

[This question paper contains 4 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 6554

K

Unique Paper Code : 2342203502

Name of the Paper : Machine Learning

Name of the Course : B.A. (Prog.)

Semester : V

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. **Section A** is compulsory.
3. Attempt any **four** questions from **Section B**.
4. Parts of a question must be answered together.
5. Use of a scientific calculator is allowed.

Section A

1. a) Discuss the major types of machine learning and list one real-world application for each type. 3
- b) Compare Linear Regression and Logistic Regression, with reference to their objectives, cost functions, and the nature of the predictions they generate. Support your comparison with suitable examples for each model. 3
- c) Consider the following labeled data points: 3

Point	Coordinates (x, y)	Class
A	(2, 4)	1
B	(4, 6)	1
C	(6, 8)	2

Determine the closest neighbor to the test point T (5, 5) among the given training data points using an appropriate distance metric.

- d) Differentiate between feature selection and feature extraction. 3
- e) Why does the R^2 score (coefficient of determination) consider a better measure of model performance compared to the Residual Standard Error (RSE). 3
- f) Consider the following data points: 3

Point	Coordinates (x, y)
A	(1, 2)
B	(2, 4)
C	(5, 8)

Using the Single Linkage method of hierarchical clustering, compute the first two clusters that will be formed based on their inter-point similarity.

- g) Explain the concept of "bias" in Machine Learning. How does the increase in the model complexity affects the "biasness" of the model. 3
- h) What is overfitting in a regression model? How does regularization help to prevent overfitting? 3
- i) Define Sigmoid activation function and explain its usage in Neural Network. 3

P.T.O.

- j) Given the precision and recall values of two medical test classifiers X and Y, determine the model that shows overall better performance using an appropriate metric. 3

Model	Precision	Recall
Model X	0.68	0.92
Model Y	0.85	0.55

Section B

2. a) Consider the following 1-dimensional dataset: 7
 {4, 8, 12, 20, 32, 36, 48}

Assume that the number of clusters $k = 2$, and the initial cluster centers for the k-means algorithm as 32 and 48. Perform k-means clustering for two iterations to determine the final cluster assignments and updated cluster centers.

- b) What is Neural Network? Why is non-linearity introduced in Neural Network? 4
 c) What is the limitation of high-dimensional dataset? Mention two advantages of dimensionality reduction. 4

3. a) The data below represents the number of hours studied (x) and the corresponding marks obtained (y): 7

Hours Studied (x)	Marks Obtained (y)
2	50
4	65
6	80
8	90

Compute the following for the given one dimensional dataset:

- i. Linear regression equation of the form: $y = mx + b$, using the least squares method.
 ii. If the actual observed marks are 78, compute the residual error.
 b) List the main steps involved in performing PCA (Principal Component Analysis). 4
 c) For the given feature values: 4
 $X = \{10, 20, 30, 40\}$
 i. Apply Min-Max Normalization on X.
 ii. Perform Z-score Standardization on X.

4. a) The dataset below represents weather conditions and the corresponding decision to play outdoors: 10

ID	Temperature	Wind	Humidity	Will Play
1	Hot	Strong	High	No
2	Hot	Weak	High	No
3	Mild	Weak	Normal	Yes
4	Cool	Strong	Normal	Yes
5	Cool	Weak	Normal	Yes
6	Cool	Strong	High	No
7	Mild	Weak	High	Yes
8	Mild	Strong	High	Yes
9	Mild	Weak	High	No
10	Hot	Weak	Normal	No

- i. Calculate the Entropy for features- "Temperature" and "Wind".
 - ii. Compute the Information Gain for features- "Temperature" and "Wind".
 - iii. Out of attributes "Temperature" and "Wind", identify the best feature to be used as splitting attribute at the root node of the Decision Tree.
 - iv. Justify your selection with proper calculations and reasoning.
- b) What are Support Vectors in SVM (Support Vector Machine)? Explain their role in defining the decision boundary. Also, describe margin and optimal hyperplane with the help of a diagram. 5

5. a) Consider the following one-dimensional data points: 7

Point	Value
A	2
B	4
C	7
D	9

Apply the Agglomerative Hierarchical Clustering (AHC) algorithm with the Complete Linkage method to generate a clustering scheme. Draw a dendrogram showing the sequence of cluster merging, based on the computed distances between data points.

- b) Describe the k-fold cross-validation process with an example where a dataset of 200 samples is divided into training and testing subsets. What will be the size of each fold when $k=4$? Can Leave-One-Out Cross-Validation be considered a special case of k-fold cross-validation? Justify your answer. 4
- c) Determine the type of Machine Learning technique to be used in each of the following scenarios. Justify your answer: 4
- i. Speech recognition trained with labeled audio data.
 - ii. A market segmentation model that discovers hidden patterns within the transaction records.
 - iii. A gaming agent that learns through rewards and penalties.
 - iv. A news app that classifies articles as "sports", "politics", or "entertainment".

6. a) Consider the following data of job applicants, where the attribute "Selected" represents the class label: 10

ID	Qualification	Experience	Age	Selected
1	Graduate	High	28	Yes
2	Undergraduate	Low	22	No
3	Postgraduate	Medium	30	Yes
4	Graduate	Low	25	No
5	Graduate	Medium	27	Yes
6	Postgraduate	High	32	Yes
7	Undergraduate	Medium	24	No
8	Graduate	High	29	Yes
9	Graduate	Low	23	No
10	Postgraduate	Medium	31	Yes

Using this dataset, train a Naive Bayes classifier model. Find the probability to predict the class label "Selected" for a new applicant with the following attribute values: {Qualification = Graduate, Experience = Medium, Age = 29}. Show every intermediate calculation clearly.

- b) Compute the output of a single-layer neural network with two input neurons and one output neuron, where the input vector is $(x_1, x_2) = (0.4, 0.7)$, the weight vector is $[w_1, w_2] = [0.6, -0.4]$, and the bias is $b = 0.2$. The output neuron uses an appropriate activation function for binary classification. Calculate the final output "y". Also, use the activation function that is commonly used in the hidden layer of modern neural networks for computation. Mention the used function explicitly. 5
7. a) 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 and Recall. 7
- b) Two features, namely *Age* (in years) and *Income* (in rupees), are used in a dataset. What is the need for data standardization before training a distance-based algorithm such as KNN on the given data? 4
- c) Explain the impact of changing the regularization parameter (λ) in L2 Regularization on the model's performance. 4
- What happens if λ tends to zero?
 - What happens if λ tends to infinity?