JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATION- 2023

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 What is the purpose of principal component analysis (PCA) in dimensionality reduction? Are there any limitations of PCA? Is PCA sensitive to outliers in the data? Given a dataset with 5 features and 100 observations, how many principal components can PCA potentially generate?

[CO4] [4]

Q.2 How feature extraction is different from feature reduction? How feature extraction is performed?

[CO1] [4]

Q.3 Why we use different performance metrics (Accuracy, Precision and F1 score). Explain each performance matrix parameters in detail with suitable examples. Find out accuracy and precision for the given confusion matrix.

[CO1, CO2]　[4]

		Predicted level		
		Class 1	Class 2	Class 3
True level	Class 1	64	46	139
	Class 2	12	237	42
	Class 3	52	79	165

Q. 4 What is a perceptron? What are the limitations of a single perceptron? Explain the working model of perceptron model used in neural network with suitable diagram. Consider a neural network with a single hidden layer containing 5 neurons and an output layer with 3 neurons. The input layer has 10 features. How many weights are there in the neural network?

[CO1, CO2] .[5]

Q. 5 What is the used of dendograms in clustering? Use complete link agglomerative clustering to group the data described by the following distance matrix. Also show the dendograms.

	A	B	C	D
A	0	1	4	5
B		0	2	6
C			0	3
D				0

[CO2, CO5]   [4]

Q. 6 In this dataset, we have 3 attributes which have sepal length, sepal width, and species. Species have a target attribute. In target attribute, we have three species (Setosa, Virginia, and Versicolor) and our target finds the nearest species which belong from three species using the k-Nearest Neighbors. Find the species if sample data having Sepal Length= 5.2, and Sepal Width= 3.1.   (Assume k=3).   [CO3, CO4]   [4]

Sepal Length	Sepal Width	Species
5.3	3.7	Setosa
5.1	3.8	Setosa
7.2	3.0	Virginica
5.4	3.4	Setosa
5.1	3.3	Setosa
5.4	3.9	Setosa
6.1	2.8	Versicolor
6.0	2.7	Versicolor

Q.7 What is the used of activation functions in neural network? Explain Sigmoid activation function with suitable diagram.   [CO2, CO3]   [4]

Q.8 What is the intuition behind the concept of support vectors in SVM? What are the main advantages of using SVM?   [CO3, CO4]   [3]

Q. 9 What are the advantages and disadvantages of using decision trees for machine learning tasks? What is the purpose of pruning in decision trees?   [CO2, CO5]   [3]