 LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

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

FIRST SEMESTER – NOVEMBER 2012

# CS 1816 - DATA MINING

 Date : 01/11/2012 Dept. No. Max. : 100 Marks

 Time : 1:00 - 4:00

**Part A**

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

1. List the steps in Knowledge Discovery in Databases.
2. Give formulae to determine Chi square.
3. What are the two approaches in implementing classification?
4. Give the gini ratio.
5. Define outliers and give an example.
6. What is Self Organizing feature Maps (SOM)?
7. Define the term Association Rule.
8. Mention any four merits of partitioning.
9. Give the taxonomy of web mining.
10. Define Spatial Association rule.

**Part B**

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

1. a) Illustrate a method to determine correlation coefficient for the two variables given below: X= <2, 4, 6, 8, 10> and Y = <1, 3, 5, 7, 9> (Or)

b) List and differentiate the use of activation functions.

1. a) Differentiate regression approaches to perform classification (Or)

b) Describe the Supervised Neural Network Learning.

1. a) Classify Clustering algorithms and explain (Or)

b) Discuss the agglomerative algorithm for Clustering.

1. a) Differentiate task parallelism and Data parallelism approaches and explain (Or)

b) Describe the methods to measure the quality of rules.

1. a) Stress the need for Web Personalization and illustrate the techniques used for the same (Or)

b) With neat diagrams, compare data structures available for Spatial Mining.

**Part C**

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

1. a) Elaborate the implementation issues in data mining.

 b) The table given below is the sample data to the height classification problem.

 Classify a new tuple t = <Adam, M, 1.95> using Naive Bayes Classification

 approach.

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Gender | Height | Output |
| Kristina | F | 1.6 m | Short |
| Jim | M | 2 m | Tall |
| Maggie | F | 1.9 m | Medium |
| Martha | F | 1.88 m | Medium |
| Stephanie | F | 1.7 m | Short |
| Bob | M | 1.85 m | Medium |
| Kathy | F | 1.6 m | Short |
| Dave | M | 1.7 m | Short |
| Worth | M | 2.2 m | Tall |
| Steven | M | 2.1 m | Tall |
| Debbie | F | 1.8 m | Medium |
| Todd | M | 1.95 m | Medium |
| Kim | F | 1.9 m | Medium |
| Amy | F | 1.8 m | Medium |
| Wynette | F | 1.75 m  | Medium |

1. a) Describe about CURE algorithm and justify its merits.

b) Illustrate Apriori Algorithm for the following sample data t1 = {Bread, jelly, PeanutButter}, t2 = {Bread, PeanutButter}, t3 = {Bread, Milk, PeanutButter}, t4 = {Beer, Bread}, t5 = {Beer, Milk} Support = 30% and Confidence = 50%.

1. a) Draw the Directed Acyclic Graph(DAG) for a corporate website with 5 web pages {A, B, C, D, E} and their visiting sequences with an integer time stamp are given as <(A,1), (B,2), (C,2), (A, 7), (C, 10), (B, 10), (C, 12), (A, 12), (C,13), (D, 14), (C,14) ,(E, 20)>. Find the general episodes where the support threshold is 30% and the time window is 3.

 b). Write an algorithm to generate rules from a decision tree and provide output

 for the sample input data given below.

 Height

≤ 1.7 m > 1.95 m

 > 1.70 m

 < 1.95 m

 Short Medium Tall

\*\*\*\*\*\*