BENNETT UNIVERSITY
THE TIMES GROUP

Enrollment No: 0771

END-TERM EXAMINATION, ODD SEMESTER DECEMBER: - 2025

Course Code : CSET211
Course Name : Statistical Machine Learning
Program Name: B.Tech.
Semester: III

Time: 02:00 Hours Max Marks: 40

General Instruction: -Do not write anything on the question paper except enrollment number.
Note: - Attempt the questions as per instruction given in each section.
Use of Scientific Calculator is allowed.

Section A: All questions in this section are compulsory. ***Attempt either A or B part of each question***. Each question carries 03 marks. **[5QX3= 15 Marks]**

Q1

1A Differentiate between Quantitative and Qualitative data with one example for each.

Or

1B Classify the following problems as supervised, unsupervised, or reinforcement learning:
(i) Spam email detection (ii) Recommender systems (iii) Automatic game playing

Q2

2A Given X = [1, 2, 3, 4], Y = [3, 5, 7, 9], calculate regression line using least squares.

Or

2B Normalize dataset [2, 4, 6, 8] using Min-Max Scaling and standardization.

Q3

3A Explain the purpose of the Elbow Method in K-Means Clustering. What is the key metric it uses to determine the optimal number of clusters?

Or

3B A model has training error = 10 and test error = 50. Identify overfitting or underfitting with justification.

Q4

4A Distinguish between Principal Component Analysis and Linear Discriminant Analysis. Also, state one use case where each method is preferred.

Or

4B Explain the concept of bagging and boosting with examples. How do these methods address overfitting and bias-variance trade-off?

Q5

5A Explain the concept of multiple linear regression and describe how it differs from simple linear regression. Give one real-world example where multiple predictors are necessary.

Or

5B Explain the fundamental idea behind the DBSCAN clustering algorithm. What are the roles of the two epsilon and minPts?

Section B: All questions in this section are compulsory. **Attempt either A or B part of each question**. Each question carries 05 marks **[3QX5=15 Marks]**

Q6

6A Explain various performance matrix for clustering using suitable examples.

Or

6B Apply k-means clustering to data points (1,1), (2,1), (4,3), (5,4) with initial centroids (1,1) and (5,4). Perform two iteration and compute new centroids.

BENNETT UNIVERSITY
THE TIMES GROUP

Q7

7A Confusion matrix: TP=45, FP=5, FN=10, TN=40. Calculate: Accuracy, Precision, Recall, F1-score.

Or

7B A dataset contains 6 positive and 4 negative samples. Calculate Gini Index.

Q8

8A Construct a Decision Tree using Information Gain for attribute selection (up to one level). Evaluate the tree by identifying whether it may lead to overfitting or underfitting, and justify your reasoning.

Weather	Temp	Play
Sunny	Hot	No
Overcast	Mild	Yes
Rain	Cool	Yes
Rain	Hot	No
Sunny	Cool	Yes
Overcast	Hot	No

Or

8B Apply Principal Component Analysis (PCA) to compute the first principal component using a suitable example. Interpret the result and evaluate how PCA would reduce dimensionality in this dataset.

Section C: Compulsory question **[1X10=10 Marks]**

Q9. A university wants to predict whether a student will Pass or Fail a Machine Learning course using a Naïve Bayes classifier.

Attendance	Study Hours	Assignment	Internal Marks	Result
High	>5	Yes	Good	Pass
High	4–5	Yes	Average	Pass
Medium	2–4	Yes	Average	Pass
Low	<2	No	Poor	Fail
Low	<2	No	Poor	Fail
Medium	2–4	Yes	Poor	Fail
High	4–5	Yes	Good	Pass
Medium	2–4	No	Average	Fail
High	>5	Yes	Good	Pass
Low	2–4	No	Average	Fail

1. Explain the working principle, formula and Assumption of Naïve Bayes Classifier.
2. Compute P(Pass), P(Fail).
3. Compute the following conditional probabilities Attendance = High and study hours>5
4. Predict the result for student with Attendance = High, Study Hours > 5, Assignment = Yes, Internal = Good
5. Comment on the model's decision and its possible limitations.

****