DOE
21/12/16.

AUTUMN END SEMESTER EXAMINATION-2016

7th Semester B.Tech

DATA MINING

IT-4003

(Regular-2013 Admitted Batch)

Time: 3 Hours **Full Marks: 60**

Answer any Six questions including question No.1 which is compulsory.
The figures in the margin indicate full marks.
Candidates are required to give their answers in their own words as far as practicable and all parts of a question should be answered at one place only.

1. a) What is data mining? How it is different from databases? [2 × 10

   b) Differentiate between agglomerative and divisive hierarchical clustering.

   c) Is 'The number of variables (or features) used for training an artificial neural network equal to the number of output nodes'? Justify your answer.

   d) How do you choose best split while constructing a decision tree?

   e) Explain the use of classification with an example.

   f) State different types of methods used for data clustering.

   g) What is discrete and continuous data in data mining world?

   h) Define and explain decision tree pruning.

   i) What is supervised learning? How is it different from unsupervised learning?

   j) State a constraint based association rule mining.

2. a) Describe three challenges of data mining methodology and user interaction issues. [4

b) What is Bayes belief network? Give a suitable example and discuss the Bayes belief network. [4

3. a) What are various OLAP operations? How it is different form OLTP? [4

b) What is distributed association rule? Explain with a suitable example. [4

4. Suppose that the data mining task is to cluster the following eight points (with(X, Y) representing location) into three clusters. [4 × 2

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 centre of each cluster, respectively. Use K-means algorithm to show the clusters.

a) The three cluster centers after the first round of execution

b) The final three clusters (with diagrammatic representation)

5. a) What is neural network? Explain the feed-forward and back-propagation of error with a suitable example. [4

b) Explain about basic decision tree induction algorithm of classification. [4

6. a) List and describe the five primitives for specifying a data mining task? [4

b) A database has five transactions. Let min_support=60% and min_confidence=80%. [4

TID	Items_bought
T100	{M,O,N,K,E,Y}
T200	{D,O,N,K,E,Y}
T300	{M,A,K,E}
T400	{M,U,C,K,Y}
T500	{C,O,O,K,I,E}

Find all frequent item sets using Apriority and FP growth respectively, Compare the efficiency of the two mining process.

7. a) What are the difference between classification analysis and predictive analysis? [4

b) What is Genetic Algorithm? What are the steps of GA? How it is applicable for classification? [4

8. Write Short notes (Any four) [4 × 2

a) K-Medoids clustering

b) Temporal mining

c) Web mining

d) Fuzzy clusters

e) Support vector Machine

– ***** –