



45954

**BRAINWARE UNIVERSITY**

*Library*  
**Brainware University**  
398, Ramkrishnapur Road, Barasat  
Kolkata, West Bengal-700125

**Term End Examination 2025-2026****Programme – M.Tech.(CSE)-DS-2025****Course Name – Machine Learning for Data Science****Course Code - MTD27503 (T)****( Semester II )****Full Marks : 40****Time : 2:0 Hours**

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

**Group-A**

(Multiple Choice Type Question)

1 x 10=10

1. Choose the correct alternative from the following :

- (i) Which metric is most suitable for a binary classification problem with highly imbalanced classes (e.g., fraud detection)?
  - a) Mean Squared Error (MSE)
  - b) Accuracy only
  - c) F1-score / Precision-Recall based metrics
  - d) R-squared
- (ii) K-Nearest Neighbors (KNN) is best described as:
  - a) A parametric model with fixed coefficients
  - b) A rule-based expert system
  - c) A lazy, instance-based learning algorithm
  - d) A reinforcement learning algorithm
- (iii) Candidate Elimination Algorithm is used to:
  - a) Find the set of most likely clusters in unlabeled data
  - b) Learn the version space by maintaining specific and general boundaries
  - c) Train deep neural networks
  - d) Optimize decision tree depth using pruning
- (iv) In a regression task such as predicting house prices, the target variable is typically:
  - a) A category label (e.g., spam/not spam)
  - b) A continuous numeric value (e.g., price in ₹)
  - c) A sequence of actions
  - d) A set of clusters
- (v) In Logistic Regression, the decision boundary is formed using:
  - a) A hyperplane in feature space
  - b) A set of random splits
  - c) A distance to nearest centroid
  - d) Tree pruning
- (vi) Unsupervised learning is primarily used when:
  - a) Output labels are available for all samples
  - b) No labeled output is available in the dataset
  - c) Only reinforcement signals are present
  - d) Data contains only text
- (vii) K-Means clustering aims to:
  - a) Maximize the distance between every pair of points
  - b) Minimize the within-cluster sum of squares (WCSS)
  - c) Maximize classification accuracy
  - d) Find principal components

(viii) A key advantage of DBSCAN over K-Means is that DBSCAN:

- a) Requires K in advance
- b) Works only on spherical clusters
- c) Can find arbitrarily shaped clusters and detect outliers
- d) Always produces the same clusters regardless of parameters

(ix) Bias-Variance tradeoff describes the relationship between:

- a) Model training time and inference time
- b) Underfitting (high bias) and overfitting (high variance)
- c) Number of clusters and number of classes
- d) Learning rate and batch size

(x) Which action is most likely to reduce variance?

- a) Increase model complexity
- b) Decrease training data
- c) Use bagging / Random Forest
- d) Increase learning rate significantly

### Group-B

(Short Answer Type Questions)

3 x 5=15

- 2. Write short notes on KNN decision boundaries and distance metrics. (3)
- 3. Write short notes on practical classification pipeline in scikit-learn. (3)
- 4. Write short notes on customer segmentation and personalization use-case. (3)
- 5. Explain the Random Forest algorithm. Describe how feature importance is computed and interpret its meaning. Define Out-of-Bag (OOB) error and explain how it serves as a validation mechanism. Illustrate both concepts with a suitable example. (3)
- 6. Write short notes on feature engineering and feature selection basics. (3)

OR

Write short notes on Kaggle-style ML pipeline and model comparison. (3)

### Group-C

(Long Answer Type Questions)

5 x 3=15

- 7. Explain in detail KNN decision boundaries and distance metrics. Discuss theoretical foundations, algorithmic steps, and practical considerations. Illustrate with a suitable real-world example. (5)
- 8. Explain in detail Naive Bayes classifier and independence assumption. Discuss theoretical foundations, algorithmic steps, and practical considerations. Illustrate with a suitable real-world example. (5)
- 9. Explain in detail Kaggle-style ML pipeline and model comparison. Discuss theoretical foundations, algorithmic steps, and practical considerations. Illustrate with a suitable real-world example. (5)

OR

Explain in detail AdaBoost algorithm and weak learners. Discuss theoretical foundations, algorithmic steps, and practical considerations. Illustrate with a suitable real-world example. (5)

\*\*\*\*\*

Library  
Brainware University  
398, Ramkrishnapur Road, Barasat  
Kolkata, West Bengal-700125