R5

III B.Tech II Semester(R05) Supplementary Examinations, January 2010
DATA WAREHOUSING AND DATA MINING
(Information Technology)
Time: 3 hoursMax Marks: 80
Answer any FIVE Questions
All Questions carry equal marks

1. What is data mining? In your answer, address the following:

   (a) Is it another hype?

   (b) Is it a simple transformation of technology developed from databases, statistics, and machine learning?

   (c) Explain how the evolution of database technology led to data mining.

   (d) Describe the steps involved in data mining when viewed as a process of knowledge discovery? [16]

2. (a) Briefly discuss about data integration.

   (b) Briefly discuss about data transformation. [8+8]

3. (a) List and describe any four primitives for specifying a data mining task.

   (b) Describe why concept hierarchies are useful in data mining. [8+8]

4. (a) What is Concept description? Explain.

   (b) What are the differences between concept description in large data bases and OLAP? [8+8]

5. Write the algorithm for Apriori. Explain with a suitable example. [16]

6. (a) Compare the advantages and disadvantages of eager classification(e.g., decision tree, Bayesian, neural network) versus lazy classification(e.g., k-nearest neighbor, case-based reasoning).

   (b) Can any ideas from association rule mining be applied to classification? Explain. [8+8]

7. (a) Suppose that the data mining task is to cluster the following eight points(with (x,y) representing location) into three clusters.
   A1(2,10), A2(2,5), A3(8,4), B1(5,8), B2(7,5), B3(6,4), C1(1,2), C2(4,9).
   The distance function is Euclidean distance. Suppose initially we assign A1,B1, and C1 as the center of each cluster, respectively. Use the k-means algorithm to show only

      i. the three cluster centers after the first round execution, and
      ii. the final three clusters.

   (b) Explain DBSCAN algorithm with suitable example.

   (c) How does CLIQUE work? [8+4+4]

8. (a) How to mine Multimedia databases? Explain.

   (b) Define web mining. What are the observations made in mining the Web for effective resource and knowledge discovery?

   (c) What is web usage mining? [10+4+2]