Reg. No.																

Question Paper Code	14363

B.E. / B.Tech. - DEGREE EXAMINATIONS, APRIL / MAY 2026

Fourth Semester

Computer Science and Engineering (AIML)

24AMPW401 - MACHINE LEARNING TECHNIQUES WITH LABORATORY

Regulations - 2024

Duration: 3 Hours | Max. Marks: 100

PART – A (i) (MCQ)
(16 × 1 = 16 Marks)
Answer ALL Questions

		Marks	K – Level	CO
1.	Which learning method uses labeled data? a) Unsupervised Learning b) Reinforcement Learning c) Supervised Learning d) Deep Learning	1	K1	CO1
2.	A bank wants to predict whether a customer will repay a loan. Which learning type is suitable? a) Supervised Learning b) Unsupervised Learning c) Clustering d) PCA	1	K2	CO1
3.	A robot learns through trial and error. Which learning model should be used? a) Naive Bayes b) Reinforcement Learning c) Decision Tree d) PCA	1	K3	CO1
4.	In LDA, the objective is to maximize: a) Intra-class variance b) Inter-class separation c) Feature correlation d) Prediction error	1	K1	CO2
5.	Why is ICA preferred over PCA in blind source separation? a) preserves variance b) extracts statistically independent components c) reduces dimensions faster d) requires less computation	1	K2	CO2
6.	An audio processing system needs to separate mixed sound signals from multiple sources. Which technique should be used? a) ICA b) PCA c) LDA d) t-SNE	1	K3	CO2
7.	Converting categorical values into numerical form is called: a) Encoding b) Scaling c) Clustering d) Regression	1	K1	CO3
8.	Which statistical measure is commonly used to detect outliers? a) Mean b) Interquartile Range (IQR) c) Accuracy d) Entropy	1	K2	CO3
9.	A survey dataset contains missing categorical values. Which technique should be applied? a) Mode Imputation b) PCA c) Normalization d) LDA	1	K3	CO3

10.	Which algorithm uses entropy or Gini index for splitting? a) SVM b) Decision Tree c) AdaBoost d) KNN	1	K1	CO4
11.	Why does Random Forest reduce overfitting compared to a single Decision Tree? a) It combines multiple trees b) It removes features c) It uses PCA d) It increases tree depth	1	K2	CO4
12.	In Support Vector Machine (SVM), for an n-dimensional feature space, the decision boundary is an _____. a) n-dimensional vector b) (n−1)-dimensional hyperplane c) (n+1)-dimensional plane d) One-dimensional line	1	K3	CO4
13.	Which regression model can represent nonlinear relationships by introducing higher-degree terms? a) Simple Linear Regression b) Multiple Linear Regression c) Polynomial Regression d) Logistic Regression	1	K1	CO5
14.	A company wants to predict employee salary using experience, education, and skill rating. Which model is most suitable? a) Simple Linear Regression b) Multiple Linear Regression c) Polynomial Regression d) Logistic Regression	1	K2	CO5
15.	In hierarchical clustering, Agglomerative clustering follows a) Top-down approach b) Bottom-up approach c) Random approach d) Parallel approach	1	K3	CO6
16.	In K-Means clustering, the objective function minimizes a) Maximum distance between clusters b) Sum of squared distances from centroids c) Number of iterations d) Density between clusters	1	K3	CO6

PART – A (ii) (Short Answer Questions)
(6 × 3 = 18 Marks)
Answer ALL Questions

17.	Identify the performance metrics used to evaluate regression and classification models.	3	K3	CO1
18.	Compare PCA and Linear Discriminant with respect to variance and class separability.	3	K4	CO2
19.	Distinguish feature extraction and feature selection techniques.	3	K4	CO3
20.	Apply the suitable process for performing classification task using SVM on non-linearly separable data.	3	K3	CO4
21.	Compare and Contrast the Lasso Regression and Ridge Regression.	3	K4	CO5
22.	Identify the differences between K-Means and K-Medoids clustering algorithms.	3	K3	CO6

PART - B (Descriptive Answer Questions)
(6 × 11 = 66 Marks)
Answer ALL Questions

23.	a) (i)	Compare different types of Machine Learning with suitable examples.	6	K4	CO1
	(ii)	Analyze the role of bias, variance, and generalization in machine learning models.	5	K4	CO1
		OR			
	b) (i)	Analyze the major challenges in machine learning such as overfitting, underfitting, and data imbalance.	6	K4	CO1
	(ii)	Compare machine learning tools based on performance and usability.	5	K4	CO1

24. a) Construct the Principal Component Analysis (PCA) for the given dataset and find the principal components. *11 K3 CO2*

Features	F_1	F_2	F_3	F_4
x	4	8	13	7
y	11	4	5	14

OR

b) Apply the Apriori algorithm Association rule to discover the strong association rules for the given dataset. Assume support level30% and confidence 60%. *11 K3 CO2*

Transaction ID	Items Bought
1	{Bread, Butter, Milk}
2	{Bread, Butter}
3	{Beer, Cookies, Diapers}
4	{Milk, Diapers, Bread, Butter}
5	{Beer, Diapers}

25. a) Analyze different strategies in training supervised machine learning models. *11 K4 CO3*

OR

b) Distinguish numerical data and categorical data with suitable examples. Also elaborate the different encoding schemes used before model training. *11 K4 CO3*

26. a) We have data from a survey with two attributes (acid durability & strength) to classify whether a special paper tissue is good or not. Training samples are *11 K3 CO4*

X1 = Acid durability (sec)	X2 = strength (kg/sq.mtr)	Y= Classification
7	7	Bad
7	4	Bad
3	4	Good
1	4	Good

Factory produces new paper tissue that has to pass laboratory test with

(i) x1=3 and x2=7.

(ii) x1=5 and x2=5

Model the classification using KNN with k=3?

OR

b) Model the working of Support Vector Machine (SVM) and explain the role of kernel functions in classification. *11 K3 CO4*

27. a) A delivery company monitors the relationship between the speed of its delivery vehicles and the fuel efficiency obtained during trips. The company observes that as vehicle speed increases, fuel efficiency decreases. A study is conducted to determine fuel efficiency as a function of vehicle speed. Recommend the expression for fuel efficiency as a linear function of vehicle speed. *11 K3 CO5*

Vehicle Speed (km/h)	20	30	40	50	60	70	80	90	100
Fuel Efficiency (km/l)	18	16.8	15.5	14.2	13.0	11.8	10.5	9.2	8

OR

b) Construct the working of Polynomial Regression and explain how it differs from Linear Regression in handling nonlinear data. *11 K3 CO5*

28. a) Divide the following 8 points into 3 clusters using K-means algorithm from: P= {(2,10), (2,5), (8,4), (5,8), (7,5), (6,4), (1,2), (4,9)}. The initial cluster centers are C1= (2,10), C2 = (5,8), and C3 = (1,2). Iterate the K-means algorithm for 2 epochs. At the end of this epoch, identify the new clusters and new centroids {C1, C2, and C3}. *11 K3 CO6*

OR

b) Construct a Hierarchical Clustering model and explain the formation of clusters using a dendrogram. *11 K3 CO6*