[This question paper contains 8 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 1745 — **L**

Unique Paper Code : 2342013602

Name of the Paper : Machine Learning

Name of the Course : **B.Sc. (H) Computer Science**

Semester : VI

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. Attempt **all** questions from **Section-A**.
3. Attempt any **four** questions from **Section-B**.
4. Attempt all parts of the question together.
5. Use of Scientific Calculator is allowed.

SECTION A

1. (a) A student trains a logistic regression model on a medical dataset to predict cancer (positive class = 1). The model achieves 98% accuracy, but the doctor rejects it immediately. Explain precisely why is the doctor right to reject it, what metric should be used instead, and give its mathematical formula. (3)

   (b) For each of the following three real-world applications (A1, A2, A3) of machine learning, identify the primary type of machine learning involved (Supervised, Unsupervised, or Reinforcement) and provide a one-sentence reason for your choice. (3)

P.T.O.

- A1: A spam filter learns to identify emails as "spam" or "not spam" using thousands of previously labelled emails.
- A2: A self-driving car learns to navigate a highway by receiving a reward for staying in its lane and a penalty for swerving, without being told every correct steering angle.
- A3: An online news website groups articles into topics (e.g., sports, politics, tech) based only on the words they contain, without any pre-assigned topic labels.

(c) What is the bias-variance trade-off? Is it feasible to minimize both bias and variance at the same time? Explain briefly. (3)

(d) Normalize the given data using standardization method

Data (X): [2, 4, 11, 23, 27, 29] (3)

(e) A classification model performs with a high accuracy on training data, but generalizes poorly to new instances. Identify the problem and briefly explain one method to overcome this problem. (3)

(f) Describe and illustrate the Sigmoid function. Determine its derivative as well. (3)

(g) Show that the XOR function is not linearly separable, and hence cannot be solved by a neural network with no hidden layer. (3)

(h) A dataset at a node of a decision tree has the following class distribution:

- Node X: 12 samples of "Yes", 0 samples of "No"
- Node Y: 6 samples of "Yes", 6 samples of "No" (3)

(i) Compute the entropy H for both Node X and Node Y. Show all the calculations clearly.

(ii) Which node is pure and which is maximally impure based on the computed values?

(i) Explain "why liner regression is not suitable for classification problems". (2)

(j) A machine learning-based classifier was tested on 440 patients, yielding the following results: 100 patients were correctly diagnosed with the disease, 300 were correctly identified as healthy, 15 healthy patients were incorrectly classified as diseased, and 25 diseased patients were incorrectly classified as healthy.

Construct the confusion matrix for the given data. Also, calculate the accuracy, precision, and F1-score. (4)

SECTION B

2 (a) Find the regression line for the given dataset using the Ordinary Least Square method. Show the computation at each step. Predict the cost of a house if number of bedrooms are 8. Note that the Property ID column is an identification label (metadata) and should not be used as a numerical feature in the regression calculations. (5)

Property ID	Bedrooms	Cost (lakhs)
H-1	2	12
H-2	1	6
H-3	3	14
H-4	4	16
H-5	5	20

(b) Describe the Best Subset Selection algorithm for feature selection. Also discuss its complexity. (5)

(c) Consider a logistic regression model $\hat{p}(X)=\frac{e^{-6+0.005X}}{1+e^{-6+0.005X}}$ to classify whether a customer is a defaulter or not. Evaluate the model for two customers, each with a balance of Rs. 8000:

- Customer A actually Defaulted (y=1).
- Customer B actually Did Not Default (y=0).

Calculate the predicted probability and the total likelihood.

P.T.O.

3. (a) Consider a fully connected feed-forward neural network with 3 input units (x_1, x_2, x_3) each set to 1.0 , a single hidden layer containing 2 units, and 1 output unit, where all connection weights are 1.0, bias at hidden layer is –1.0, and the output layer unit has a bias of +1.0. All neurons utilize the ReLU activation function . Given a ground truth value of y=1.0, answer the following :

- Draw the schematic diagram for this network, clearly labelling all weights on the connections and the biases at each node.
- Calculate the total number of trainable parameters (total weights and total biases) in the network.

Perform a step-by-step forward propagation to determine the final predicted output $\hat{y}$. (5)

(b) What is regularization? When would you choose Ridge Regression? What is the effect of the following on the model :

(i) The regularization parameter $\lambda \to 0$

(ii) The regularization parameter $\lambda \to \infty$ (5)

(c) Suppose we have a classification dataset consisting of 2 classes A and B with 100 and 50 samples respectively. We use stratified sampling to split the data into train and test sets. Which of the following train-test splits would be appropriate? Justify your answer. (5)

ST. Train- {A : 80 samples, B : 30 samples}, Test- {A : 20 samples, B : 20 samples}

S2: Train- {A : 20 samples, B : 20 samples}, Test- {A : 80 samples, B : 30 samples}

S3: Train- {A : 80 samples, B : 40 samples}, Test- {A : 20 samples, B : 10 samples}

S4: Train- {A : 20 samples, B : 10 samples}, Test- {A : 80 samples, B : 40 samples}

4. (a) A medical diagnostic laboratory uses K-Nearest Neighbors (KNN) to classify whether a patient has a "High Risk" or "Low Risk" of a specific metabolic condition based on two features: Glucose Level (x1) and BMI (x2). The training dataset of 5 patients is given below: (5)

Patient Id	Glucose (x1)	BMI (x2)	Risk Level (y)
P1	100	18	Low
P2	110	22	Low
P3	105	25	Low
P4	135	32	High
P5	150	35	High

A new patient arrives with Glucose (x1) = 130 and BMI (x2) = 28. Using the Euclidean Distance metric and (K=3), determine the Risk Level for the new patient. Show all the steps.

(b) Distinguish between hard margin and soft margin in SVM. Write the constraint that must be satisfied to identify support vectors. Also analyse the role of C and ξ_i in controlling margin violations. (5)

(c) Why is it important to perform standardization before applying Principal Component Analysis (PCA) method? Give the formal equation for computing first principal component and its interpretation in PCA method. (5)

5 (a) What is the key assumption of a Naive Bayes Classifier? Using Naive Bayes classification method for the following data, predict whether a working individual with Age = Old and Income = High will buy a computer or not. (5)

Id	Age	Income	Working	Buy Computer
1	Young	High	No	No
2	Young	High	No	No
3	Middle	High	No	Yes
4	Old	Medium	No	Yes
5	Old	Low	Yes	Yes
6	Old	Low	Yes	No
7	Middle	Low	Yes	Yes
8	Young	Medium	No	No

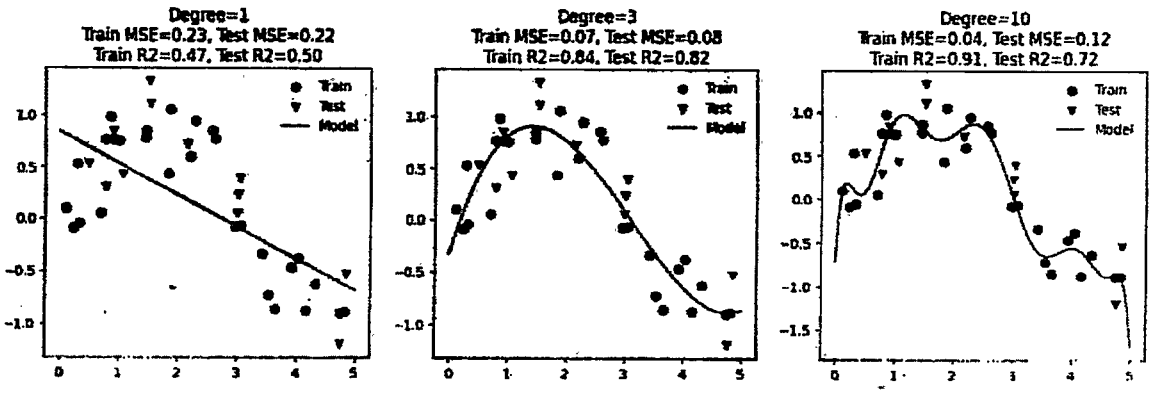
(b) You are using a One-vs-All (OVA) approach for a 10-class problem. You notice the confidence scores are inconsistent. Logically, what is the primary weakness of OVA compared to One-vs-One (OVO)? (5)

(c) Given a dataset of five points : A(1, 2), B(2, 1), C(4, 4), D(5, 4), and E(2, 5), perform hierarchical clustering using the single linkage method.

- Compute the initial Euclidean distance matrix.
- Show the stepwise merging of clusters and the updated distance calculations.
- Draw the final dendrogram, clearly labelling the merge distances on the vertical axis. (5)

6. (a) A neuron has two inputs x1=0.5 and x2=0.8 with weights w1=-0.2 and w2=0.5. If the bias b=-0.1, calculate the weighted sum(z) and find output using a Sigmoid activation function (output 1 if threshold $\geq$ 0.5 else 0). (5)

(b)



Consider polynomial regression models of degree 1, 3, and 10. Based on the given plots and their associated performance metrics (training and test error/R^2), answer the following :

- Which model is the most suitable for deployment?
- Explain the primary mathematical evidence (from the metrics above the plots) that justifies this selection over the other two options. (5)

(c) Given a dataset with four points: A(1,1), B(2,1), C(4,3), and D(5,4).

The points are grouped into two clusters:

- Cluster 1: {A, B}
- Cluster 2: {C, D}

Calculate the centroid for each cluster and the total SSE for this clustering scheme. (5)

7. (a) Using the ID3 Decision Tree algorithm, determine the root node based on information gain from the given dataset considering 'Play Game' as the target variable: (5)

Weather	Temperature	Play Game
Sunny	Hot	No
Sunny	Hot	No
Overcast	Hot	Yes
Rainy	Mild	Yes
Rainy	Cool	Yes
Sunny	Mild	No

(b) A linear regression model is used to predict house prices. The following table shows the Actual prices (y) and the Predicted prices ($\hat{y}$) for 4 houses (in lakhs) : (5)

House No.	Actual prices (y)	Predicted prices ($\hat{y}$)
1	10	11
2	20	18
3	30	31
4	40	40

Compute the following Regression Evaluation Metrics using above table data

- Mean Squared Error (MSE).
- R^2 Score.

(c) What happens to the training process if the learning rate is extremely high? Explain how it affects loss convergence and model stability with a suitable diagram. (5)