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

07MCA542

Fifth Semester MCA Degree Examination, December 2011

Data Mining

Time: 3 hrs. Max. Marks:100

Note: ***Answer any FIVE full questions.***

1 a. Explain data mining. Describe the knowledge discovery process in data mining. **(05 Marks)**

b. Discuss the different types of data sets in data mining. **(10 Marks)**

c. Explain the various types of data mining tasks. **(05 Marks)**

2 a. Define data preprocessing. Mention the approaches of preprocessing data. **(05 Marks)**

b. i) Calculate SMC and Jaccard similarity for the binary vectors :
X = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0) and Y = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1)

ii) Calculate cosine similarity for the data objects :
X = (3, 2, 0, 5, 0, 0, 0, 2, 0, 0) and Y = (1, 0, 0, 0, 0, 0, 0, 1, 0, 2) **(10 Marks)**

c. Discuss the dissimilarities between data objects. **(05 Marks)**

3 a. Explain classification process. Write algorithm for decision tree induction technique. **(10 Marks)**

b. Explain the characteristics of rule base classifier. **(05 Marks)**

c. Write an algorithm for K-nearest neighbor classifier and explain. **(05 Marks)**

4 a. Write apriori algorithm for finding frequent item sets. Find the frequent pattern generated using apriori for the following set and transactions : **(10 Marks)**

TID	List of item - IDs
T_{100}	L_1, L_2, L_5
T_{200}	L_2, L_4
T_{300}	L_2, L_3
T_{400}	L_1, L_2, L_4
T_{500}	L_1, L_3
T_{600}	L_2, L_3
T_{700}	L_1, L_3
T_{800}	L_1, L_2, L_3, L_5
T_{900}	L_1, L_2, L_3

b. Draw F-P tree for the given set of transactions. Find all frequent item sets using F-P growth.

TID	List of item - IDs
T_{100}	{M, O, N, K, E, Y}
T_{200}	{D, O, N, K, E, Y}
T_{300}	{M, A, K, E}
T_{400}	{M, U, C, K, Y}
T_{500}	{C, O, O, K, I, E}

(05 Marks)

c. Write a note on alternate methods for generating frequent item sets. **(05 Marks)**

5 a. Write a note on evaluation of association patterns. **(05 Marks)**

b. Explain the effect of skewed support distribution. **(05 Marks)**

c. Describe the ripper algorithm. **(05 Marks)**

d. Explain merging sequence criteria, with an example. **(05 Marks)**

6 a. Describe cluster analysis. Briefly explain different types of clustering. **(10 Marks)**
b. Write algorithm for density based clustering and explain in brief. **(10 Marks)**

7 a. Explain data mining for biomedical and DNA data analysis. **(10 Marks)**
b. Briefly describe the various trends in data mining. **(10 Marks)**

8 Write short notes on the following :
a. Origins of data mining
b. Association rule mining
c. Dimensionality reduction
d. Spatial data mining. **(20 Marks)**

* * * * *