Fifth Semester MCA Degree Examination, December 2012

Data Mining

Time: 3 hrs.   Max. Marks:100

Note: *Answer any FIVE full questions.*

1   a.   What is data mining? Explain categories of data mining tasks with an example. **(10 Marks)**
    b.   Explain different attribute types with examples. **(08 Marks)**
    c.   What is curse of dimensionality? Explain. **(02 Marks)**

2   a.   What are missing values? Give the strategies for dealing missing data used. **(08 Marks)**
    b.   Explain feature subset selection, preprocessing step involved in data mining. **(08 Marks)**
    c.   Write a note on binarization. **(04 Marks)**

3   a.   Explain the various measures for selecting the best splits in classification with examples. **(08 Marks)**
    b.   Explain direct methods for rule extraction. **(08 Marks)**
    c.   For the following vectors x and y, calculate the indicated similarity measures:
       $x = (1, 1, 0, 1, 0, 1)$ $y = (1, 1, 1, 0, 0, 1)$
      i) Cosine   ii) Jaccard **(04 Marks)**

4   a.   Define and give an example to each: i) support of rule, ii) confidence of rule. **(04 Marks)**
    b.   Write and explain frequent item set generation in the Apriori algorithm. **(10 Marks)**
    c.   Explain support count in Apriori algorithm using a Hash tree. **(06 Marks)**

5   a.   Consider the following transaction data set:

Tid	Items	Tid	items
1	{a, b}	6	{a, b, c, d}
2	{b, c, d}	7	{a}
3	{a, c, d, e}	8	{a, b, c}
4	{a, d, e}	9	{a, b, d}
5	{a, b, c}	10	{b, c, c}

      Construct the FP tree. Show the tree separately after each transaction. **(10 Marks)**
    b.   Explain the effect of skewed support distribution with cross support pattern example. **(06 Marks)**
    c.   Explain criteria for merging sequence with an example. **(04 Marks)**

6   a.   Explain the different types of clustering and clusters. **(10 Marks)**
    b.   Explain K-means clustering method. **(10 Marks)**

7   a.   Explain the generalization of structured data with examples. **(10 Marks)**
    b.   Write a short note on different approaches to text mining. **(10 Marks)**

8   a.   What are features to assess data mining systems? Explain. **(10 Marks)**
    b.   Write a short note on data mining for the telecommunication industry. **(10 Marks)**

* * * * *