

B.E. COMPUTER SCIENCE AND ENGINEERING
FOURTH YEAR FIRST SEMESTER - 2019

SUBJECT
PATTERN RECOGNITION

Time : 3 hours

Full marks : 100

Answer question no.1 and any four from the rest

All questions carry equal marks

1. Answer true or false stating reasons 2x10 = 20
- (a) A Naïve Bayes Classifier can only work on categorical data.
 - (b) 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.
 - (c) The success of a pattern classification scheme using decision function depends only on finding the appropriate form of $d(x)$.
 - (d) Each node in a typical ANN can have any number of outgoing connections, where the signals in all of these may not be the same.
 - (e) The *Tanimoto measure* is used to find a measure of dissimilarity between patterns where the features are having binary values.
 - (f) K-means algorithm cannot cluster data sets well with large differences in densities.
 - (g) Partitional clustering methods help in exploring data at different levels of granularity.
 - (h) In a typical PR problem, the dimensionality of measurement space is either greater than or less than that of feature space.
 - (i) A Syntactic Pattern Classifier infers a grammar from a set of training patterns.
 - (j) In Bayesian Network the nodes represent variables and the edges represent conditional dependencies.
2. (a) Explain briefly the necessary steps involved in a typical Pattern Classifier. 6
- (b) Design a classifier that can classify a set of given text documents into a number of classes depending on their contents where the number of such classes is known *a priori*. Comment on the merits and demerits of such a classifier. 12 + 2
3. What is partitional clustering? State and explain one such method which is based on the principle of minimization of intra-class distances. Discuss the merits and demerits of the method. 2 + 12 + 6

4. (a) Explain briefly a typical Naïve Bayes Classifier. What are the advantages of such a classifier over KNN and Decision Tree? 6 + 4

- (b) Consider the following hypothetical data for possible buyers of laptops.

Age Group	Income	Student	Credit_rating	Class: buys_laptop
Youth	High	No	Fair	No
Youth	High	No	Excellent	No
Middle_aged	High	No	Fair	Yes
Senior	Medium	No	Fair	Yes
Senior	Low	Yes	Fair	Yes
Senior	Low	Yes	Excellent	No
Middle_aged	Low	Yes	Excellent	yes
Youth	Medium	No	Fair	No
Youth	Low	Yes	Fair	Yes
Senior	Medium	Yes	Fair	Yes
Youth	Medium	Yes	Excellent	Yes
Middle_aged	Medium	No	Excellent	Yes
Middle_aged	High	Yes	Fair	Yes
Senior	Medium	No	Excellent	No

Use a Naïve Bayes Classifier to determine whether a person represented by the following tuple (X) will purchase a laptop.

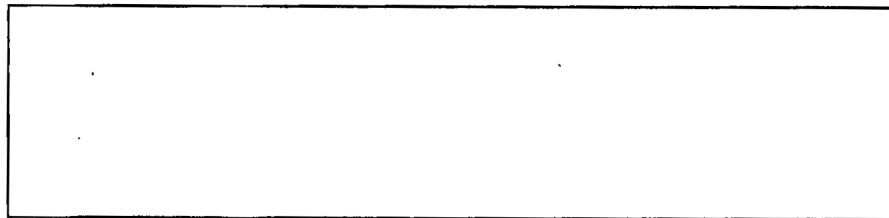
$X = (\text{Age_group} = \text{Youth}, \text{Income} = \text{Medium}, \text{Student} = \text{Yes}, \text{Credit_rating} = \text{Fair})$

10

5. (a) What is Syntactic pattern recognition ? Why is it important? 2+2

- (b) With necessary diagram explain briefly a typical Syntactic Pattern Recognition System. 8

- (c) 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. 8

6. (a) What is data mining ? How is it different from a typical DBMS ? 3+5

(b) Discuss whether or not each of the following activities is a data mining task.

- (i) Dividing the customers of a company according to their profitability.
- (ii) Predicting the future stock price of a company using historical records.
- (iii) Monitoring the heart rate of a patient for abnormalities.
- (iv) Monitoring seismic waves for earthquake activities.
- (v) Extracting the frequencies of a sound wave.
- (vi) To detect a group of people who stage accidents to gain money from insurance companies.

2 X 6

7. Write short notes on any two of the following. 2 X 10

- (a) Bayesian Belief Network
- (b) TeMplate matching vs parsing in Syntactic PR
- (c) Importance of non-metric similarity measure in clustering.