero)

Movie 4: Hitchhiker's guide to the galaxy (Comedy, Adventure, Sci-fi)

Movie5: Batman Begins (Superhero)

Movie 6: Spider man (Comedy, Super Hero)

Suppose

M1	2/10
M2	10/10
M3	8/10

Create User profile

Comedy	Adventure	Sup	
0	1	1	
1	1	1	
1	0	1	0

Movie Matrix

⇓ Weighted Genre Matrix

	Comedy	Adventure	S.H	Scifi
M1	0	2	2	0
M2	10	10	10	10
M3	8	0	8	0
User Profile	18 ③	12 ③	20 5	10 5 = Avg 60/4 = 15
Normalize	0.3	0.2	0.33	0.16

18
12
20
10
60/4 = 15

3/9 = .3

50/25 = .2

9+9+25+25 / 4

12

	Comedy	Advt	S.H	Scifi
M4	1	1	0	1
M5	0	0	1	0
M6	1	0	1	0
			0.33	0.16
User Profile	0.3	0.2		

M4 = 0.3 + 0.2 + 0.16 = 0.66 ✓ Recommended to User

M5 = 0.33

M6 = 0.3 + 0.33 = 0.63

Test movie → Lady Bird - Drama

↓

Never Recomde / Because not in feature set

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Inverse document frequency (IDF)

$$IDF_i = \log_e \frac{N}{N_i}$$

N → Total no. of docs.
N_i → no. of document

$$TF - IDF (\text{score } w_{ij}) = TF_{ij} * IDF_i$$

Example:- Let Document contain 10,000 terms and specific term apper 25 times in document

$$Tf = \frac{25}{10,000} = 0.0025$$

You have collection of 10,000 document and term ~~IDF~~ appear 500 of these document

$$Idf = \log \frac{10,000}{500} = 1.30$$

$$Tf \times IDf = 0.0025 \times 1.30$$

$$= 0.00325$$

Example:- Term 'cat' appear 25 times in a document which contain 1000 words

~~TF = 25/1000 = 0.025~~

- Collection of related document are 15,000 and 30 document contain cat IDF = $\log 15000/30 = 1.69$

$TF \times IDF = 0.025 \times 1.69 = 0.04225$

Swami Keshvanand Institute of Technology, Management & Gramothaan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

In content based filtering, first we create profile for each item. For this we perform TF (Term frequency) vectorizer. Means word is the no. of times it appears in the document. The IDF (Inverse Document frequency) of a word is the measure of how significant that term in the whole corpus.

TF-IDf Vectorizer:→

1) Term frequency

$$TF_{ij} = \frac{f_{ij}}{\max_k f_{kj}}$$

→ freq of term (feature) i in document j

Let Doc1: Ben studies about Computer in Compute lab

tf-idf weight is a weight often used in information retrieval & text mining.

Doc1	Ben	Studies	Computer	Lab
tf	1	1	2	1

vector [1 1 2 1]

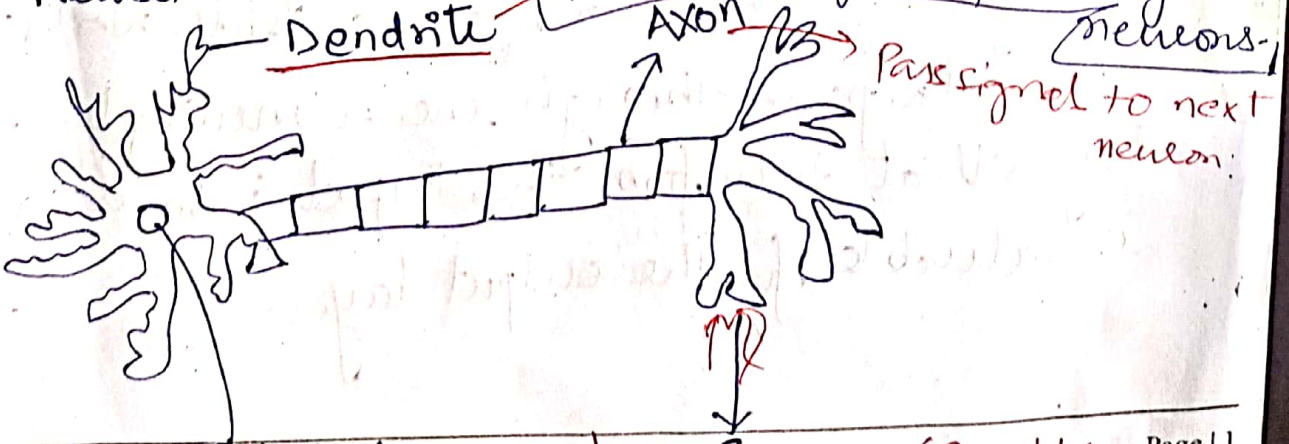
$\max_k f_{kj} = \|D\| \rightarrow$ Total no. of term in the document
$= 7$

normalized tf $= \left[\frac{1}{7} \frac{1}{7} \frac{2}{7} \frac{1}{7}\right] = [0.143 \quad 0.143 \quad 0.286 \quad 0.143]$

SKIT
असतो मा सद्गमय
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

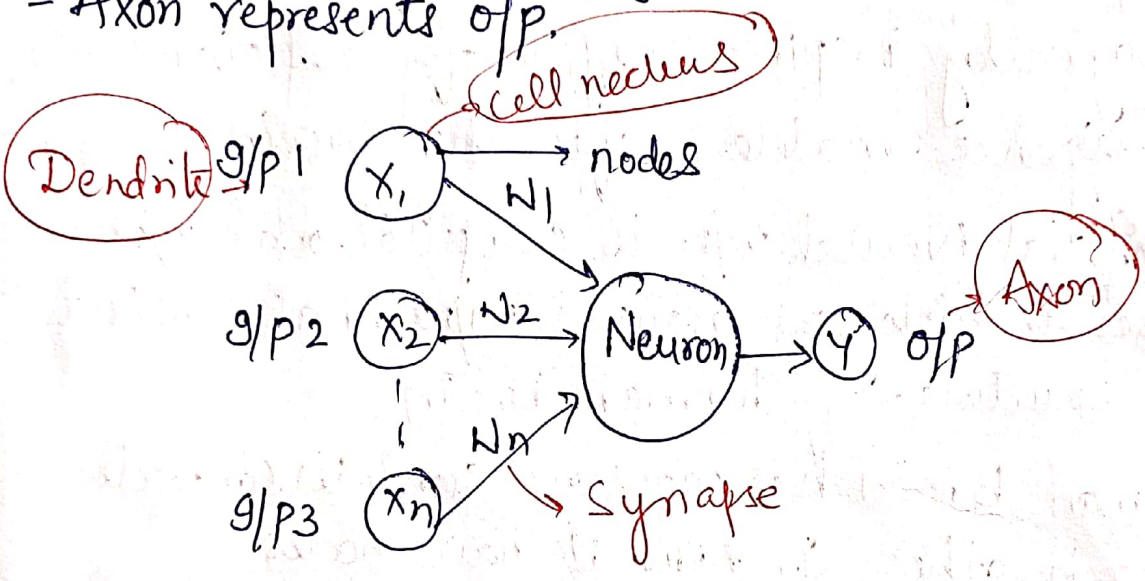
Artificial Neural Network :→

- Artificial Neural Network refers to a biologically inspired subfield of artificial intelligence modeled after the brain
- Artificial Neural n/w is compitational n/w based on biological neural n/w that construct the structure of heman brain
- Human brain has neurons or interconnected to each other, In ANN We have nodes.
- ANN term is derived from Bidgical neural network that develop the structure of a human brain.
- Given figure show the diagram of Biological Neural Network.



→ Main body of the neuron which accumulates the signals coming from the different dendrites

- Dendrites from Biological Neural Network represents inputs in Artificial Neural n/w.
- Cell nucleus represents Nodes,
- Synapse reprents weights
- Axon represents o/p.

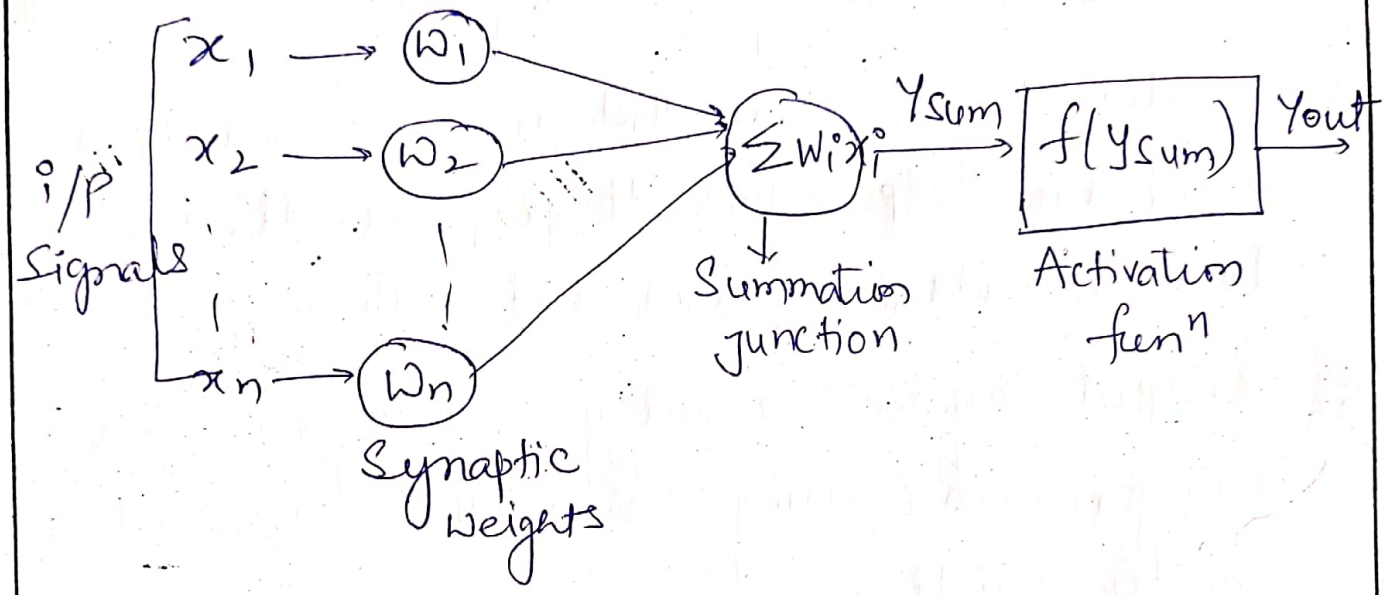


- ANN contain neurons which are called units.
- A layer can have only a dozen units or millions of units.
- ANN has an input layer, o/p layer & hidden layer.
- I/p layer receive data from outside world.
- This data passes through one or multiple hidden layers that transform the input into data that is valuable for the output layer.

Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

Structure of Artificial neuron



Input signal $x_i (x_1, x_2 \ldots x_n)$ comes to artificial neurons. Each neuron has three major components:

(i) Set of i- synapse having weight w_i.
Value of w_i may be positive or negative.
$w_i \rightarrow$ +ve value to increase likelihood to o/p
$w_i \rightarrow$ −ve value decrease likelihood to o/p

(ii) summation junction It is linear combiner or adder of the weighted input signals.

$$y_{sum} = \sum_{i=1}^{n} w_i x_i$$

(iii) Threshold activation fun: - results in o/p signal

$$y_{out} = f(y_{sum})$$

Or we can say that ANN has these three layers.

① Input layer:- It accept i/p in several different formate provided by the programmer

② Hidden layer:- The Hidden layer presents in b/w input and o/p layers. It performs all the calculations to find Hidden features and patterns.

③ Output layer:→ Input goes through a ~~set~~ series of transformations using the hidden layer, which finally result in o/p.

→ So for the competation of ANN

$$y_{sum} = \sum_{i=1}^{n} w_i \cdot x_i + b$$

where b is the bias, which adjust the i/p of the activation fun.

Working of ANN:-

- first inputs are given and multiplied by their corresponding weights.
- Weighted inputs are summarized inside the computing Unit

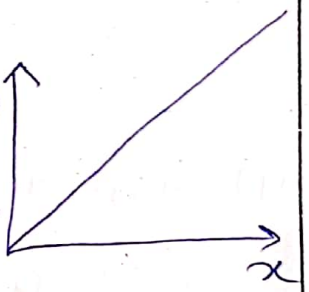
if weighted sum = 0

bias is added to make the output non-zero.

- Compare the output with the desired o/p, we change the value of weights and passes through the activation fun^c.
- Activation fun^c refers to set of transfer fun^n used to achieve the desired o/p.
- There is a different kind of activation fun^c.

Types of Activation fun^c :-

(1) Identity fun^c / Linear fun^n



It use linear fun^c. o/p remain same as i/p

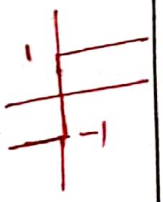
$$y_{out} = f(x) = x, \text{ for all } x.$$

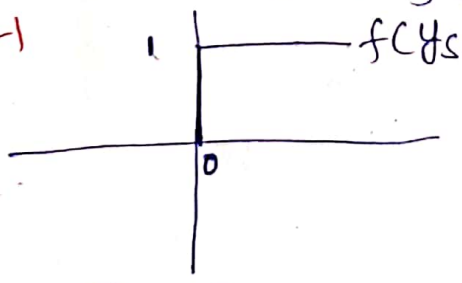
(2) Threshold / step fun^n

It gives 1 as o/p if the input is either 0 or positive. If the o/p is negative it give 0 as o/p.

$$f(x) = f(y_{sum}) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & x < 0 \end{cases} \quad \begin{matrix} \geq 0 \\ < 0 \end{matrix}$$

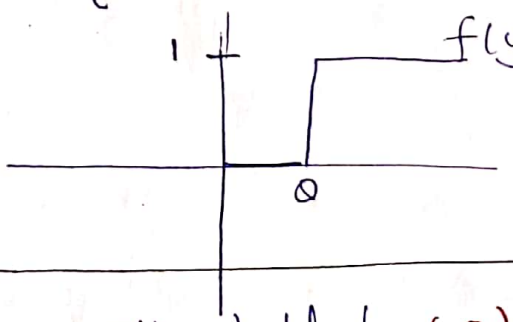
Bipolar → value 1 or -1





Step fun

→ mostly used in single layer nets

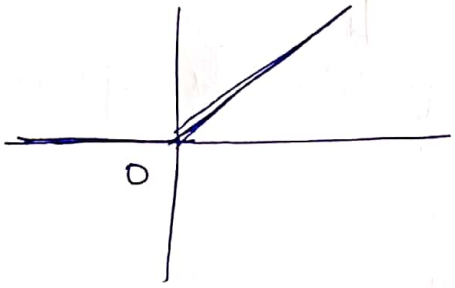


Threshold fun (0) → threshold value

(3) ReLU (Rectified Linear Unit) func

ReLU is most popular used activation fun in the area of convolutional NN and deep learning

$$f(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$$

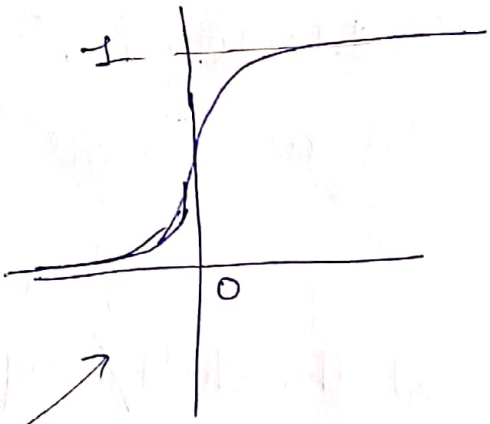


f(x) is zero when x is less than zero and f(x) equal to x above than zero.

(4) Sigmoid fun

- most commonly used fun in NN.

There are two type of sigmoid func.



(1) Binary sigmoid fun

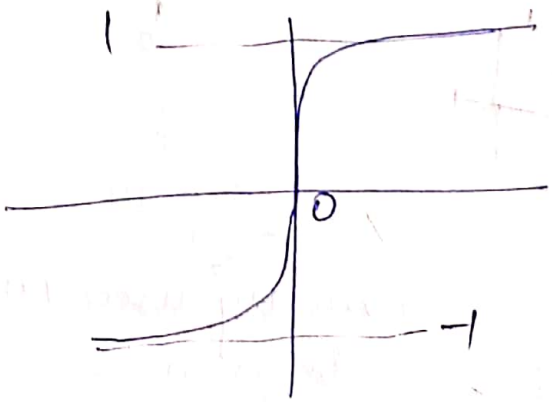
$$y_{out} = f(x) = \frac{1}{1+e^{-kx}}$$

range (0,1) ← k = Steepness or Slope parameter

if k value high, It work as threshold func

(2) Bipolar Sigmoid fun

$$y_{out} = f(x) = \frac{1+e^{-kx}}{1+e^{-kx}}$$



(5) Hyperpolic tangent func

- Continuous activation fun^c, bipolar in nature.
- It's used widely for back propagation n/w

$$Yout = f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}$$

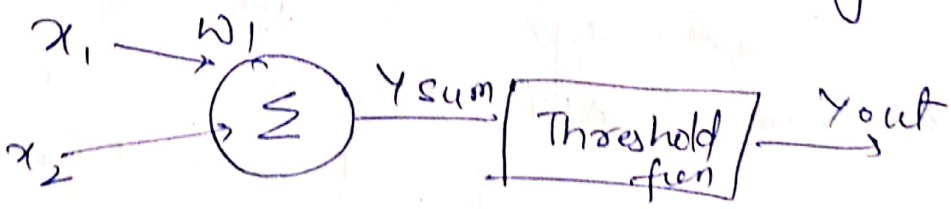
- Similar to bipolar sigmoid fun^n.

Type of Neural N/W: →

1) Perceptron: -

It is one of the simplest and oldest model of Neuron. It is smallest unit of neural n/w that does certain computation to detect features

- Perceptron is also known as TLU (Threshold logic Unit)
- Perceptron is a supervised learning algo that classifies the data into two categories.



Raining x_1	Sunny x_2	Yact	Yout
0	0	0	0
0	1	1	1
1	0	1	1
1	1	2	1

OR

Weights are associated

$w_0 = -2, \quad w_1 = \frac{1}{2} \quad w_3 = \frac{1}{4}$

$P_1 = (5, 2)$

$P_2 = (-1, 12)$

$P_3 = (3, -5)$

$P_4 = (-2, -1)$

$x_0 w_0 + x_1 w_1 + x_2 w_2 = 0$

$x_0 = 1$

$P_1 = -2 + 5 \times \frac{1}{2} + \frac{1}{4} \times 2 =$

$= \frac{-4+5+1}{2} = 1 - C1$

$P_2 = 0.5 - C1$

$P_3 = -1.75 - C2$

$P_4 = -3.25 - C2$

if $\hat{y}out \geq 0$ = class 1

< 0 = class 2

Perceptron Advantage

- It can implement logic gates like AND, OR, NAND
- It separate linearly pattern only.

Disadvantages

- It can does not work for non-linear problem such as boolean XOR problem

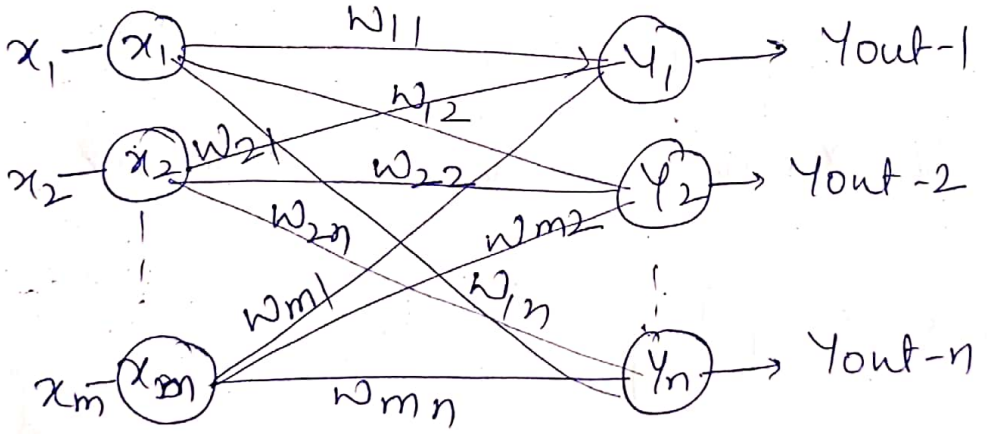
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

There exist five basic type of neuron connection architecture.

1) Single-layer feed forward n/w
2) Multilayer feed-forward n/w
3) Single node with its own feedback
4) Single-layer recurrent n/w
5) Multilayer recurrent n/w

(1) Single layer feed forward n/w

- Simplest & most basic architecture of ANN.



- It consist of only two layer (i/p layer & o/p layer)

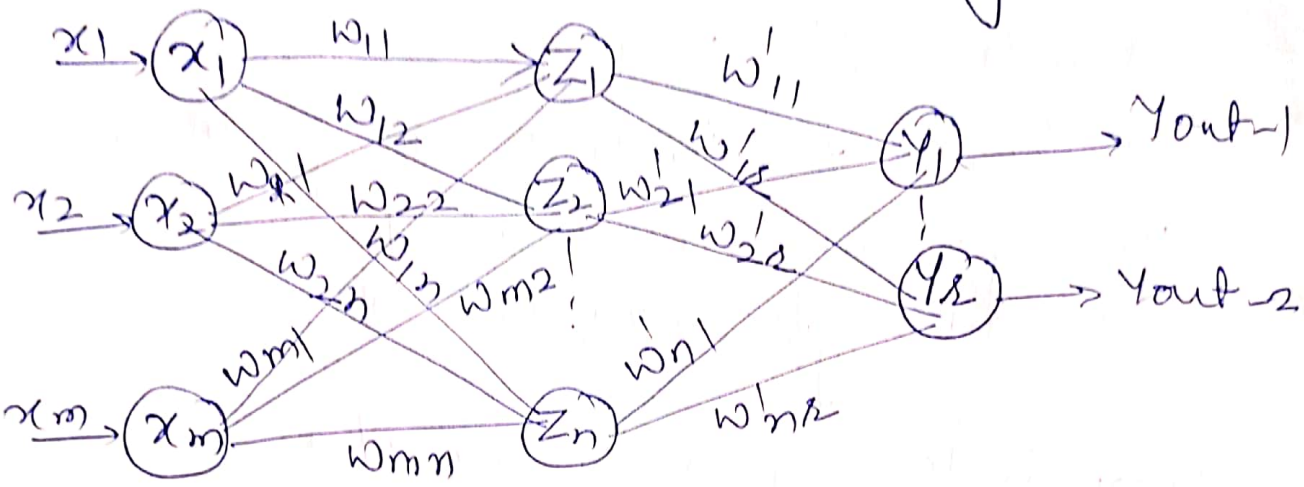
→ In this signal always flow from the i/p layer to o/p layer. Hence this n/w is known as feed forward

→ $y_{in-k} = x_1 w_{1k} + x_2 w_{2k} + \cdots + x_m w_{mk} = \sum_{i=1}^{m} w_{ik} x_i$

kth neuron

2) Multi-layer feed forward ANNs

- Quite similar to single layer feed forward n/w
- There are one or more intermediate layers of neurons b/w the input & o/p layer.



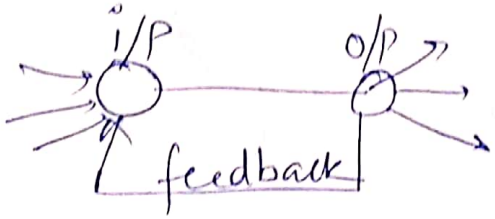
- i/p to hidden layer

$$Z_{in-k} = x_1 w_{1k} + x_2 w_{2k} + \cdots + x_m w_{mk} = \sum_{i=1}^{m} x_i w_{ik}$$

→ Signal at o/p layer

$$Y_{in-k} = Z_1 \cdot w'_{11} + Z_2 w'_{21} + \cdots + Z_n \cdot w'_{n2} = \sum_{i=1}^{n} Z_i \cdot w'_{ik}$$

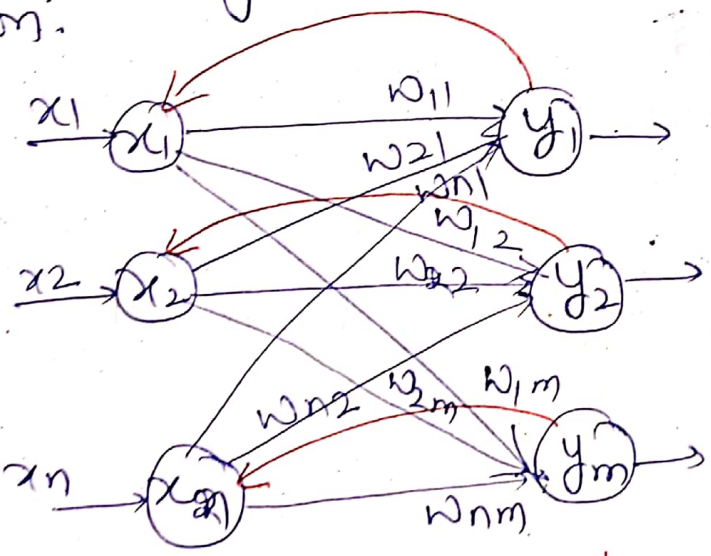
3) Single node with own feedback: →



Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

4) Single-layer Recurrent n/w :→

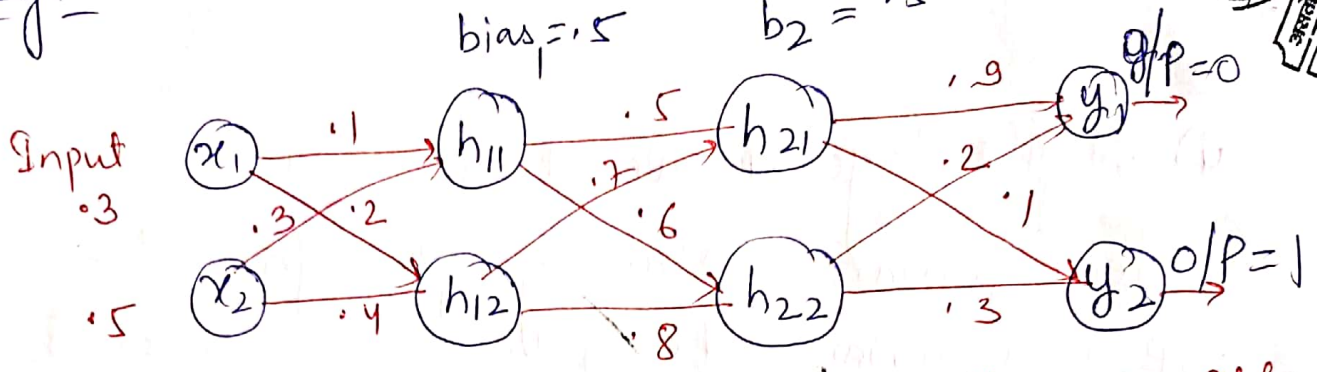
- Feed forward n/w always flow from the i/p layer towards the o/p layer (one direction)
- In case of Recurrent neural n/w, there is a small deviation
- There is a feedback loop from the neurons ~~in~~ output layer to the input layer neuron.



5) Multilayer Recurrent n/w :-

- In this have similar architecture
- but we have hidden layer in between and feedback from o/p layer to i/p layer.

Eg:-



$h_{11} = .1 \times .3 + .3 \times .5 + .5$

$= .03 + .15 + .5 = .68 = \theta$

$h_{12} = .3 \times .2 + .5 \times .4 + .5$

$= .06 + .2 + .5$

$= .76$

Here we are using sigmoid fun for activation

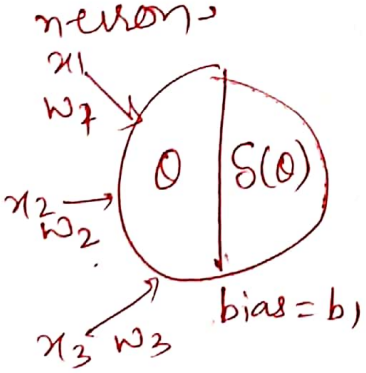
$S(\theta) = S(h_{11}) = \frac{1}{1+e^{-\theta}}$

$= \frac{1}{1+e^{-.68}}$

$= \boxed{0.66}$

$S(h_{12}) = \frac{1}{1+e^{-\theta}}$

$= \frac{1}{1+e^{-.76}} = \boxed{.68}$



Now calculate for next layer

$h_{21} = .66 \times .5 + .68 * .7 + .5 = 1.306$

$S(\theta) = \boxed{.786}$

$h_{22} = .66 \times .6 + .68 \times .8 + .5 = 1.504 = \boxed{0.818}$

Now Calculate for o/p layer

$y_1 = .786 * .9 + .818 * .2 = \cancel{0.1636}\ 0.871$ $S(\theta) = \boxed{\cancel{.54}\ .70}$

$y_2 = .786 * .1 + .818 \times .3 = .324$ $S(\theta) = \boxed{.58}$

Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

असतो मा सद्गमय

We will find Error

Mean Squred Error =

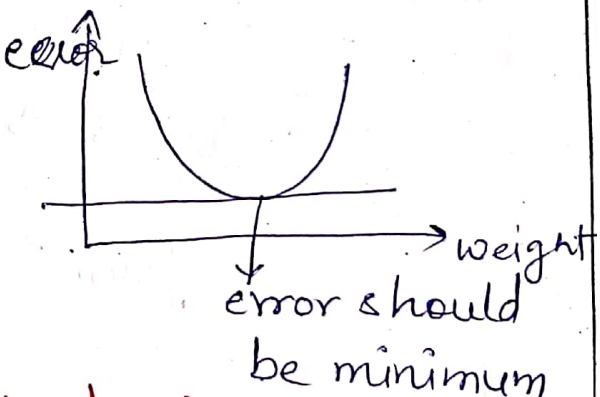
$$= \frac{1}{2}\left[(y'_P - y'_A)^2 + (y^2_P - y^2_A)^2\right]$$

$$= \frac{1}{2}\left[(.54 - 0)^2 + (.58 - 1)^2\right]$$

$$= \frac{1}{2}[.2916 + .1764] = .234 \; 0.33$$

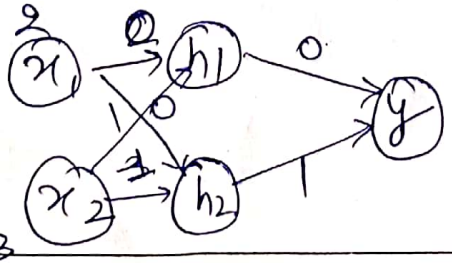
This is first epoch, we will repeat it until we did not get the reduced value of error

Q Why we need ff
- It approximate some fun^n f



Q = x = (2, 3)
w = [0, 1]
b = 0
] find its dog or cat if ≥0.5 → dog, <0.5 cat.

Af = Sigmoid, only one hidden layer.



$h_1 = 2 \cdot 0 + 3 \cdot 1 + 0 = 3 = 0$ & $\sigma(0) = 0.95$

$h_2 = 2 \cdot 0 + 3 \cdot 1 + 0 = 3 = .95$

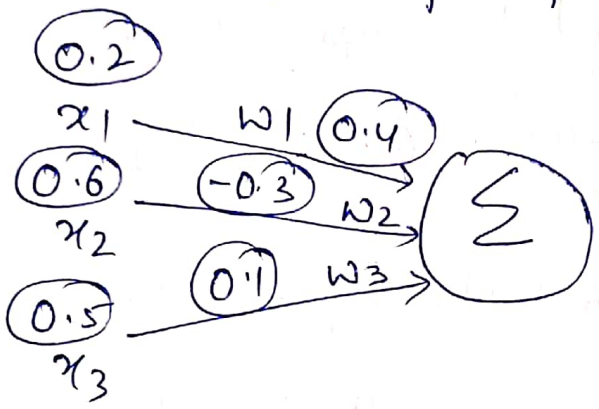
b = 0

$y = 0 \times .95 + 1 \times .95$

$= .95 \quad \& (.95) = .7216$

if ≥0.5 → Dog
<0.5 → cat

Q1: Consider a single perceptron with sign activation fun^n. The perceptron is represented by weight vector $[0.4 \quad -0.3 \quad 0.1]$ and a bias $\theta = 0$. If the input vector to the perceptron is $X = [0.2 \quad 0.6 \quad 0.5]$ then the o/p of the perceptron is



$y_{sum} = 0.2 \times 0.4 + 0.6 \times -0.3 + 0.5 \times 0.1$

$= .08 - .18 + .05$

$= .13 - .18 \Rightarrow -.05$

if $P > 0 \Rightarrow 1$

$P = 0 \Rightarrow 0$

$P < 0 \Rightarrow -1$

$y_{sum} \Rightarrow -0.05 < 0 \rightarrow o/p = -1$

SKIT
असतो मा सद्गमय
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Backpropagation in Machine Learning :→

- Backpropagation is an algorithm that backpropagates the error from the output nodes to the input nodes.
- It's uses in vast application of neural n/w like character Recognition, Signature verification etc.
- It is widely used algorithm for training feedforward NN.
- It computes the gradient of the loss fun^n with respect to the network weights.
- It is very efficient, rather than naively directly computing the gradient concerning each weight.
- Gradient descent or stochastic gradient are used to update the weights for minimize the loss.

features of Backpropagation :→

1) Gradient descent method is used.
2) Training is done in 3 stages
   (i) feed-fwd of input training pattern
   (ii) Calculation and back-propagation of error
   (iii) Updation of weight

Working of Backpropagation: -

1. Inputs X, through preconnected path.
2. Input is modeled using true weights W, Weights are usually chosen randomly.
3. Calculate o/p of each neuron from i/p layer to the hidden layer to o/p layer. $Z_{inj} = V_{oj} + \sum x_i W_{ij}$
4. Calculate error

$$\delta_k = (t_k - y_k) + y_{ink}$$

Backpropagation Error = Actual o/p - Desired o/p

5. From o/p layer, go back to the hidden layer to adjust the weights to reduce the error.
6. Repeat the process until the desired o/p is achieve

$X = \{x_1, x_2 \dots x_n\}$

t = target vector $\{t_1, t_2 \dots t_n\}$

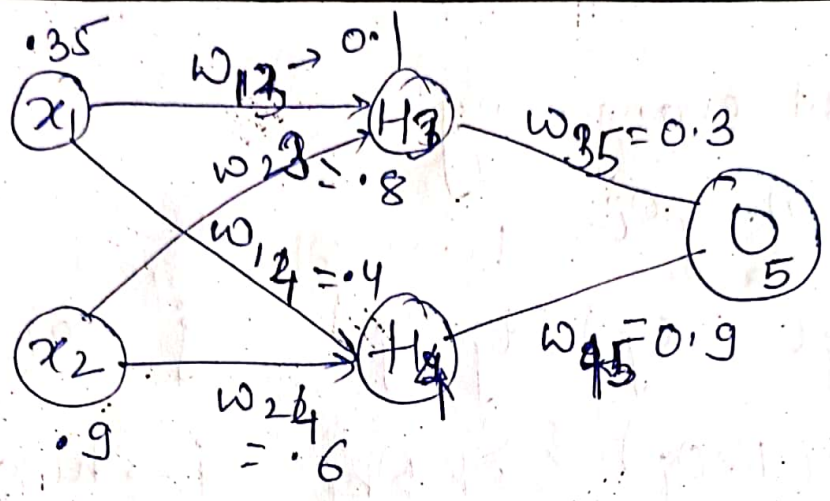
δ_k = error at o/p unit

δ_j = error at hidden layer

α - learning rate

V_{oj} = bias of hidden unit j

Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in



Assume that the neurons have a sigmoid Af, perform forward pass & backward pass. Actual o/p $y=0.5$ & learning rate = 1

$H_3 = x_1 w_{13} + w_{23} \cdot x_2$

$= .35 \times .1 + .8 \times .9 = 0.035 + .72 = 0.755$

$f(H_3) = \frac{1}{1+e^{-0.755}} = \frac{1}{1.47} = \boxed{.68}$

$H_4 = x_1 \cdot w_{14} + x_2 \cdot w_{24}$

$= .35 \times .4 + .9 \times .6 = .68$

$f(H_4) = \boxed{.6637}$

$O_5 = .68 \times .3 + .9 \times .66$

$= .801$

$f(O_5) = 0.69$ — predicted value

Error = $y_{target} - y_p$

$= 0.5 - 0.69 = -0.19$

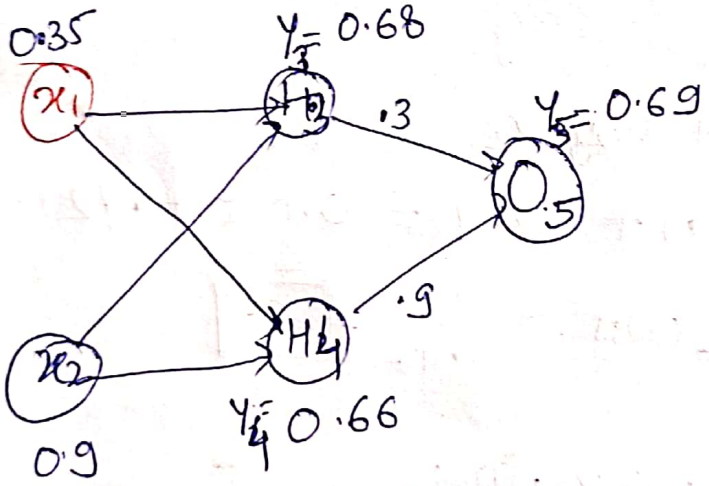
Each Weight changed by

$\Delta w_{ij} = \eta \cdot \delta_j o_i$ (η: learning rate, o_i: correct o/p for unit j)

$\delta_j = O_j(1-O_j)(t_j - O_j)$ $j \rightarrow$ o/p unit $\delta(h)$

$\delta_j = O_j(1-O_j)\sum_k \delta_k w_{kj}$ $j \rightarrow$ hidden unit (h)

↓ error measure for unit j



1) Calculate at O

$\delta_5 = y_5(1-y_5)(y_{target} - y_5)$

$= 0.69(1-0.69)(0.5-0.69) = -0.0406$

2) Hidden layer

$\delta_3 = y_3(1-y_3) \times \sum_k \delta_k w_{35} \delta_5$

$= .68(1-.68) * (0.3 * -0.0406) = -0.00265$

Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

$$\delta_4 = y_4(1-y_4) \cdot w_{45} \cdot \delta_5$$
$$= 0.66 \times (1-0.66) \cdot (0.9 \times -0.0406)$$
$$= -0.0082$$

$$\Delta w_{45} = \eta \cdot \delta_5 \cdot y_4$$
$$= 1 * -0.0406 * 0.66$$
$$= -0.0269$$

$$w_{45}(new) = \Delta w_{45} + w_{45}(old)$$
$$= -0.0269 + 0.9 = 0.8731$$

$$\Delta w_{14} = \eta \cdot \delta_4 \cdot x_1$$
$$= 1 * -0.0082 * 0.35$$
$$= -0.00287$$

$$w_{14}(new) = \Delta w_{14} + w_{14}(old)$$
$$= -0.00287 + 0.4 = 0.3971$$

i	j	w_{ij}	δ_j	x_i	W_{ij}(new)
1	3	0.1	-.00265	0.35	0.0991
2	3	0.8	-.00265	.9	0.7976
1	4	0.4	-0.0082	.35	0.3971
2	4	0.6	-0.0082	.9	0.5926
3	5	0.3	-0.0406	.68	0.2724
4	5	0.9	-0.0406	.6637	0.8731

Step2 (After Updated Weights)

$$y_3 = (w_{13} \times x_1) + (w_{23} \times x_2)$$
$$= (0.0991 \times 0.25) + (0.7976 \times 0.9)$$
$$= 0.7525$$

$$f(y_3) = 0.6797$$

$$f(y_4) = 0.6620, \quad y_4 = 0.6723$$

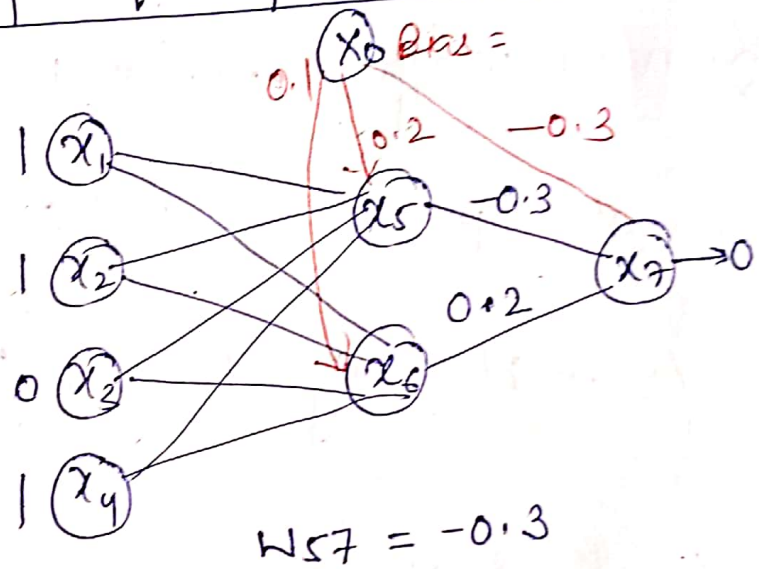
$$y_5 = 0.7631 \quad f(y_5) = 0.6820$$

$$\text{Error} = y_{target} - f(y_5)$$
$$= 0.5 - 0.6820 = -0.182$$

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Q MLP consist of I/P layer, One H/L, O/P layer
i/p layer - 4 value, 2 neuron at hidden layer,
Learning rate 0.8 and bias value =

x_1	x_2	x_3	x_4	$O_{Desired}$
1	1	0	1	1



$W_{15} = .3$
$W_{25} = -0.2$
$W_{35} = 0.2$
$W_{45} = 0.1$

$W_{16} = 0.1$
$W_{26} = 0.4$
$W_{36} = -0.3$
$W_{46} = 0.4$

$W_{57} = -0.3$
$W_{67} = 0.2$

$\theta_5 = 0.2$ $\theta_6 = 0.1$ $\theta_7 = 0.3$

Step 1 : Calculate o/p at x_5 & x_6

$x_5 = .3\times1 - 0.2\times1 + 0\times1\times0.1 + \theta = 0.2 = 0.4$

$O_5 = \frac{1}{1+e^{0.4}} = \boxed{0.599}$

$x_6 = .1\times1 + 0.4\times1 + 0\times0.4\times1 = +\theta(.1)$

$= 1.0$

$O_6 = \frac{1}{1+e^{-1}} = \boxed{.76}$

Step2 Calculate for x_7

$$x_7 = .599 \times -0.3 + 0.76 \times 0.2 + -0.3$$

$$= -0.326$$

$$O_7 = 0.419$$

Error = $Y_T - Y_P$

$$= 1 - 0.419 = 0.581$$

- We need propagation for weight (error reduce)

Epoche → 1

Error at O/P $= O_j(1-O_j)(t_j - O_j)$

Error at H/L $= O_j(1-O_j) \sum_k Error_k O_{kj}$

Step3 : Error at x_7

$$E_7 = 0.419(1-0.419)(1-0.419)$$

$$= 0.141$$

Step4: hidden layer (x_5, x_6)

$$E_6 = .76(1-.76) * 0.2 \times 0.141$$

$$= 0.005$$

$$E_5 = .599(1-.599) \times -0.3 \times 0.141$$

$$= -0.0101$$

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Steps:- Update weight using learning rate $\alpha = 0.8$

$$\Delta w_{ij} = \alpha \cdot E_j * O_i$$

$$\Delta w_{15} = 0.8 \times -0.01012 \times 1$$

$$= -.00808$$

$$w_{15}(new) = \Delta w_{15} + w_{15}(old)$$

$$= -0.00808 + .3 = \boxed{.292}$$

Same for

$$w_{16}(new) = w_{16} + \alpha \cdot E_6 \cdot O_1$$

$$= 0.1 + 0.8 \times .005 \times 1$$

$$\boxed{= 0.104}$$

$$w_{25}(new) = w_{25} + \alpha \cdot E_5 \cdot O_2$$

$$= -0.2 + 0.8 * -0.01012 \times 1$$

$$= -.208$$

$$w_{26}(new) = w_{26} + \alpha \cdot E_6 \cdot O_2$$

$$= 0.4 + 0.8 \times 0.005 \times 1$$

$$= .404$$

w_{35}	w_{36}	w_{45}	w_{46}	w_{57}	w_{67}
0.2	-0.3	0.092	0.404	-0.232	0.287

Step6:- Update Bias

$$\theta_j = \theta_j + \alpha \times Error_j$$

$$\theta_5 = \theta_5 + \alpha \cdot E_5$$
$$= 0.2 + 0.8 \times -0.0101 = 0.192$$

$$\theta_6 = \theta_6 + \alpha \cdot E_6$$
$$= 0.1 + 0.8 \times 0.005 = 0.104$$

$$\theta_7 = \theta_7 + \alpha \cdot E_7$$
$$= -0.3 + 0.8 * 0.41 = -0.187$$

Now we calculate bias as well as weights.

Now calculate the o/p for next ~~epochs~~ epochs

$x_5 = 0.368$ $O_5 = 0.591$

$x_6 = 0.204$ $O_6 = 0.7692$

$x_7 = -0.326$ $O_7 = 0.474$

Error $= Y_T - Y_P$

$= 1 - 0.474 =$ **0.526**

Now, when we compare the error, we get in the previous iteration and in current iteration

Error reduced $= 0.581 - 0.526 = 0.055$

- We will repeat the steps till we will not get the target o/p.

Swami Keshvanand Institute of Technology, Management & Gramothan,
Ramnagaria, Jagatpura, Jaipur-302017, INDIA
Approved by AICTE, Ministry of HRD, Government of India
Recognized by UGC under Section 2(f) of the UGC Act, 1956
Tel. : +91-0141- 5160400 Fax: +91-0141-2759555
E-mail: info@skit.ac.in Web: www.skit.ac.in

Need for Backpropagation :→

- It is useful for training neural networks.
- It is fast, easy to implement and simple.
- It's a flexible method because no prior knowledge of n/w is required.

Type of Backpropagation

1) Static backpropagation :→
Produce a mapping of a static input for static o/p. (OCR)

2) Recurrent Backpropagation :→
- In data mining fed fwd until a fixed value is achieved. After that, the error is compiled and propagated backward.

Advantages

1. Don't have parameter to tune except no. of i/ps
2. No need of n/w knowledge
3. User friendly, fast & easy
4. Don't need any special fun^n.

Disadvantages:

1. Sensitive to noisy data
2. Performance is highly dependent on input data

Deep Learning :-

- DL is a part of ML which is based on neural n/w.
- In DL, there is no need of explicit program.
- Basically, it is ML class that makes use of numerous non-linear processing units, so it perform feature extraction as well as transformation.
- In DL output from each preceding layer is taken as input by each one of the sucessive layer.
- DL & ML are subset of AI.

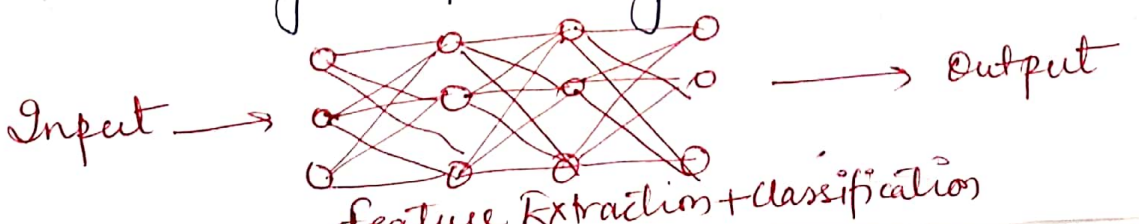
"The idea of DL to build such algorithm that can mimic the brain."

→ "Deep Learning is a collection of statistical techniques of machine learning for learning feature hierarchies that are actually based on artificial neural n/w."

Architecture of DL :→

① Deep Neural Network :→

- It is a neural n/w with a certain level of complexity (having multiple hidden layers in b/w input & o/p layers). They are capable of modeling and processing non-linear relationship



② Deep Belief N/N (DBN):—

- It is a class of Deep neural n/w.
- It is multi-layer belief n/w.
- Steps of DBN

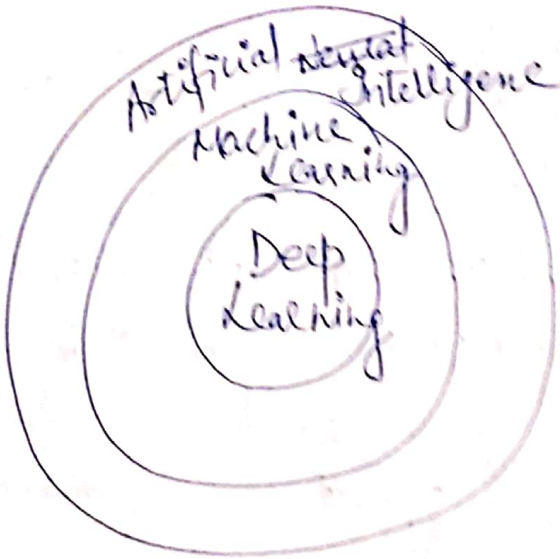
① Learn a layer of features from visible units using Contrastive Divergence algorithm

② Treat activations formerly trained features as visible units, which perform learning of features.

③ When the learning of Hidden layer done, then whole DBN is trained.

③ Recurrent Neural Network (RNN):→

- It allow for parallel and sequential computation.
- Due to large feedback n/w of connected neurons they remember important things about the input they received and enable them to be more precise.



Feed forward Neural Network :-

- In this nodes do'nt form a cycle.
- Also known as Artificial Neural network.
- Hidden layer connected with the previous and next layer only and are fully connected.
- Dont have visible or invisible connection b/w the nodes in same layer.
- Dont have back-loops

Application :-

- Data compression
- Computer Visim
- Handwritten characters Recognition

③ Convolutional Neural Network :-

- Special kind of neural n/w
- mainly used for image classification

Application :-

- Image Recognition
- Drug Discovery
- Identify faces, street signs

④ Restricted Boltzmann Machine :→

- In this neurons present in the input layer and the hidden layer encompasses symmetric connections
- No internal association within the respective layer.

Application :-

- filtering
- Risk Detection
- feature learning.

Difference b/w ML and DL

Machine Learning	Deep Learning
① Work on low small amount of Dataset	① Works on large amt of dataset for feature extraction
② Depend on Low-end m/c	② Depend on high-end m/c.
③ Divide the tasks into sub-tasks, solve them individually and finally combine the results.	③ Solve problem end to end.
④ Training time is less	④ Training time is large
⑤ Testing time increase	⑤ Testing time decrease

Types of Deep Learning networks? →

① RNN (Recurrent Neural Network):- Another version of feed-fwd n/w

- In RNN, hidden layer neurons receive input from the previous layer with a specific delay in time
- Due to this it has slow computational speed as well as it does not contemplate any future input for the current state.

Application ① Music composition
② Speech Recognition
③ Robot Control
④ Time-Series Prediction

Deep learning Applications :→

1. Self-driving cars
2. Voice-controlled Assistance
3. Automatic Image Caption Generation
4. Automatic Machine Translation

Limitations :-

1. Only learn through observation
2. Comprises of biases issues.

Advantage of DL :-

- Easily identifies difficult defects
- Eliminates unnesary costs
- Reduced cost for feature engineering

Disdvantages of DL :-

- Required large ant. of data
- Computationally expensive to train
- No strong theortical foundation