SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES (AUTONOMOUS)

QUESTION BANK

Year / Semester: II B.Tech IV Semester **Regulation: R23**

Subject and Code: Machine Learning 23CSM241

SYLLABUS

23CSM241	MACHINE LEARNING	L T P C 3 0 0 3

PRE-REQUISITES: Artificial Intelligence

COURSE EDUCATIONAL OBJECTIVES:

1. Define machine learning and its different types (supervised and unsupervised) and understand their applications.
2. Apply supervised learning algorithms including decision trees and k-nearest neighbors (k-NN).
3. Implement unsupervised learning techniques, such as K-means clustering.

UNIT I: INTRODUCTION TO MACHINE LEARNING (9)

Evolution of Machine Learning, Paradigms for ML, Learning by Rote, Learning by Induction, Reinforcement Learning, Types of Data, Matching, Stages in Machine Learning, Data Acquisition, Feature Engineering, Data Representation, Model Selection, Model Learning, Model Evaluation, Model Prediction, Search and Learning, Data Sets.

UNIT 2: NEAREST NEIGHBOUR-BASED MODELS (9)

Introduction to Proximity Measures, Distance Measures, Non-Metric Similarity Functions, Proximity Between Binary Patterns, Different Classification Algorithms Based on the Distance Measures ,K-Nearest Neighbor Classifier, Radius Distance Nearest Neighbor Algorithm, KNN Regression, Performance of Classifiers, Performance of Regression Algorithms.

UNIT 3: MODELS BASED ON DECISION TREES (9)

Decision Trees for Classification, Impurity Measures, Properties, Regression Based on Decision Trees, Bias–Variance Trade-off, Random Forests for Classification and Regression. The Bayes Classifier: Introduction to the Bayes Classifier, Bayes' Rule and Inference, The Bayes Classifier and its Optimality, Multi-Class Classification | Class Conditional Independence and Naive Bayes Classifier (NBC)

UNIT 4: LINEAR DISCRIMINANTS FOR MACHINE LEARNING (9)

Introduction to Linear Discriminants, Linear Discriminants for Classification, Perceptron Classifier, Perceptron Learning Algorithm, Support Vector Machines, Linearly Non-Separable Case, Non-linear SVM, Kernel Trick, Logistic Regression, Linear Regression, Multi-Layer Perceptrons (MLPs), Backpropagation for Training an MLP.

UNIT 5: CLUSTERING (9)

Introduction to Clustering, Partitioning of Data, Matrix Factorization |Clustering of Patterns,

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES (AUTONOMOUS)

QUESTION BANK

Year / Semester: II B.Tech IV Semester **Regulation: R23**

Subject and Code: Machine Learning 23CSM241

Divisive Clustering, Agglomerative Clustering, Partitional Clustering, K-Means Clustering, Soft Partitioning, Soft Clustering, Fuzzy C-Means Clustering, Rough Clustering, Rough K-Means Clustering Algorithm, Expectation Maximization-Based Clustering, Spectral Clustering.

Total Hours: 45

COURSE OUTCOMES:

On successful completion of the course, students will be able to		Pos
CO1	Identify machine learning techniques suitable for a given problem.	PO6, PO7, PO8, PO9, PO12
CO2	Solve real-world problems using various machine learning techniques.	PO6, PO7, PO8, PO9, PO12
CO3	Apply Dimensionality reduction techniques for data preprocessing.	PO6, PO7, PO8, PO9, PO12
CO4	Explain what is learning and why it is essential in the design of intelligentmachines.	PO6, PO7, PO8, PO9, PO12
CO5	Evaluate Advanced learning models for language, vision, speech, decisionmaking etc.	PO6, PO7, PO8, PO9, PO12

TEXT BOOKS:

1. "Machine Learning Theory and Practice", M N Murthy, V S Ananthanarayana, Universities Press (India), 2024

REFERENCE BOOKS:

2. "Machine Learning", Tom M. Mitchell, McGraw-Hill Publication, 2017.
3. "Machine Learning in Action",Peter Harrington, DreamTech
4. "Introduction to Data Mining", Pang-Ning Tan, Michel Stenbach, Vipin Kumar, 7th Edition, 2019.

CO-PO MAPPING:

CO\PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	-	-	-	-	-	2	2	3	2	-	-	3
CO2	-	-	-	-	-	2	2	3	2	-	-	3
CO3	-	-	-	-	-	2	2	3	2	-	-	3
CO4	-	-	-	-	-	2	2	3	2	-	-	3
CO5	-	-	-	-	-	2	2	3	2	-	-	3
CO6	-	-	-	-	-	2	2	3	2	-	-	3
CO*	**-**	**-**	**-**	**-**	**-**	**2**	**2**	**3**	**2**	**-**	**-**	**3**

Max Marks: 10

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: II B.Tech IV Semester **Regulation: R23**

Subject and Code: Machine Learning 23CSM241

S.No.	CO	Questions	BT
Unit I: (INTRODUCTION TO MACHINE LEARNING)			
1	1	a. What is machine learning? What are the goals of Machine Learning? b. Illustrate the applications and Challenges of Machine Learning.	L4
2	1	a. Explain the importance of vectors and matrices in machine learning. Provide examples of their usage. b. What is data preprocessing? Why is data preprocessing important?	L3
3	1	a. What is feature selection. b. Differentiate between Supervised, Unsupervised and reinforcement Learning.Explain the importance of vectors and matrices in machine learning. Provide examples of their usage.	L4
4	1	Describe dimensionality reduction in Machine Learning and list commonly used techniques.	L3
5	1	a. Describe over fitting in machine learning, and how does it affect model performance? b. What is model validation in machine learning and why is it important?	L5
6	1	Describe the different types of data used in Machine Learning. Explain how each type influences data preprocessing using appropriate examples.	L4
7	1	A dataset used for training a machine learning model contains missing values,outliers, andfeatures with significantly different value ranges. (a). Describe how you would handle missing values. (b). Explain how outliers can affect model performance. (c). Why is feature scaling important when dealing with variables of different ranges? Compare normalization and standardization with suitable use cases.	L3
8	1	Discuss the different paradigms of machine learning. Highlight their key characteristics, differences, and typical applications. Support your answer with appropriate examples.	L5
9	1	Illustrate Stages in Machine Learning in detail with figure	L4
10	1	Explain Paradigms for ML(Learning by Rote, Learning by Induction, Reinforcement Learning) with example	L3
11	1	Discuss about Types of Data, Matching.	L3

S.No.	CO	Questions	BT

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: II B.Tech IV Semester **Regulation: R23**

Subject and Code: Machine Learning 23CSM241

Unit II: (NEAREST NEIGHBOUR-BASED MODELS)			
1	2	Given the following two vectors A=(2,3,5) B=(4,1,7) (i). Calculate Euclidean distance between A and B, (ii). Calculate Manhattan distance between A and B, (iii). Calculate the cosine similarity and cosine distance between vectors A and B, (iv). Briefly discuss where cosine distance might be preferred over Euclidean distance in Machine Learning.	L4
2	2	Define a proximity measures in machine learning, and why are they important in machine learning.	L3
3	2	Name the performance metrics used to evaluate classification and regression algorithms and explain their purpose briefly.	L4
4	2	Apply K nearest neighbor classifier to predict the diabetic patient with the features BMI, Age.	L3
5	2	Given the two binary vectors: X=[1,0,1,1,0,0,1,0] Y=[1,1,0,1,0,1,1,1] (i). calculate the hamming distance between the two vectors. (ii). Interpret what the distance indicates in the context of similarity.	L5
6	2	Consider two data points: A=(2,4,5), B=(5,1,7) (i). Compute the minkowski distance between the two points using : a). Manhattan distance(p=1) . b). Euclidean distance(p=2), c).Minkowski diatance(p=3). (ii). Briefly explain how the value of 'p" affects the distance.	L4
7	2	Use the Radius Distance Nearest Neighbor algorithm to classify whether a patient is at health risk based on their vital statistics.	L3
8	2	Explain how K-Nearest Neighbors (KNN) Regression can be used to predict the price of a house based on features such as area (in sq. ft.), number of bedrooms, and age of the house (in years).	L5
9	2	Elaborate Distance Measures(Non-Metric Similarity Functions, Proximity Between Binary Patterns)	L4
10	2	Explain K-Nearest Neighbor Classifier with example.(Initially explain NN algorithm then KNN algorithm)	L3
11	2	What different ways to measure Performance of Classifiers and Performance of Regression Algorithms give example?	L4

S.No.	CO	Questions	BT

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: II B.Tech IV Semester **Regulation: R23**

Subject and Code: Machine Learning 23CSM241

Unit III: (Unit Name)			
1	**3**	**a. What is the bias-variance trade off in machine learning and why is it important in machine learning task? b. What is conditional independence in the context of a Naïve Bayes classifier, and why is it an important assumption?**	**L4**
2	**3**	**Explain decision tree algorithm with neat sketch.**	**L3**
3	**3**	**Explain Naïve Bayes algorithm with neat sketch.**	**L4**
4	**3**	**What are impurity measures in machine learning, why are they important in building decision tree?**	**L3**
5	**3**	**What is multi-class classification in machine learning, and how does it differ from binary classification? Provide one example.**	**L5**
6	**3**	**Explain Random Forest Model with neat sketch.**	**L4**
7	**3**	**Identify the first splitting attribute using Entropy and Information Gain by considering data set of your choice.**	**L3**
8	**3**	**What are the ten properties of Decision tree?**	**L5**
9	**3**	**Demonstrate Bayes rule and inference**	**L4**
10	**3**	**Illustrate Bayes classifier and its optimality**	**L3**

S.No.	**CO**	**Questions**	**BT**
Unit IV: (Unit Name)			
1	**4**	**a. What is a linear discriminant in machine learning, how is it used to separate data classes in a feature space? b. What is logistic regression in machine learning, and how is it used to model binary classification problem?**	**L4**
2	**4**	**Illustrate support vector machine(SVM) algorithm with neat diagram**	**L3**
3	**4**	**Consider the values of independent variable X and dependent value Y given below:** **X: 15 23 20 18 23 24 22 19 16 24 11 24 16 23** **Y: 49 63 58 60 58 61 60 63 52 62 30 59 49 68** **Compute the leat square regression line Y=a+Xb. Estimate the value of Y when X is 20.**	**L4**

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(AUTONOMOUS)

QUESTION BANK

Year / Semester: II B.Tech IV Semester **Regulation: R23**

Subject and Code: Machine Learning 23CSM241

4	4	What is the role of Kernel functions in support vector Machines and how do diffent types of kernel functions impact model performance? List any suitable real world examples for each.	L3
5	4	You are given a simple multilayer perceptron with one hidden layer that uses sigmoid activation functions. The input layer consists of two neurons, the hidden layer contains two neurons, and there is one output neuron. All weights are initialized with small values. (a). Explain the working principle of the back propagation algorithm used to train an MLP. (b). Discuss how learning rate and activation function choices affect the performance of the Back propagation algorithm.	L5
6	4	Discuss about Logistic Regression. Explain aboutBuilding Logistic Regression model briefly.	L4
7	4	Define the perceptron learning rule. How it is used for classification?	L3
8	4	Explain non-linear Support Vector Machine with example.	L5
9	4	Differentiate Logistic Regression and Linear Regression with suitable examples.	L4
10	4	Explain the working principle of Linear Discriminant with necessary equations.	L3
11	4	Explain the Perceptron learning algorithm with example.	L3

<table>
<tr><th>S.No.</th><th>CO</th><th>Questions</th><th>BT</th></tr>
<tr><td colspan="4">Unit V: (Unit Name)</td></tr>
<tr><td>1</td><td>5</td><td>a. What is Matrix Factorization, and how is it commonly used in recommendation systems?
b. What is Rough Clustering, and how does it handle uncertainty in the data compared to traditional clustering methods?</td><td>L4</td></tr>
<tr><td>2</td><td>5</td><td>Given a data set of the six objects characterized by two features. Use K-means clustering algorithm to divide the following data intotwo clusters and find the updated clustercentres after two iteration (Use Euclidean distance measure)
<table>
<tr><td>X</td><td>1</td><td>2</td><td>2</td><td>2</td><td>4</td><td>5</td></tr>
<tr><td>Y</td><td>1</td><td>1</td><td>3</td><td>3</td><td>3</td><td>5</td></tr>
</table>
Initial cluster centers are C1(2,1) and C2(2,3)</td><td>L3</td></tr>
</table>

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES (AUTONOMOUS)

QUESTION BANK

Year / Semester: II B.Tech IV Semester **Regulation: R23**

Subject and Code: Machine Learning 23CSM241

3	5	**Explain the Fuzzy C-Mean Clustering algorithm in detail.**	**L4**
4	5	**a. Describe the key difference between FCM and traditional clustering methods like K-means.** **b. Describe the objective function used in FCM and the role of fuzziness parameter(m)** **c. Explain the steps involved in the FCM algorithm including how membership values and cluster centers are updated.**	**L3**
5	5	**a. What is soft clustering, and how does it differ from Hard Clustering?** **b. What is Spectral Clustering, and how does it differ from traditional clustering methods like k-Means?**	**L5**
6	5	**Given a one dimensional data set {1,5,8,10,2 }use agglomerative clustering algorithms with the complete link to establish a hierarchical grouping relationship. Note: Use Eclidean distance measure for calculation (a). By using the cutting threshold of 5, how many clusters are there?** **(b). What is their membership in each cluster?** **©. Use the Dendrogram representation to show the grouping.**	**L4**
7	5	**Explain the expectation maximization algorithm for clusteringin detail.**	**L3**
8	5	**a. Describe the conceptual difference between EM-based clustering and K-means clustering** **b. Illusttate the probabilistic model used in expectation maximization algorithm and what is the role of Gaussian Mixture Models.**	**L5**
9	5	**Explain the steps involved in the expectation maximization algorithm for clustering.**	**L4**
10	5	**What is Rough Clustering? What are the steps involved in Rough K-means Clustering?**	**L3**

Note: L1-Remembering, L2-Understanding, L3-Applying, L4-Analyzing, L5-Evaluating, and L6-Creating

Instruction to Faculty Members:

The Six Levels of Bloom's Taxonomy:

1. **Remembering:** Retrieving, recognizing, and recalling relevant knowledge from long-term memory (e.g., list, define, name, locate).

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES (AUTONOMOUS)

QUESTION BANK

Year / Semester: II B.Tech IV Semester **Regulation: R23**

Subject and Code: Machine Learning 23CSM241

2. **Understanding:** Constructing meaning, explaining ideas, or concepts (e.g., summarize, interpret, classify, compare).
3. **Applying:** Using information in new situations or implementing procedures to solve problems (e.g., solve, use, demonstrate, implement).
4. **Analyzing:** Breaking material into constituent parts, determining how the parts relate to one another and to an overall structure (e.g., contrast, categorize, distinguish, diagram).
5. **Evaluating:** Making judgments based on criteria and standards through checking and critiquing (e.g., judge, critique, justify, defend, argue).
6. **Creating:** Putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure (e.g., design, construct, develop, formulate).