[This question paper contains 4 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 7384 **L**

Unique Paper Code : 2353010026

Name of the Paper : Machine Learning: A Mathematical Approach

Name of the Course : **B.Sc. (Hons.) Mathematics – DSE**

Semester : VIII

Duration : 3 Hours Maximum Marks : 90

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. **All** questions are compulsory. Attempt any **two** parts from each question.
3. Each question carries equal marks.

1. (a) (i) Define the fundamental difference between supervised and unsupervised learning. Give an example to illustrate your answer.
(ii) Why can't Linear regression be used for classification? Explain with help of an example.

(b) Given $X=\{-2, -1, 0, 1, 2\}$ and the corresponding $Y=\{3.1, 4.2, 7.3, 5.4, 6.5\}$, find the parameters β_0 and β_1 of the regression line $Y = \beta_0 + \beta_1 X$ that's best fits the data.

(c) Derive an error function of polynomial regression of degree 2 for given dataset: $(x,y)=\{(1,2),(2,3),(3,5),(4,10)\}$ and also evaluate optimal weights w_0, w_1 and w_2 so that $y = w_0 + w_1 x + w_2 x^2$ best fits the data.

2. **(a)** Create a confusion matrix for given problem:
In a covid test of 1000 patients, there were 45 positive tests out of which 30 were covid and 15 were falsely tested positive . Of the 955 were negative tested and 5 were incorrect. These patients had covid but were tested negatively. Compute all the parameters like accuracy , precision, and recall.

(b) Suppose we want to predict the occurrence of heart-attacks based on person cholesterol level, blood pressure and age. Data for two people given in table below:

Person	Cholesterol(mg/dL)	Blood pressure(mmHg)	Age(in years)
a	180	110	21
b	230	140	40

P.T.O.

Given a weight vector w=(1,−2,1), compute the sigmoid function to evaluate the probability that person a and person b have heart attacks.

(c) Explain the Bias-Variance trade off with the help of expected prediction error decomposition.

3. (a) Consider the following two-dimensional dataset:

Observation	X_1	X_2	Y
1	3	4	Red
2	2	2	Red
3	4	4	Red
4	1	4	Red
5	2	1	Blue
6	4	3	Blue
7	4	1	Blue

i. Sketch the observations.

ii. Based on the sketch, provide an equation of the optimal separating hyperplane for the support vector machine (SVM).

iii. Identify the support vectors for the SVM classifier applied to this dataset

iv. Define the term margin in the context of SVM.

(b) Differentiate between a Regression Tree and Classification Tree. Consider the following dataset for binary classification problem:

Instance	A	B	C	Target Class
1	T	T	1	+
2	T	T	6	+
3	T	F	5	-
4	F	F	4	+
5	F	T	7	-
6	F	T	3	-
7	F	F	8	-
8	T	F	7	+
9	F	T	5	-

Calculate the information gain when splitting on A and B. Which attribute would the decision tree induction algorithm choose from A and B?

(c) Consider the following dataset that records student behavior and whether they pass or fail.

Study Hours	Attendance	Sleep Quality	Part time job	Pass/Fail
Low	Poor	Bad	Yes	Fail
High	Good	Good	No	Pass
Medium	Good	Good	Yes	Pass
Low	Poor	Bad	No	Fail
High	Good	Bad	No	Pass
Medium	Poor	Good	Yes	Fail
High	Good	Good	Yes	Pass
Low	Good	Bad	Yes	Fail
Medium	Good	Good	No	Pass
High	Poor	Good	No	Pass

Using the Naïve Bayes classifier, predict whether the following student will pass or fail:

Study Hours=Medium, Attendance=Good, Sleep Quality=Good, Part-time Job = No.

4 (a) What is meant by Dimensionality Reduction? What is the curse of Dimensionality problem? Explain any one feature subset selection technique to deal with this problem?

(b) For the unsupervised data set:

'Income': [1500, 1800, 2000, 3000, 3500, 4000, 4500],
'Spending': [200, 250, 300, 500, 600, 800, 1000]

Perform k means algorithm to obtain clusters in python. Take k=2.

(c) For the given data set, perform Principal Component Analysis (PCA).

X	Y
5	10
3	7
10	15

2	4

5. (a) Mathematically define an artificial neuron. Explain the role of the weights (w), the bias (b), and the pre-activation value $z = \sum w_i x_i + b$.

(b) Show that the logistic function σ satisfies the inequality $0 < \sigma'(x) \leq \frac{1}{4}$ for all $x \in R$.

(c) Define Mean Squared Error (MSE) and Cross-Entropy. Explain the mathematical conditions under which Cross-Entropy is preferred for classification.

6(a) What is the role of the learning rate in optimization algorithms? Derive the gradient descent update rule starting from a loss function.

(b) Define adaptive learning rate methods. Give two examples.

(c) Discuss limitations of the Universal Approximation Theorem.