SPRING END SEMESTER EXAMINATION-2014

8th Semester B.Tech

DATA MINING & DATA WAREHOUSING IT-811

(Regular-2010 & Back of Previous Batches)

Full Marks: 60 **Time: 3 Hours**

Answer any SIX questions including Question No.1 which is compulsory.

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable and all parts of a question should be answered at one place only.

1. a) Define Predictive Model? Discuss data mining tasks belongs to predictive model. [2 × 10]

   b) Define similarity and dissimilarity matrix with example.

   c) State why data pre-processing is an important issue for data mining and data warehousing.

   d) Define with example what do you mean by correlation analysis.

   e) Given a dataset D, number of attributes 'm', the number of training tuples |D|. Show that the computational cost of growing tree is at most $m \times |D| \times \log(|D|)$.

   f) What is the difference between classification and prediction?

   g) Which classifier try to classify data by maximizing $P(X/C_i)P(C_i)$, where 'i' is the index of class.

   h) What is Baysian classification?

i) Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8). Compute Manhattan distance & Minkowski distance between two objects. (q = 3).

j) What is information gain? Write its use.

2. a) Consider 5 points {x1, x2, x3, x4, x5} with the following co-ordinates as a two dimensional Sample for clustering: [4

x1 = (0, 2), x2 = {0, 0}, x3 = {1.5, 0}, x4 = (5, 0), x5 = (5, 2)

Perform the K-Means clustering algorithm using the above data points.

b) Discuss constraint based cluster analysis with suitable example. [4

3. a) Compute Apriori algorithm on the given transaction set and find out frequent item sets. Consider minimum support count is 3. [4

TID	Item Set
101	{S, L, O}
102	{S, BF}
103	{BF, V}
104	{S, L, O}
105	{B, O}
106	{B}
107	{B, BF}
108	{S, L, A, O}
109	{A, V}
110	{S, L}

b) Explain major issues in data mining. [4

a) Differentiate between DBMS and Data Mining? Discuss KDD process with suitable diagram. [4

b) Discuss different pre processing stages in Data Mining. [2

c) Define dimensionality reduction? Explain any two methods. [2

5. Describe each of the following approaches to clustering with example: Hierarchical Method and Density based method. [8

6. a) Briefly explain the major steps of decision tree classification with suitable example. [4

b) How to measure interestingness of a pattern? Discuss four different types of objective measures with example. [8

7. a) What is the role of OLTP and OLAP in data warehousing? State at least six difference between OLTP and OLAP. [4

b) What is the role of machine learning in data mining? Discuss different types of learning with suitable example. [4

8. Write Short Notes: (Any Two). [8

a) Genetic Algorithm.

b) Spatial Data Mining

c) FP – Tree

d) Factors involved in choosing a Data Mining System.

XXXXX

a) Differentiate between DBMS and Data Mining? Discuss KDD process with suitable diagram. [4

b) Discuss different pre processing stages in Data Mining. [2

c) Define dimensionality reduction? Explain any two methods. [2

5. Describe each of the following approaches to clustering with example: Hierarchical Method and Density based method. [8

6. a) Briefly explain the major steps of decision tree classification with suitable example. [4

b) How to measure interestingness of a pattern? Discuss four different types of objective measures with example. [8

7. a) What is the role of OLTP and OLAP in data warehousing? State at least six difference between OLTP and OLAP. [4

b) What is the role of machine learning in data mining? Discuss different types of learning with suitable example. [4

8. Write Short Notes: (Any Two). [8

a) Genetic Algorithm.

b) Spatial Data Mining

c) FP – Tree

d) Factors involved in choosing a Data Mining System.

XXXXX