SR 22

St. Peter's Engineering College (Autonomous) Dullapally (P), Medchal, Hyderabad – 500100 QUESTION BANK						Dept.	:	CSE(AIML)
						Academic Year 2025-26		
Subject Code	:	AS22-66PC05	Subject	:	MACHINE LEARNING			
Class/Section	:	B. Tech.	Year	:	III	Semester	:	I

BLOOMS LEVEL					
Remember	L1	Understand	L2	Apply	L3
Analyze	L4	Evaluate	L5	Create	L6

Q. No	Question (s)	Marks	BL	CO
	UNIT - I			
1	Define Machine Learning?	1M	L1	313.1
2	List the types of machine learning.	1M	L1	313.1
3	State the goals of PAC learning?	1M	L1	313.1
4	Give an example of a standard learning task.	1M	L2	313.1
5	Compare supervised learning with unsupervised Learning.	1M	L2	313.1
6	Define a learning problem in the context of ML.	1M	L2	313.1
7	State the goal of a machine learning algorithm.	1M	L2	313.1
8	Name any one application of machine learning.	1M	L1	313.1
9	Explain the key idea behind statistical learning.	1M	L2	313.1
10	Differentiate between classification and regression in one sentence	1M	L4	313.1

Q. No	Question (s)	Marks	BL	CO
1.	Explain the need for machine learning in today's world with examples.	5M	L2	313.1
2.	Discuss various real-world scenarios where machine learning is applied.	5M	L2	313.1
3.	Compare and contrast the different types of learning: supervised, unsupervised, semi-supervised, and reinforcement learning.	5M	L4	313.1
4.	Demonstrate standard learning tasks in machine learning.	5M	L2	313.1
5.	Summarize Statistical Learning Framework? Explain with suitable diagrams and examples.	5M	L2	313.1
6.	Explain the concept of Probably Approximately Correct (PAC) learning and its significance.	5M	L2	313.1

7.	Discuss the steps involved in solving a machine learning problem.	**5M**	**L3**	**313.1**
8.	Differentiate between supervised and unsupervised learning with examples.	**5M**	**L4**	**313.1**
9.	Explain any two standard learning tasks and their use cases.	**5M**	**L2**	**313.1**
10.	What are the key components of a machine learning system? Illustrate with a flow diagram.	**5M**	**L2**	**313.1**
11.	Explain the relationship between data, model, and learning in ML.	**10M**	**L2**	**313.1**
12.	Describe a complete machine learning pipeline from data collection to model evaluation.	**10M**	**L3**	**313.1**
13.	State the assumptions underlying Statistical Learning Theory and discuss their relevance in model development.	**10M**	**L4**	**313.1**
14.	Illustrate how PAC learning helps in evaluating the performance of learning algorithms.	**10M**	**L3**	**313.1**
15.	Explain the role of hypothesis space in learning. How does the size and structure of the hypothesis space affect learning performance?	**10M**	**L4**	**313.1**
16.	Define a learning problem formally. How is it different from a traditional programming problem?	**10M**	**L3**	**313.1**
17.	How does machine learning differ from traditional software development? Illustrate your answer with a real-world problem example.	**10M**	**L4**	**313.1**
18.	Discuss the challenges in designing a machine learning system.	**10M**	**L5**	**313.1**
19.	Identify the limitations of the PAC learning model and explain any two extensions or alternatives that address these limitations.	**10M**	**L2**	**313.1**
20.	Elaborate on how performance of learning algorithms can be measured and improved.	**10M**	**L6**	**313.1**

S.No	Questions	Marks	BTL	CO
	UNIT- II			
	1 MARK QUESTIONS (10)			
1.	State the key difference between linear and non-linear classification.	**1M**	**L2**	**313.2**
2.	Define multi-label classification.	**1M**	**L1**	**313.2**
3.	Expand the acronym ID3 as used in decision tree algorithms	**1M**	**L1**	**313.2**
4.	Identify the function used in logistic regression to model probability.	**1M**	**L2**	**313.2**
5.	Explain the role of an activation function in neural networks.	**1M**	**L2**	**313.2**
6.	How many layers does a perceptron consist of?	**1M**	**L1**	**313.2**
7.	Name the algorithm that uses information gain for splitting data.	**1M**	**L1**	**313.2**
8.	Name a kernel function used in SVM.	**1M**	**L1**	**313.2**
9.	Define the meaning of 'K' in K-Nearest Neighbors.	**1M**	**L1**	**313.2**
10.	Identify whether linear regression is suitable for classification problems?	**1M**	**L2**	**313.2**

Long Questions (20)				
1.	Differentiate between linear, non-linear, multi-class, and multi-label classification with suitable examples.	**5M**	**L4**	**313.2**
2.	Explain the process of learning a class from examples. Include training data, hypothesis space, and learning algorithm.	**5M**	**L2**	**313.2**
3.	Discuss the mathematical formulation of logistic regression and its application in binary classification.	**5M**	**L3**	**313.2**
4.	Compare and contrast classification and regression models with real-world examples.	**5M**	**L4**	**313.2**
5.	Describe the ID3 algorithm in detail. Include how information gain is used to build the tree.	**5M**	**L2**	**313.2**
6	Explain the steps involved in building a CART (Classification and Regression Tree) with an example dataset.	**5M**	**L3**	**313.2**
7	Discuss the advantages and disadvantages of decision trees. How is overfitting handled?	**5M**	**L5**	**313.2**
8	Explain the assumptions, formulation, and applications of linear regression.	**5M**	**L2**	**313.2**
9	Describe multiple linear regression and differentiate it from simple linear regression with an example.	**5M**	**L4**	**313.2**
10	Compare logistic regression with linear regression. When would you prefer one over the other?	**5M**	**L5**	**313.2**
11	Define perceptron? Explain its architecture, working mechanism, and limitations.	**10M**	**L2**	**313.2**
12	Describe the architecture and training process of a multilayer perceptron (MLP). Include forward and backpropagation.	**10M**	**L3**	**313.2**
13	Explain the role of learning rate, activation functions, and loss functions in neural networks.	**10M**	**L3**	**313.2**
14	Explain the working of a linear SVM with a clear diagram and mathematical representation.	**10M**	**L3**	**313.2**
15	Describe how SVM can be extended for non-linear classification using kernel functions.	**10M**	**L4**	**313.2**
16	Discuss different kernel functions used in SVM and their applications. Include polynomial, RBF, and sigmoid kernels.	**10M**	**L2**	**313.2**
17	Describe the K-NN algorithm for classification. Include distance metrics and the effect of K value.	**10M**	**L2**	**313.2**
18	Compare K-NN with other classification algorithms like decision trees and logistic regression.	**10M**	**L4**	**313.2**
19	Compare the performance and suitability of decision trees, logistic regression, SVM, neural networks, and K-NN for classification problems.	**10M**	**L5**	**313.2**
20	Suppose you are solving a medical diagnosis problem. Explain how you would choose between different algorithms i.e, decision trees, logistic regression, SVM, neural networks, and K-NN, and justify your selection based on data type, interpretability, and accuracy.	**10M**	**L6**	**313.2**

UNIT – III				
Q. No.	**Question(s)**	**Marks**	**BL**	**CO**
	PART-A			
1.	Define clustering in machine learning?	**1M**	**L1**	**313.3**
2.	Identify an algorithm that follows the partitional clustering approach.	**1M**	**L1**	**313.3**
3.	Define PCA.	**1M**	**L1**	**313.4**
4.	State the key idea of Expectation Maximization?	**1M**	**L2**	**313.4**
5.	Differentiate between hierarchical and partitional clustering.	**1M**	**L2**	**313.3**
6.	How does DIANA differ from AGNES?	**1M**	**L2**	**313.3**
7.	State is the objective of Self-Organizing Maps?	**1M**	**L2**	**313.3**
8.	Explain the underlying assumption of Gaussian Mixture Models regarding data distribution.	**1M**	**L2**	**313.4**
9.	State one application of Locally Linear Embedding.	**1M**	**L3**	**313.4**
10.	Mention the data type for which K-Modes clustering is particularly well-suited.	**1M**	**L3**	**313.3**
	PART-B			
1.	Explain the process of dimensionality reduction using PCA with an example.	**5M**	**L3**	**313.4**
2.	Evaluate the strengths and weaknesses of Gaussian Mixture Models.	**5M**	**L5**	**313.4**
3.	Create a comparative chart for hierarchical vs. partitional clustering methods.	**5M**	**L6**	**313.3**
4.	Implement a basic clustering algorithm in pseudocode for numeric data.	**5M**	**L6**	**313.3**
5.	Analyze a case study where clustering helped in market segmentation.	**5M**	**L4**	**313.3**
6.	Analyze the challenges in clustering high-dimensional data.	**5M**	**L4**	**313.3**
7.	Compare Self-Organizing Maps and K-means with visualization examples.	**5M**	**L4**	**313.3**
8.	Create a conceptual framework for choosing between Self-Organizing Maps and GMMs.	**5M**	**L6**	**313.4**
9.	Apply the K-modes algorithm to a categorical dataset and describe the steps involved.	**5M**	**L3**	**313.3**
10.	Apply dimensionality reduction using Factor Analysis on a multivariate dataset.	**5M**	**L3**	**313.4**
11.	Apply the K-means algorithm to a small dataset and show the clustering process.	**10M**	**L3**	**313.3**
12.	Describe how DIANA and AGNES differ in forming hierarchical clusters.	**10M**	**L3**	**313.3**

13.	Apply the Expectation-Maximization algorithm on a two-cluster Gaussian distribution example.	**10M**	**L3**	**313.4**
14.	Compare PCA, LLE, and Factor Analysis for reducing high-dimensional data.	**10M**	**L4**	**313.4**
15.	Investigate the limitations of partitional clustering techniques like K-means.	**10M**	**L4**	**313.3**
16.	Estimate when Self-Organizing Maps are preferable over other clustering methods.	**10M**	**L5**	**313.3**
17.	Compare and contrast K-means and K-modes with suitable data types.	**10M**	**L4**	**313.3**
18.	Justify the use of PCA before clustering in high-dimensional datasets.	**10M**	**L5**	**313.4**
19.	Design a flowchart for selecting an appropriate clustering technique for mixed-type datasets.	**10M**	**L6**	**313.3**
20.	Evaluate the impact of different distance metrics on hierarchical clustering results.	**10M**	**L5**	**313.3**

UNIT – IV				
Q. No.	**Question(s)**	**Marks**	**BL**	**CO**
	PART-A			
1.	Define ensemble learning? Name any two ensemble techniques.	**1M**	**L1**	**313.5**
2.	Differentiate between Bagging and Boosting.	**1M**	**L2**	**313.5**
3.	State the consequences of using too many decision trees in Random Forest.	**1M**	**L3**	**313.5**
4.	Illustrate how Voting contributes to decision-making in ensemble learning.	**1M**	**L2**	**313.5**
5.	Explain error-correcting output codes in brief.	**1M**	**L2**	**313.5**
6.	**Define the concept of Stacking in ensemble learning.**	**1M**	**L2**	**313.5**
7.	**State the primary goal of Bayesian learning in the context of machine learning.**	**1M**	**L3**	**313.5**
8.	**Identify the key assumptions made by the Bayes Optimal Classifier.**	**1M**	**L2**	**313.5**

9.	Differentiate between a Bayesian Network and a Decision Tree.	**1M**	**L2**	**313.5**
10.	Identify where can Frequent Pattern Mining be applied?	**1M**	**L3**	**313.5**

PART-B				
1.	Implement a simple version of AdaBoost with weak learners.	**5M**	**L3**	**313.5**
2.	Apply Naïve Bayes on a text classification problem and show calculations.	**5M**	**L3**	**313.5**
3.	Analyze error-correcting output codes and their use in multiclass classification.	**5M**	**L4**	**313.5**
4.	Justify the use of ensemble methods over individual classifiers.	**5M**	**L5**	**313.5**
5.	Evaluate the impact of weak learners on AdaBoost performance.	**5M**	**L5**	**313.5**
6.	Explain the process of stacking with an example of base and meta classifiers.	**5M**	**L3**	**313.5**
7.	Analyze the limitations of Naïve Bayes and how it can be improved.	**5M**	**L4**	**313.5**
8.	Compare Naïve Bayes with Optimal Bayes Classifier in terms of assumptions and performance.	**5M**	**L4**	**313.5**
9.	Suggest a real-world use case of frequent pattern mining and explain why it fits.	**5M**	**L5**	**313.5**
10.	Explain the structure of a Bayesian Belief Network and illustrate with an example how it represents conditional dependencies among variables.	**5M**	**L4**	**313.5**
11.	Apply Random Forest on a dataset and explain how trees are combined.	**10M**	**L3**	**313.5**
12.	Explain the trade-offs between accuracy and interpretability in ensemble models.	**10M**	**L5**	**313.5**
13.	Evaluate the performance of Random Forest vs. AdaBoost on noisy data.	**10M**	**L5**	**313.5**
14.	Compare Bagging, Boosting, and Stacking in terms of learning process and accuracy.	**10M**	**L4**	**313.5**
15.	Create a flowchart showing how ensemble learning techniques are selected based on problem type.	**10M**	**L6**	**313.5**
16.	Design a model using Voting Classifier with at least three different algorithms.	**10M**	**L6**	**313.5**
17.	Explain the concept of stacking in ensemble learning. How can Bayesian Learning be incorporated into stacking models to improve classification accuracy?	**10M**	**L4**	**313.5**
18.	Describe the working of the Bayes Optimal Classifier. What are its assumptions and how does it compare with practical classifiers like Naive Bayes?	**10M**	**L5**	**313.5**
19.	Explain Naive Bayes classifier with an example	**10M**	**L2**	**313.5**

20.	Describe about the Stacking lr12 Bayesian Learning.	**10M**	**L4**	**313.5**

UNIT - V				
S.NO	**Question(s)**	**Marks**	**Bloom's Level**	**CO**
	PART-A			
1.	State the typical activation function used in the output layer for binary classification?	**1M**	**L1**	**313.6**
2.	Define an Artificial Neural Network (ANN) in machine learning?	**1M**	**L2**	**313.6**
3.	Explain the importance of weight factor in ANN.	**1M**	**L1**	**313.6**
4.	Write the command to install TensorFlow using pip.	**1M**	**L3**	**313.6**
5.	List the operating systems that support TensorFlow installation.	**1M**	**L1**	**313.6**
6.	Identify the commonly used Python package manager for installing TensorFlow.	**1M**	**L3**	**313.6**
7.	State the typical file extension used for writing and executing TensorFlow code in Python.	**1M**	**L1**	**313.6**
8.	Explain the purpose of the `tf.data.Dataset` API in TensorFlow.	**1M**	**L2**	**313.6**
9.	Name the TensorFlow function used to shuffle a dataset.	**1M**	**L1**	**313.6**
10.	How can missing values in a dataset be handled during preprocessing?	**1M**	**L4**	**313.6**
	PART-B			
1.	Apply early stopping in Keras and explain its role in training deep neural networks.	**5M**	**L3**	**313.6**
2.	Explain how model evaluation is done using validation data in Keras.	**5M**	**L3**	**313.6**
3.	Analyze the effect of increasing the number of layers in an MLP built using Keras.	**5M**	**L4**	**313.6**
4.	Evaluate model performance using metrics like accuracy, precision, and recall in Keras.	**5M**	**L5**	**313.6**
5.	Compare sequential and functional API models in Keras with code examples.	**5M**	**L4**	**313.6**
6.	Design an MLP in Keras for a multi-class classification problem using softmax output.	**5M**	**L6**	**313.6**
7.	Explain the difference between model.fit(), model.evaluate(), and model.predict() in Keras.	**5M**	**L3**	**313.6**
8.	Evaluate a Keras model on test data and explain the meaning of the metrics returned.	**5M**	**L5**	**313.6**
9.	Develop a Keras model to classify hand-written digits (MNIST) and summarize the steps.	**5M**	**L6**	**313.6**
10.	Design a Keras model that uses both numerical and text inputs	**5M**	**L6**	**313.6**

	and combine their outputs.			
11.	Explain how overfitting and underfitting can occur in neural networks. How can Keras help in preventing overfitting?	**10M**	**L3**	**313.6**
12.	Explain the architecture and working of an Artificial Neural Network (ANN). How can such a network be implemented using Keras?	**10M**	**L3**	**313.6**
13.	Explain the step-by-step process of building and training a Multilayer Perceptron (MLP) model using Keras for a classification problem	**10M**	**L6**	**313.6**
14.	Discuss different activation functions used in ANN. How are these implemented in Keras?	**10M**	**L3**	**313.6**
15.	Compare and contrast a single-layer perceptron and a multilayer perceptron. How does Keras simplify the implementation of MLPs	**10M**	**L4**	**313.6**
16.	Describe how dropout and batch normalization can be used in an MLP implemented using Keras	**10M**	**L3**	**313.6**
17.	Demonstrate the steps involved in installation of Tensorflow	**10M**	**L4**	**313.6**
18.	Explain about the loading and preprocessing data with Tensor flow	**10M**	**L3**	**313.6**
19.	Briefly explain about Tensorflow.	**10M**	**L2**	**313.6**
20.	Define and compare two preprocessing strategies in TensorFlow: using **dataset.map()** vs. Keras preprocessing layers inside the model.	**10M**	**L3**	**313.6**