JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATION- 2024

B.Tech- VI Semester (ECE)

COURSE CODE(CREDITS): 19B1WEC636 (3) MAX. MARKS: 35

COURSE NAME: Machine Learning for Data Analysis

COURSE INSTRUCTORS: Dr. Alok Kumar MAX. TIME: 2 Hours

Note: *(a)All questions are compulsory.*

(b)Marks are indicated against each question in square brackets.

(c) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

---

Q.1 How does the choice of single and complete linkage method affect the hierarchical structure of the clusters in hierarchical clustering? Use the similarity matrix given in Table 1 to perform single linkage hierarchical clustering. Show your result by drawing a dendrogram. The dendrogram should clearly show the order in which the points are merged.

	P1	P2	P3	P4	P5
P1	1.00	0.10	0.41	0.55	0.35
P2	0.10	1.00	0.64	0.47	0.98
P3	0.41	0.64	1.00	0.44	0.85
P4	0.55	0.47	0.44	1.00	0.76
P5	0.35	0.98	0.85	0.76	1.00

Table 1

[CO3, CO4] **[4 Marks]**

Q.2 Consider the dataset containing the following points in a two-dimensional space:
(2,3), (4,6), (5,8), (7,7), (8,5), (9,4), (10,6), (11,3)
cluster these points into two clusters using the k-means clustering algorithm. Start with initial cluster centers at points (3, 4) and (9, 5). Perform two iterations of the k-means algorithm and determine the final cluster centers and the points assigned to each cluster.

[CO3, CO4] **[4 Marks]**

Q.3 What is the role of activation functions in neural networks? Explain the problem of the vanishing gradient and how the ReLU activation function helps alleviate it?

[CO4, CO5] **[4 Marks]**

Q.4 Consider the following dataset for a binary class problem. [CO3] **[4 Marks]**

A	B	Class Label
T	F	+
T	T	+
T	T	+
T	F	-
T	T	+
F	F	-
F	F	-
F	F	-
T	T	-
T	F	-

a) Calculate the information gain when splitting on A and B. Which attribute would be decision tree induction algorithm choose?

b) Calculate the gain in the Gini index when splitting on A and B. Which attribute would be decision tree induction algorithm choose?

Q.5 Discuss the concept of regularization techniques in neural networks and their role in preventing overfitting. [CO2, CO3] **[4 Marks]**

Q.6 Describe the steps involved in performing PCA on a dataset. What is the role of Eigenvalues and Eigenvectors in PCA? How are they related to variance and covariance?

[CO3] **[4 Marks]**

Q.7 What is the difference between LDA (Linear Discriminant Analysis) and PCA (Principal Component Analysis)? Compute the within class scatter matrix used in LDA for the following two dimensional dataset.

Samples for class 1 = $X_1 = \{(4,2), (2,4), (2,3), (3,6), (4,4)\}$

Samples for class 2 = $X_2 = \{(9,10), (6,8), (9,5), (8,7), (10,8)\}$ [CO3, CO4] **[4 Marks]**

Q.8 Describe the architecture of a basic RNN. Explain the vanishing gradient problem in RNNs.

[CO3] **[4 Marks]**

Q.9 Derive the mathematical relationship between correlation and Euclidean distance when each data object has an L2 length of 1. [CO1, CO2] **[3 Marks]**