Or

What K-means Algorithm is for ? Describe its working. Why is it necessary to normalize the attributes before applying k-means algorithm ? Discuss the limitations of K-means algorithm. **11**

Subject Code—0429

M.C.A. (Fifth Year) EXAMINATION

(5 Years Integrated Course)

(Batch 2009 Onwards)

DATA MINING AND DATA WAREHOUSING

MCA-505

Time : 3 *Hours* *Maximum Marks :* 70

Section A

Note : Attempt any *Seven* questions. **7×5=35**

1. Differentiate between data mart, data warehouse and a data mining system.

2. How data mining query language is different than SQL ?

3. What is association rule mining and how is it beneficial for any business ?

4. Describe Bayes theorem. How is it used in predictive modelling ?

5. How can we discover outliers from data ?

6. Give formulas for computing information gain and gain ratio for an attribute ? What advantages does the gain ration has over information gain ?

7. Write five applications of data mining.

8. How do you computer distance between two objects while clustering data ? Assume that all the attributes of the object are the categorical type.

9. What are multi-dimensional association rules ? Give *two* examples.

10. Describe techniques for data normalization.

Section B

Note : Attempt all the questions.

11. What is purpose of the task of classification ? Describe decision tree algorithm in detail.

Or

How do we use back propagation neural networks for the task of classification ? Give algorithm and necessary formulas. You may take help of an example while describing the working of the algorithm. **12**

12. Describe architecture for multi-dimensional data warehouse model.

Or

Explain FP growth algorithm for association rule mining. What are its advantages of FP growth algorithm over Apriori algorithm ? **12**

13. What is the need for cleaning and integrating data before mining information from it ? What are the ways to deal with missing values ?