INDIAN STATISTICAL INSTITUTE

(Pattern Recognition)

Final SEMESTER EXAMINATION (July 2021)

(MSQMS, SQC)

Duration: 180 minutes   **Maximum Marks: 100**

1. Describe the K-medoid and DBSCAN clustering algorithms. **[6]**

2. Compare the Advantages and disadvantages of K-medoid and DBSCAN algorithms. **[6]**

3. Differentiate between the principal component analysis (PCA) and linear discriminant analysis? Describe the step-wise process of finding the PCA of a data set. **[6+4]**

4. What is ROC curve? Describe the method of finding and interpretation of this curve. **[2+3]**

5. Define the relationship between Euclidean, normalised Euclidean and Mahalanobis distance with equations. **[3]**

6. If X is a data sets with 9 samples and 2 features each. Find the Euclidean and Mahalanobis distances between 2nd and 5th samples of X. Explain the reasons for the different distance values.

$$XX = \begin{bmatrix} 4 & 5 \\ 3 & 2 \\ 3 & 2 \\ 1 & 6 \\ 0 & 2 \\ 5 & 7 \\ 7 & 9 \\ 3 & 7 \\ 2 & 1 \end{bmatrix};$$ **[10]**

7. Describe with examples and equations; the measures of Location, spread, Shape and dependency. **[2+2+2+2]**

8. Give short notes on standardization and normalization of a data set. Give one equation for finding each of them. **[2+2+2+2]**

9. Construct the Box and Whisker Plot for vector **X**.

   a. **X**=[12, 5, 22, 30, 7, 36, 14, 42, 15, 53, 25] **[6]**

10. What is Bayes decision rule for classification of a data set. Discuss the advantages and disadvantages of this rule. **[4+2+2]**

11. List the common properties of a distance METRIC. **[2]**

12. Describe the whole steps of operation of KNN classification model. Discuss the advantages and disadvantages of this model. **[4+4]**

13. In a database of 30 samples, 20 samples belong to DOG category and 10 samples belong to CAT category. The model M classifies 15 DOGs and 8 CATs correctly. Develop the confusion matrix and find the Precision, recall and F-measure of the model. **[4+2+2+2]**

14. Describe briefly the process of Semisupervised and Reinforcement learning algorithms. Also mention the motivations for the need of these algorithms. **[3+3+2+2]**