

[This question paper contains 8 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 2336

I

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 should be attempted together.
5. Use of Scientific calculator is allowed.

Section A

1. (a) Explain the difference between unsupervised and supervised machine learning techniques. Additionally, provide a real-world example for each. (5)

P.T.O.

(b) Describe the k-fold cross-validation method with a suitable example. What are the ideal choices for the value of k? Can we say that Leave-one-out cross-validation is a special case of k-fold cross-validation? Justify your answer. (5)

(c) A medical test was conducted on 500 individuals, out of which 400 were healthy, and 100 were truly sick. Among the sick individuals, only 70 tested positive test result, while the rest received a negative result. For the healthy individuals, 25 tested positive, and 375 received a negative result. Construct a confusion matrix for the results mentioned above. Calculate the True Positive Rate (TPR), Specificity, and Accuracy of the test. (5)

(d) Given the following Boolean functions, construct decision trees to represent each function. Clearly label the decision nodes and branches to show how each variable contributes to the output.

(i) $F(A,B)=A \wedge B$

(ii) $G(A,B)=A \vee B$

(5)

- (e) Apply the K-nearest neighbor algorithm on the following dataset with $K=3$ to predict the class label for a new data point X with Feature A=3 and Feature B = 6. (5)

(Use the Euclidean distance metric)

Feature A	Feature B	Class
5	8	Negative
6	5	Negative
2	3	Positive
1	4	Positive

- (f) Describe the steps followed to cluster data using k-means algorithm. What is the main drawback of this algorithm? (5)

Section B

2. (a) Define the Naive Bayes Classifier and its mathematical formulation. Enumerate the key assumption underlying the Naive Bayes classifier. (5)
- (b) Consider the following table of customer transaction data, with the "Purchase" attribute representing the class label: (10)

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 dataset, train a Naive Bayes classifier and predict the class label "Purchase" for an instance with the following attribute values: Product = Headphones, Price = low, Quantity = 2, Day = Weekday.

3. (a) Explain how changing the tuning parameter λ in ridge regression affects the model.

(i) What happens if $\lambda \rightarrow 0$?

(ii) What happens if $\lambda \rightarrow \infty$?

- (b) Given the following dataset of ages and test scores for a group of students, fit a linear regression model using the ordinary least squares method. Determine the residual error for the best-fit line. Create a linear regression model for this data and predict the test score for a student with an age of 21. (10)

Age (x)	Test Score (y)
16	85
18	90
20	92
22	88
24	95
26	92

4. (a) What do you understand by Cross-Validation? How is it different from the Hold-out method? (5)

Consider a dataset consisting of 50 instances. In the context of 5-fold cross-validation, answer the following :

- (i) How many iterations of training will be conducted?

- (ii) What will be the size of the training and testing sets for each iteration?
- (iii) Describe how the accuracy of the classifier is calculated during 5-fold cross-validation.
- (b) Consider the following dataset representing customer behaviour metrics for an online retail platform :

Customer	Average Order Value	Purchase Frequency
C1	0.85	0.62
C2	0.45	0.78
C3	0.92	0.55
C4	0.38	0.43
C5	0.73	0.81
C6	0.56	0.67

Using the Complete Linkage hierarchical clustering method calculate the distance matrix using Euclidean distance and construct a dendrogram showing the hierarchical clustering process

5. (a) For a maximal margin classifier, which points of the training set are called as support vectors? Discuss the capability of this classifier for data which is not linearly separable.

- (b) Describe the structure of a multilayer neural network with a suitable example. What are general strategies employed to prevent a neural network from overfitting? (5)
- (c) The provided regression coefficients show a multiple logistic regression model used to predict the risk of sudden death. Using these coefficients calculate the predicted probability of death for a man with Diastolic blood pressure of 130 mmHg, Cholesterol of 280 mg/100mL, and Glucose of 120 mg/100mL. (5)

Risk Factor	Regression Coefficient
Constant term (β_0)	-17.2
Blood Pressure (mm Hg) (β_1)	0.112
Cholesterol (mg/100 mL) (β_2)	0.0071
Glucose (mg/100 mL) (β_3)	0.0075

6. (a) Discuss the importance of feature selection in Machine Learning. Also, describe the various strategies followed for feature selection. (5)
- (b) Consider the following customer purchase dataset for an e-commerce platform: (10)

Customer ID	Purchase	Age Group	Income Level	Purchase
1	Yes	Young	High	Yes
2	No	Middle	Low	No
3	Yes	Senior	High	Yes
4	No	Young	Low	No
5	Yes	Middle	High	No
6	No	Senior	Low	Yes
7	Yes	Young	High	Yes
8	No	Middle	Low	No

Using the Entropy criterion construct a decision tree.

7. (a) What do you understand by Principal Component Analysis (PCA)? Enumerate the steps involved in performing PCA. (5)
- (b) Explain how SVM can be used for multiclass classification. (5)
- (c) Calculate the output of a single-layer neural network with two input neurons and one output neuron. The input feature vector is $(x_1, x_2) = (0.7, 0.3)$, the weight vector is $[w_1, w_2] = [0.5, -0.2]$, and the bias is 0.1. Use the sigmoid function as the activation function. (5)