JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -2 EXAMINATION- 2024

B.Tech-VII Semester (CSE/IT/ECE)

COURSE CODE (CREDITS): 19B1WCI731(2) MAX. MARKS: 25

COURSE NAME: Computational Data Analysis

COURSE INSTRUCTORS: VKS MAX. TIME: 1 Hour 30 Minutes

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

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

Q.No	Question	CO	Marks
Q1.	a) Consider a dataset containing observations of different species of flowers, with features including sepal length, sepal width, petal length, and petal width. After performing PCA on the standardized dataset, you find that the first principal component (PC1) explains **70%** of the variance, while the second principal component (PC2) explains **20%**. What does it imply about the structure of the data that PC1 captures such a large proportion of variance?	CO-3	[3]
	b) Derive objective function of PCA using distance minimization formulation.		[2]
Q2.	a) One drawback of ensemble methods is their complexity, which can reduce interpretability. Discuss the trade-off between performance and interpretability in ensemble learning. How might you address interpretability issues when using ensemble models?	CO-3	[2.5]
	b) Explain crowding problem of **TSNE** with the help of an example.		[2.5]
Q3.	You have the following dataset of points in a 2D space:	CO-4	[5]
	Point A: (1, 1), Point B: (1, 2), Point C: (2, 2), Point D: (8, 8), Point E: (8, 9) and Point F: (25, 25) Apply the **DBSCAN** algorithm with the following parameters:		
	Epsilon (ϵ\epsilon): **1.5** & Minimum points (MinPts): **2**		
Q4.	You have the following dataset consisting of five points in a 2D space:	CO-4	[5]
	Point A: (1, 2), Point B: (2, 3), Point C: (6, 5), Point D: (8, 8) and		P.T.O

	Point E: (9, 7) Apply the **k-medoids** algorithm to cluster these points into **k = 2** clusters. Randomly choose the initial medoids and use euclidean distance metric to perform the computations. Update both the clusters after each iteration		
Q5.	Use AdaBoost algorithm to classify a binary dataset with the following five training instances: 	Instance	Feature 1	Feature 2	Class
1	1	2	+1
2	2	3	+1
3	3	1	-1
4	4	5	-1
5	5	4	-1		CO-4	[5]