Hall Ticket No | | | | | | | | | | | Question Paper Code: ACAC03

INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)
Dundigal-500043, Hyderabad

B.Tech VI SEMESTER END EXAMINATIONS (REGULAR) - JULY 2023
Regulation: UG-20
FOUNDATIONS OF MACHINE LEARNING

Time: 3 Hours **CSE(DATA SCIENCE)** **Max Marks: 70**

Answer ALL questions in Module I and II
Answer ONE out of two questions in Modules III, IV and V
All Questions Carry Equal Marks
All parts of the question must be answered in one place only

MODULE – I

1. (a) Describe the concept of standard learning tasks in machine learning. Discuss the specific goals and techniques associated with each task. [BL: Understand| CO: 1|Marks: 7]

   (b) In the statistical learning framework, what challenges and limitations might arise when dealing with non-identically distributed data or when the underlying data distribution changes over time? How can these challenges be mitigated to maintain the performance and reliability of the learning algorithms? [BL: Apply| CO: 1|Marks: 7]

MODULE – II

2. (a) Summarize the ID3 algorithm for decision tree construction, including the information gain metric and the recursive process of attribute selection. How does ID3 handle categorical and continuous attributes in classification tasks? [BL: Understand| CO: 2|Marks: 7]

   (b) For a given problem, what trade-offs would you consider between linear regression, multiple linear regression, and logistic regression in selecting the appropriate regression algorithm. How would you determine which algorithm is best suited for a particular regression task based on the data characteristics and problem requirements? [BL: Apply| CO: 2|Marks: 7]

MODULE – III

3. (a) What is ensemble learning, and how does it differ from using a single machine learning model? Explain the concept of model combination schemes in ensemble learning. [BL: Understand| CO: 3|Marks: 7]

   (b) How would you determine the optimal number of decision trees to include in the random forest? Does the number of trees impact the model's performance, computational complexity, and generalization capabilities? Elucidate. [BL: Apply| CO: 3|Marks: 7]

4. (a) Discuss about boosting, specifically the Adaboost algorithm, work in ensemble learning. Describe the process of iteratively training weak learners and assigning weights to correctly and incorrectly classified instances. [BL: Understand| CO: 4|Marks: 7]

   (b) The predictions of individual base learners are combined using a meta-learner in stacking. How would you select the appropriate meta-learner and determine its architecture (e.g., linear regression, neural network)? What factors would you consider in designing an effective meta-learner for the stacking ensemble? [BL: Apply| CO: 4|Marks: 7]

MODULE – IV

5. (a) Explain the working of K-means clustering algorithm work. List the steps involved in assigning data points to clusters and iteratively updating the cluster centers.
[BL: Understand| CO: 5|Marks: 7]

   (b) What situations would hierarchical clustering algorithms like AGNES and DIANA be preferred over partitional clustering methods like K-means? Conversely, when would partitional clustering be more suitable? [BL: Apply| CO: 5|Marks: 7]

6. (a) Identify the purpose of principal component analysis (PCA) in clustering, and how does it help in dimensionality reduction and visualization of high-dimensional data?
[BL: Understand| CO: 5|Marks: 7]

   (b) Evaluate the quality and validity of clustering results obtained from different techniques such as hierarchical clustering, partitional clustering, or probabilistic models like Gaussian mixture models in practical applications. What metrics or validation methods could be employed to assess the performance of clustering algorithms in real-world scenarios?
[BL: Understand| CO: 5|Marks: 7]

MODULE – V

7. (a) Examine the primary differences between support vector machines (SVMs) used for linear classification and those used for non-linear classification. [BL: Understand| CO: 6|Marks: 7]

   (b) Consider a dataset of flowers, labeled as either ”rose” or ”tulip,” with features like petal length and petal width. Using a support vector machine, how would you classify a new flower as either a ”rose” or a ”tulip” based on its petal measurements? [BL: Analyze| CO: 6|Marks: 7]

8. (a) Outline the concept of a multilayer perceptron and how it enhances the capabilities of a single perceptron? [BL: Understand| CO: 6|Marks: 7]

   (b) Imagine you have a dataset of students, each labeled as either ”pass” or ”fail,” with features like hours studied and test scores. How would you apply the K-nearest neighbors algorithm to determine whether a new student will pass or fail based on their hours studied and test score?
[BL: Understand| CO: 6|Marks: 7]