Ex/CSE/T/414A/2021

B.E. COMPUTER SCIENCE AND ENGINEERING
FOURTH YEAR FIRST SEMESTER - 2021

SUBJECT
PATTERN RECOGNITION

Time : 4 hours Full marks : 70

Answer Question No. 1 (COMPULSORY) and any four form the rest

1. Answer true or false stating reasons 2x11 = 22

(a) A Naïve Bayes Classifier always assumes that each feature x_i is conditionally dependent of every other feature x_j for $j \neq i$.
(b) The success of a pattern classification scheme using decision function depends on finding the geometrical properties of the pattern classes under consideration.
(c) A training phase is essential in a typical Syntactic Pattern classification system.
(d) The support vectors of a linear SVM can be any points in the training set.
(e) Density- based clustering algorithms provide a natural protection against outliers.
(f) The main demerit of a NB classifier is that, the assumption of class conditional independence not always valid for real life problems.
(g) Syntactic Pattern Recognition attempts to classify patterns based on a set of extracted features and an underlying statistical model for the generation of these patterns.
(h) Partitional clustering methods help in exploring data at different levels of granularity.
(i) A Hopfield net is mainly used for prediction.
(j) For a supervised pattern classification problem having M classes, where the patterns are pair-wise separable, the classifier needs to compute M number of decision surfaces to perform classification.
(k) Data preprocessing is required to ensure only the accuracy of the data.

2. Consider M pattern classes and assume that these classes are represented by prototype patterns $Z_1, Z_2, \ldots, Z_M$. Derive the expression of decision function utilizing the distance of pattern vectors from their prototype patterns. Show that such a decision function is a linear one. 12

3. What is Bayes theorem? Derive an analytical expression for Naïve Bayes classifier for a set of data $D \in \mathbb{R}^n$ to be assigned into k classes $C_1, C_2, \ldots, C_k$. Assume that a feature can be either categorical or continuous valued and each feature is conditionally independent of every other feature. 12

4. Let us define the "natural groups" present in a given data set $S = \{x_1, x_2, ..., x_n\} \subseteq \Re^2$, to be the groups that one perceives by viewing the scatter diagram of S. Assume that the number of such "natural groups" present in S is not known *a priori*. Describe a suitable clustering technique that can extract such natural groups present in S. 12

5. Suppose you need to classify the following four alphabets.

Define the necessary set of primitive strings and an appropriate grammar to solve the problem using syntactic pattern recognition method. 12

6. Write short notes on any one of the following. 12

   (a) SVM Classifier
   (b) The importance of ANN in PR