

[This question paper contains 4 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 8062

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) Differentiate between supervised and unsupervised learning. Additionally provide real world application for each.

(ii) What does a high Area Under the ROC Curve (AUC) indicate a model's performance?

(b) Find the line of best fit using least square regression method, given the following data:

X	Y
2	6
6	11
5	9
10	17
12	21

Predict the value of Y for the test data $X=11$.

(c) Find the Linear regression coefficients using Gradient descent method for the following dataset when learning rate = 0.1. Carry out the process for 2 iterations.

P.T.O.

X	Y
0	2
1	3
2	5
3	4
4	6

Q2. (a) Answer the following for 5-fold cross validation on a training set of 45 tuples:

- How many rounds of learning will be performed?
- State the size of the training and testing set for each round of learning.
- How would the accuracy of the classifier be calculated in 5-fold cross validation?

(b) A multiple logistic regression model is used to predict the probability of stroke with the following coefficients

Variable	Coefficient
Intercept β_0	-14.2
Systolic β_1	0.097
Smoking β_2	0.008
Age β_3	0.065

- Identify the dependent variable.
- Write Logistic regression equation
- Find the probability of the stroke.
- Find the predicted class of the output.

(c) A spam detection model classifies 200 emails. Out of these, 160 are actually spam and 40 are not. The model correctly identifies 150 spam emails and incorrectly marks 10 non-spam emails as spam. Construct the confusion matrix. Calculate the following metrics: Accuracy, Precision, Recall, True Positive Rate (TPR) and False Positive Rate (FPR).

3 (a) What is the assumption of the Naïve Bayes classifier? What are the advantages of the assumption? Consider the following dataset with the attribute purchase as the class label.

Product	Price	Quantity	Day	Purchase
Laptop	high	1	Weekday	Yes
Headphones	low	2	Weekend	No

Mouse	low	3	Weekday	Yes
Keyboard	low	1	Weekend	Yes
Monitor	high	1	Weekday	No
Laptop	high	1	Weekend	Yes
Headphones	low	1	Weekday	No
Mouse	low	2	Weekend	Yes
Keyboard	low	1	Weekday	No
Monitor	high	1	Weekend	Yes

Using this data set train a Naïve Bayes classifier and predict the class label "Purchase" for a new instance with the following attribute values: Product = Headphones, Price = low, Quantity = 1, Day = Weekday.

(b) Discuss the importance of feature selection in Machine Learning. Also, discuss the various strategies followed for feature selection. Differentiate between Principal Component Analysis, Linear Discriminant Analysis and Quadratic Discriminant Analysis.

(c) What is meant by Tree Pruning? Mention different types of Tree Pruning with 1 characteristic each. Why is it needed? Identify the first splitting attribute for decision tree with the following dataset

Field	Experience	Hired
IT	Coding	No
IT	Coding	No
IT	Administration	Yes
IT	Administration	Yes
Business	Coding	Yes
Business	Coding	Yes
Business	Administration	No
Business	Administration	No

4. (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. Sketch the optimal separating hyperplane and provide equation of the hyperplane.
 - iii. On your sketch, indicate the margin for the maximal margin hyperplane.
 - iv. Define the term kernel in the context of SVM.
- (b) Use K-means clustering to cluster the following data into two groups: $\{2, 4, 10, 12, 3, 20, 30, 11, 25\}$. Assume cluster centroid are $m_1 = 2$ and $m_2 = 4$. Give two drawbacks of this method.
- (c) Explain the concept of Gini Index in detail and illustrate how it is used in decision tree construction with an example.
- 5(a) Define the Hyperbolic Tangent function. Compare its range to the Sigmoid function and explain why being "zero-centered" is an advantage for optimization.
- (b) Define the Mean Squared Error (MSE) and Cross-Entropy loss functions. Under what mathematical conditions is one preferred over the other?
- (c) Define logistic function σ_c with parameter c . Show that it satisfies the inequality $\sigma'_c(x) \leq \frac{c}{4}$ for all $x \in \mathbb{R}$.
- 6 (a) How does normalization affect gradient flow in deep networks?
- (b) Show that L1 regularization leads to non-differentiability at zero.
- (c) How does network width affect approximation capability?