Total number of printed pages: 3 **PG/2nd/PCSE214**

2024

SUBJECT NAME Machine Learning

Full Marks : 100

Time : Three hours

The figures in the margin indicate full marks for the questions.

*Answer **any five** questions.*

1. a) What is Machine Learning? How it is related with Artificial Intelligence. What is the goal of Machine Learning? 5

   b) Explain the concept of model selection in machine learning. Discuss different strategies like cross-validation. 5

   c) Provide an example of a real-world scenario where you would choose unsupervised learning over supervised learning. Justify your choice. What type of Machine Learning algorithm would you use to allow a robot to walk in various unknown terrains? 5

   d) Describe various metrics (Accuracy, precision, recall, F1-score, R-squared) for evaluating the performance of machine learning models for classification and regression tasks. 5

2. a) Briefly discuss techniques for regularization to address overfitting in machine learning models. 5

   b) Imagine you are working on a project which is a binary classification problem. You trained a model on training dataset and get the below confusion matrix on validation dataset. 5

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

   Based on the above confusion matrix, calculate the accuracy

   c) Explain the importance of considering bias-variance trade-off when selecting and evaluating machine learning models. 5

d) P1: Suppose you are working on stock market prediction, and you would like to predict the price of a particular stock tomorrow (measured in dollars). You want to use a learning algorithm for this. 5

P2: Suppose you are working on stock market prediction. You would like to predict whether or not a certain company will declare bankruptcy within the next 7 days (by training on data of similar companies that had previously been at risk of bankruptcy).

P3: Suppose you are working on stock market prediction, typically tens of millions of shares of Microsoft stock are traded (i.e., bought/sold) each day. You would like to predict the number of Microsoft shares that will be traded tomorrow.

P1 is a _____ problem, P2 is a _____ problem and P3 is a _____ problem

3. a) Explain the concept of Linear Regression and its assumptions. 5

b) Derive the cost function used for training a Linear Regression model. 5

c) Describe the process of Gradient Descent for optimizing the cost function in Linear Regression. 5

d) Discuss the limitations of Linear Regression and how Logistic Regression addresses them. 5

4. a) Explain the Logistic Regression model for binary classification 5

b) Derive the sigmoid function used in Logistic Regression and explain its role. 5

c) Describe the Maximum Likelihood Estimation (MLE) approach for training a Logistic Regression model. 5

d) Discuss how to evaluate the performance of a Logistic Regression model for classification tasks. 5

5. a) Explain the concept of Linear Regression and its assumptions. 5

b) Derive the cost function used for training a Linear Regression model. 5

c) Describe the process of Gradient Descent for optimizing the cost function in Linear Regression. 5

d) Discuss the limitations of Linear Regression and how Logistic Regression addresses them. 5

6. a) Explain the concept of Principal Component Analysis (PCA) and its purpose in dimensionality reduction. 5

b) Derive the formula for computing the principal components (eigenvectors) of a data matrix. 5

c) Explain how to choose the number of principal components to retain for optimal information preservation. 5

d) Discuss the applications of PCA in various machine learning tasks, such as image processing or anomaly detection. 5

********** End **********