Total No. of Pages 04

Roll No. .....

FIFTH SEMESTER

B. Tech.ICSEI

END SEMESTER EXAMINATION ***Nov-Dec 2024***

CO327 MACHINE LEARNING

Time: 3:00 Hours ***Max. Marks: 40***

Note: Answer ***ALL*** questions.
Assume suitable missing data, if any.
CO# is course outcome(s) related to the question.

1[a] Classify the following tasks as supervised, unsupervised, or reinforcement learning. For each, explain the type of data used (labeled, unlabeled, or interactive feedback), the task's goal (e.g., prediction, grouping, or decision-making), and the key challenge in achieving the goal, considering factors like data quality, model generalization, or exploration-exploitation trade-offs. [Keep your answer pointwise (each one sentence only)]. **[1x4=4] [CO1]**

(i) Develop a model to predict the likelihood of a customer purchasing a product based on their demographic and browsing data. Is this prediction model likely to generalize well to unseen customers?

(ii) Analyze a dataset of unlabeled social media posts to uncover key topics or themes and group them into related clusters. What metrics could be used to evaluate the quality of these clusters?

(iii) Train an autonomous vehicle to navigate through an unfamiliar environment and reach a target destination while avoiding obstacles. How can the vehicle balance exploration of the environment and exploitation of learned strategies?

(iv) Use historical real estate data, including features like location, size, and age of houses, to predict the sale price of new properties. How can the model account for differences in market conditions over time?

[b] A hospital is using a decision tree model to classify patients into "high-risk" and "low-risk" categories for a specific medical condition based on their health metrics. To evaluate the model, the effects of pruning are analyzed for certain nodes in the decision tree, focusing on changes in both training and validation accuracies. The table below provides the relevant data for analysis. Identify which nodes should be pruned to

improve the decision tree's generalization and justify your answer mathematically. **[4] [CO3, CO4]**

Table I

Metric	Node A	Node B	Node C	Node D	Node E
Training Accuracy (%) Before Pruning	97	96	98	95	94
Training Accuracy (%) After Pruning	95	94	96	93	91
Validation Accuracy (%) Before Pruning	84	83	82	81	80
Validation Accuracy (%) After Pruning	86	85	84	83	82

2. Answer *any TWO* of the followings

[a] A space research organization is using a Naive Bayes classifier to classify a newly observed celestial object as either a Comet or an Asteroid. The classification is based on three binary features: "Highly Reflective", "Irregular Orbit", and "Ice Presence" (1 for true, 0 for false). The following table summarizes the relevant counts from the training dataset: **[4] [CO3]**

Table II

Class	No. of Objects with Highly Reflective = 1	No. of Objects with Irregular Orbit = 1	No. of Objects with Ice Presence = 1	Total Objects in Class
Comet	30	50	40	60
Asteroid	20	10	5	40

A newly detected object has the following features: Highly Reflective = 1, Irregular Orbit = 1, and Ice Presence = 1. Using the Naive Bayes classifier, classify the object as a Comet or an Asteroid.

[b] A medical research team is using a Support Vector Machine (SVM) to classify patients as "High Risk" or "Low Risk" for a condition based on two health indicators: X_1 (blood pressure) and X_2 (cholesterol level). The decision boundary identified by the SVM is $0.4 \times X_1 + 0.6 \times X_2 - 120 = 0$, with two support vectors: one at ($X_1 = 140$, $X_2 = 200$) for "High Risk" and another at ($X_1 = 130$, $X_2 = 220$) for "Low Risk." Classify a new patient with $X_1 = 145$ and $X_2 = 210$. Explain the role of support vectors in the SVM. **[2+2] [CO3]**

[c] A logistic regression (LR) model is used to predict whether a student will pass an exam (Pass = 1) or fail (Pass = 0) based on two features: Hours Studied (X_1) and Number of Practice Tests Taken (X_2). The coefficients of trained LR model are: $w_0 = -4$, $w_1 = 0.6$, $w_3 = 0.8$. For a student who studied for 10 hours and took 5 practice tests, determine whether the student will pass or fail based on the probability $P(Pass = 1)$ and a decision threshold of 0.5. What would the classification be if the decision threshold is increased to 0.7? **[4] [CO3, CO4]**

3. Answer *any TWO* of the followings

[a] A telecom company is using K-means clustering to segment customers based on monthly usage patterns. The company tested different values for the number of clusters, K, and recorded the **inertia** for each. In K-means, inertia measures the sum of squared distances between each data point x_i and the centroid c_k of its assigned cluster C_k: **[4] [CO1, CO6]**

$$Inertia = \sum_{k=1}^{K} \sum_{i \in C_k} \|x_i - C_k\|^2$$

The inertia values recorded for different cluster counts K are shown below:

Table III

K	2	3	4	5	6
Inertia	3500	2500	1800	1200	1000

Explain inertia in K-means clustering and why it decreases as K increases. Determine the optimal K based on the inertia values, describing how the "elbow method" can help.

. **[b]** A PCA was performed on a dataset with four features: Temperature, Humidity, Wind Speed, and Precipitation. The first principal component (PC1) captures 65% of the total variation in the dataset, and the second principal component (PC2) captures 25%. What does it mean that PC1 captures 65% of the variation in the dataset? Should the dataset be reduced to two principal components based on the variation captured? Justify your answer. **[2+2] [CO6]**

[c] A classification model predicts whether a basketball player will score more than 20 points in a game (High Scorer = 1) or not (High Scorer = 0). In a test dataset of 500 games, 100 were actual high-scoring games. The model predicted 120 games as high-scoring, correctly identifying 80, while 20 high-scoring games were misclassified as low-scoring. Construct the confusion matrix and calculate Precision, Recall, Specificity, and F1-Score. Discuss how increasing the threshold affects Recall. **[1+2+1] [CO4]**

4. Answer *any TWO* of the followings

[a] An MLP predicts customer churn (Churn = 1, No Churn = 0) based on Monthly Spend (x_1) and Customer Support Tickets (x_2). The MLP has 2 input neurons (layer 0), 2 hidden neurons with ReLU activation (layer 1), and 1 output neuron with sigmoid activation (layer 2), The weights and biases are: **[4] [CO3]**

$$\boldsymbol{W}^{(1)} = \begin{bmatrix} 0.5 & -0.2 \\ 0.7 & 0.1 \end{bmatrix}, \boldsymbol{W}^{(2)} = \begin{bmatrix} 0.8 \\ -0.3 \end{bmatrix}, \boldsymbol{b}^{(1)} = \begin{bmatrix} 0.1 \\ -0.3 \end{bmatrix}, \boldsymbol{b}^{(2)} = -0.2$$

Given the input $\mathbf{X} = [120, 3]$, compute the hidden layer and output layer values using ReLU and sigmoid activations, respectively. Classify the customer as "Churn" or "No Churn" using a threshold of 0.5.

[b] For the MLP described in the above question, let the true label for the customer $\mathbf{X} = [120, 3]$ be Y = 1. Compute the error using binary cross-entropy loss. Then, derive the gradient of the loss with respect to the activation from the output layer. **[2+2] [CO3]**

[c] For the MLP described in the above question, Construct the computational graph showing the flow of calculations. Then, derive the gradient of the loss with respect to $\boldsymbol{W}^{(2)}$. **[2+2] [CO3]**

5[a] A two-player zero-sum game tree (Fig. 1) has a root node as the current state and leaf nodes with scores (some as unknown "?"). Player Max aims to maximize, and Player Min aims to minimize the score. Apply the minimax algorithm with alpha-beta pruning to determine the root node's optimal value. Identify pruned branches and justify the pruning decisions. **[4] [CO6]**

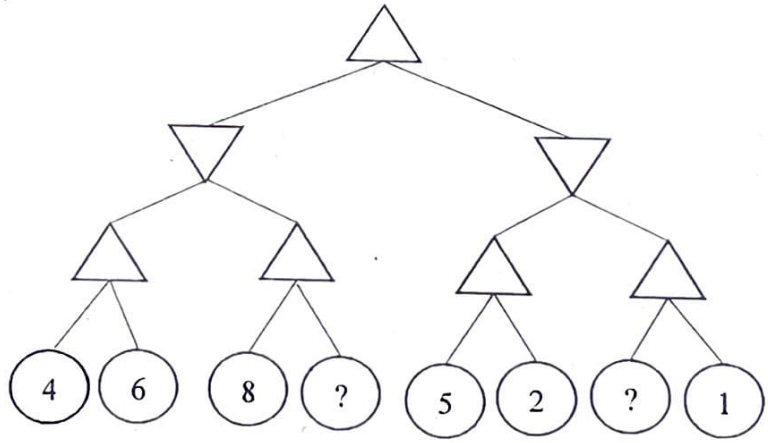

Fig. 1

[b] In a 4x4 grid world, a robot starts at S (1, 1) and aims to reach the charging station C (4, 4). Each step incurs a reward of −2, and reaching C provides a reward of +20. The discount factor is $\gamma = 0.8$. Using the value estimates for neighboring states given in the table below, compute the expected returns for each action (up, down, left, right) from (1, 1) using the Bellman update equation. Determine the optimal action at (1, 1) and explain the impact of γ on the decision. **[4] [CO6]**

Table IV

State	(1,1)	(1,2)	(2,1)	(2,2)	(3,3)	(4,4)
Value	4.0	4.8	5.2	5.6	8.0	20.0

----Best of Luck----