6E1552

Total No. of Questions : 22 Total No. of Pages : 04

Roll No. : .....

6E1552

B.Tech. VI-Sem. (Back) Exam. - 2025

COMPUTER SCIENCE AND ENGINEERING

6CS4-02 / Machine Learning

CS, IT

Time : 3 Hours **Maximum Marks : 120**

Attempt all ten questions from Part A, five question out of seven from Part B and four questions out of five from Part C.

Schematic diagrams must be shown wherever necessary. Any data you feel missing suitably be assumed and stated clearly. Units of quantities used/ calculated must be stated clearly.

Use of following supporting material is permitted during examination. (Mentioned in form No.205)

1. ..... 2. .....

PART-A [10x2=20]

(Answer should be given up to 25 words only)
All questions are compulsory

Note : All questions are **compulsory** :

1. State the assumption behind Linear regression and its mathematical form.
2. What is the Gini index in a decision tree ?
3. What is the role of the Sigmoid function in Logistic regression ?

4. What is the role of the support, confidence and life metrics in Association Rule Mining ?

5. Differentiate between feature extraction and feature selection.

6. What is a Markov Decision Process (MDP) ?

7. Differentiate between Q-learning and SARSA algorithms.

8. What is the purpose of Policy Iteration ?

9. How does the Apriori algorithm generate frequent item sets ?

10. What is the FP-Growth algorithm ?

PART-B

[5x8=40]

(Analytical/Problem solving questions)

Attempt any Five questions

Q.1. In a Spam detection model, the probability of an email being spam is 30%, and the likelihood of a keyword appearing in spam emails is 0.8, but 0.1 in non-spam. Use Bayes' theorem to calculate the probability that an email is spam if the keyword is present.

Q.2. Give these three data points : (2, 2), (4, 4), (6, 6) and initial centroids (2, 2) and (6, 6), perform one iteration of the K-means algorithm.

Q.3. KNN classifier (k = 3) predicts the class of a point (2, 2) using the following labeled data :

X	Y	Class
1	1	A
2	3	B
3	3	A
3	2	B

What is the predicted class ?

Q.4. Give the data set :

X	Y
1	2
2	3
3	4

Compute the first principal component.

Q.5. A model gives the following confusion matrix :

	Predicted Positive	Predicted Negative
Actual Positive	40	10
Actual Negative	20	30

Calculate precision, recall and F1-score.

Q.6. How SARSA differs from Q-learning in terms of exploration and exploitation ? Explain.

Q.7. Explain the steps of backpropagation and compute the gradient for a single hidden layer neural network.

PART-C

[4x15=60]

(Descriptive/Analytical/Problem Solving/Design questions)

Attempt any four questions

Q.1. Given a logistic regression model with the decision boundary :

$$P\left(Y=\frac{1}{X}\right)=\frac{1}{1+e^{-(2X_1+3X_2-4)}}$$

Determine whether the point $(X_1=1, X_2=0)$ will be classified as 1 or 0.

Q.2. Given the following dataset, calculate the information gain if the root node splits on "Weather" :

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

Q.3. Solve the Bellman equation for the following policy :

$$V(s) - R(s) + \gamma \max_{\alpha} Q(s,a)$$

Given :

(i) $R(s) = 5$

(ii) $\gamma = 0.9$

(iii) $Q(s,a) = [3,4]$

Q.4. Given the following transaction :

Transaction ID	Item
1	{A, B, C}
2	{A, B}
3	{B, C}
4	{A, C}

Find frequent itemsets using Apriori algorithm with min-supoort = 2.

Q.5. Design a Convolutional Neural Network (CNN) architecture for image classication with at least two convolutional layers and a softmax output.

-------X-------