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

06IS74

Seventh Semester B.E. Degree Examination, December 2011
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. What are the challenges in methodology of data mining technology? **(08 Marks)**

b. What defines a data mining task? Explain at least five primitives. **(07 Marks)**

c. Distinguish between categorical and numerical attributes. **(05 Marks)**

2 a. What is Euclidian distance? Write the generalized Minkowski distance metric for different values of r. **(08 Marks)**

b. Explain the similarity and dissimilarity between two objects. **(04 Marks)**

c. What are the issues related to proximity measures? **(08 Marks)**

3 a. Explain hunts algorithm for the inducing decision trees. **(08 Marks)**

b. What are the various characteristics of the decision tree induction? **(07 Marks)**

c. Explain the characteristics of nearest neighbor classifiers. **(05 Marks)**

4 a. Develop the Apriori algorithm for the generating frequent itemset generation. **(08 Marks)**

b. Consider the transaction data set for an super market:

Tid	1	2	3	4	5	6	7	8	9
List of item-ids	I1, I2, I5	I2, I4	I2, I3	I1, I2, I4	I1, I3	I2, I3	I1, I3	I1, I2, I3, I5	I1, I2, I3

Generate all the frequent itemsets. Also generate all the association rules by considering minimum confidence threshold 70% and minimum support thruhold 20%. **(12 Marks)**

PART – B

5 a. Apply Fp-growth algorithm to generate frequent itemset for the figure 1 transaction data set. **(08 Marks)**

b. Write an algorithm to construct conditional Fp-tree, with an example. **(07 Marks)**

c. Write short notes on sequential pattern discovery. **(05 Marks)**

6 a. What is cluster analysis? What are the different types of clusters? **(08 Marks)**

b. Explain the bisecting K-means algorithm to the generate clusters. **(07 Marks)**

c. Compare K-means, with the DBSCAN algorithm. **(05 Marks)**

7 a. Explain the different dimensions in a spatial data mining. **(08 Marks)**

b. Explain the different text mining approaches. **(07 Marks)**

c. Explain the need for mining the world wide web. **(05 Marks)**

8 Write short notes on:

a. Statistical data mining

b. Multimedia data mining

c. Trends in data mining

d. Data mining applications. **(20 Marks)**

* * * * *