USN | | | | | | | | | | |

06CS755

Seventh Semester B.E. Degree Examination, December 2010
Data Mining

Time: 3 hrs. Max. Marks:100

Note: ***Answer any FIVE full questions, selecting at least TWO questions from each part.***

PART – A

1 a. Define data mining. List and explain the different challenges of data mining. **(10 Marks)**

b. What is an attribute? Explain the different types of attributes, with examples. **(10 Marks)**

2 a. List the different purposes of dimensionality reduction and the different techniques used to reduce dimensionality. **(10 Marks)**

b. Calculate the distance matrix using Euclidean distance for the following data points: **(10 Marks)**

	P_1	P_2	P_3	P_4	P_5
x	4	8	15	24	24
y	4	4	8	4	12

3 a. List the different characteristics of decision tree induction. **(10 Marks)**

b. Explain the nearest neighbor classification technique, giving its algorithm. List the characteristics. **(10 Marks)**

4 a. Discuss the Apriary algorithm for frequent item set generation. **(10 Marks)**

b. Discuss the different factors affecting computational complexity of the Apriary algorithm. **(10 Marks)**

PART – B

5 a. Explain the different steps involved in subsequent extension of FP tree. **(10 Marks)**

b. Explain how association patterns are evaluated. **(10 Marks)**

6 a. What is clustering? What are the different types of clusters? Explain the basic K-means cluster algorithm. **(10 Marks)**

b. Perform a hierarchical clustering of five samples using the single-linkage algorithm and two features x and y. Draw the dendogram. **(10 Marks)**

	1	2	3	4	5
x	4	8	15	24	24
y	4	4	8	4	12

7 a. Briefly discuss the approach to mining multidimensional analysis and discipline mining of complex data objects. **(10 Marks)**

b. Write a note on text mining. **(10 Marks)**

8 Write notes on any two of the following : **(20 Marks)**

a. Trends in data mining

b. Social impacts of data mining

c. Data mining applications.

* * * * *