JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST-3 EXAMINATION-2025

B.Tech-VII Semester (CSE/IT)

COURSE CODE (CREDITS): 19B1WCI731 (2) MAX. MARKS: 35

COURSE NAME: Computational Data Analysis

COURSE INSTRUCTORS: Ekta Gandotra MAX. TIME: 2 Hours

Note: *(a) All questions are compulsory.*
(b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems
(c) Calculator is allowed.

Q. No.	Question	CO	Marks
Q1.	Consider the following distance matrix for six data points. | | A | B | C | D | E | F | |---|---|---|---|---|---|---| | **A** | 0.00 | | | | | | | **B** | 1.85 | 0.00 | | | | | | **C** | 4.72 | 3.95 | 0.00 | | | | | **D** | 7.20 | 6.55 | 2.45 | 0.00 | | | | **E** | 6.95 | 6.25 | 2.25 | 1.80 | 0.00 | | | **F** | 5.65 | 4.85 | 3.80 | 3.25 | 2.95 | 0.00 | Perform Agglomerative Hieratchical Clustering on the given 2D data points using the Single Linkage method. Show all intermediate merging steps and construct a dendrogram that illustrates the sequence of cluster merges along with the distances at which each merge occurs.	3	6
Q2.	a. Given the following two clusters of 2D points: Cluster 1: (1, 2), (2, 3), (3, 3) Cluster 2: (6, 7), (7, 8), (8, 8) Using Manhattan distance, compute the Dunn Index for these two clusters. Also, analyze the quality of the clustering based on the value of the Dunn Index obtained. b. Describe how varying the parameters epsilon (ε) and MinPts in the DBSCAN algorithm influences the formation of clusters and the detection of noise.	3	4 2
Q3.	Consider the following 8 data points with (x, y) representing their locations: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9). Use k-Means clustering algorithm to determine the three cluster centers after the second iteration. Take A1, A2, and A7 as the initial cluster centers, and use Manhattan distance to measure distances between points.	3	5

Q4.	Explain the process by which AdaBoost algorithm adjusts the weights of training samples to enhance model accuracy. Illustrate your explanation with a small numerical example.	5	5
Q5.	a. Explain the parameter update rule used in Gradient Descent and analyze how different choices of the learning rate influence the convergence behavior of the algorithm. b. State two conditions under which the Pearson correlation coefficient can give misleading results.	1	3 2
Q6.	A data scientist trains two different machine-learning models on a dataset: • Model A: Training accuracy = 97%, Test accuracy = 53% • Model B: Training accuracy = 63%, Test accuracy = 60% Evaluate the performance of both models in terms of overfitting and underfitting. Recommend strategies to improve their generalization.	5	4
Q7.	Formulate the Elastic Net Regression loss function and analyze how the combined L1 and L2 regularization terms influence model complexity, feature selection, and generalization performance.	5	4