LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

M.Sc. DEGREE EXAMINATION - **COMPUTER SC.**

FIRST SEMESTER – **NOVEMBER 2013**

CS 1822 - DATA MINING

Date : 05/11/2013   Dept. No.   Max. : 100 Marks
Time : 1:00 - 4:00

Part A

Answer all the questions.   **10 x 2 = 20 Marks**

1. Write down the steps involved in KDD process.

2. How 'Single Cross Over' is differing from 'Multiple Cross Over'?

3. Give the formulae to determine 'gini' index.

4. Notify the significance of combining techniques in clustering.

5. Define Outliers.

6. What is Kohonen network?

7. How association rules are used in the telecom industry?

8. List out the advantages of partitioning algorithm.

9. Present the characteristics of suffix tree.

10. What is 'Minimum Binding Rectangle'?

Part B

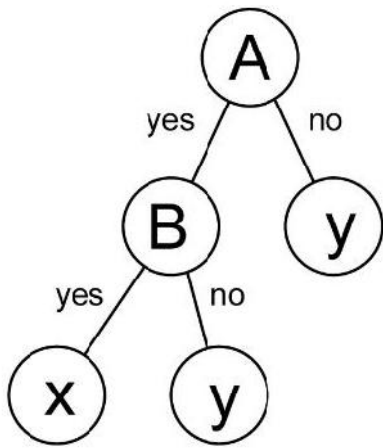**Answer all the questions.**   **5 x 8 = 40 Marks**

11 a) Illustrate the expectation maximization method to find the missing values in the following: Xobs = {5,1,4,10} when n=6, k=4 and the initial guess=3 (Or)
 b) Discuss briefly about the implementation issues in Data Mining.

12 a) Illustrate the ID3 algorithm to find out splitting attribute for the following classification problem

Id	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Age	Y	Y	Y	Y	Y	M	M	M	M	M	O	O	O	O	O
Has_Job	F	F	T	T	T	F	F	F	T	F	F	F	T	T	F
Own_House	F	F	F	T	F	F	F	T	T	T	T	T	F	F	F
Credit_rating	Fair	Good	Fair	Fair	Fair	Fair	Good	Good	Excel	Excel	Excel	Good	Good	Excel	Fair
Class	No	No	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No

Where Y= Young, M=Middle, O=Old, T=True, F=False.

(Or)   b) Derive the rules from the following Decision Tree.

13 a) Illustrate the K-means clustering algorithm to cluster the following items
   {2,4,10,12,3,20,30,11,25} when k =2 (Or)

b) With a neat diagram, explain the use of Dendrograms in clustering.

14 a) Discuss how sampling algorithm is suitable for handling large databases (Or)

   b) Describe how data parallelism is differing from task parallelism with an example.

15 a) Elaborate how personalization helps the web site to keep up the expectations of the user community
   (Or)

   b) Describe a suitable method to mine spatial data with an example.

Part C

Answer any two questions.   **2 x 20 = 40 Marks**

16 a) Discuss various methods in predictive type data mining with an example. [10 Marks]

   b) Propose a Neural Network Classification Model to read the values of gender and height and to
      propagate the output {Tall, Medium, Short} based on the input received. [10 Marks]

17 a) Describe the basic steps involved in CURE algorithm with a suitable example. [10 Marks]
   b) A grocery shop transactions are recorded as below: t1={Bread, Jelly, Peanutbutter}, t2={Bread,
   Peanutbutter}, t3={Bread, Milk, Peanutbutter}, t4={Beer, Bread}, t5={Beer, Milk}. Illustrate
   Apriori algorithm to determine Large Item set for the above sample transactions. [10 Marks]

18 a). Describe in detail about pattern discovery and its significance in the web mining. [10 Marks]
   b). Explain how the K –Nearest Neighbor algorithm of Classification differs from the K –Nearest
      Neighbor algorithm of Clustering. [10 Marks]

                                    ****