IV B.Tech. Examinations – August 2017

**DATAWAREHOUSING AND DATAMINING**

Time: **3** hours (CSE) Max. Marks: **60**

# SECTION – A

(Short Answer Questions)

**Answer all ten questions 10×1M=10M**

1. If the number of dimensions are three then total number of cuboids or group bys is\_\_\_
2. If we consider concept hierarchy of dimensions then total number of cuboids\_\_\_\_\_
3. Give an example for single and multi dimensional association rule.
4. In, ---------each value in a bin is replaced by the mean value of the bin.
5. The ---------- operation performs aggregation on a data cube.
6. All nonempty subsets of a --------- must also be frequent.
7. \_\_\_\_\_\_is the learning of decision trees from class – labeled training tuples.
8. Neutral network learning is also referred to as \_\_\_\_\_\_\_\_\_\_

a) Connectionist b) machine learning c) All the above d) none of the above

1. \_\_\_\_ method works by grouping data objects into a hierarchy or ‘tree’ of clusters.

a) Hierarchical clustering b) K-means clustering c) Clustering d) K-nearest neighbor

10. \_\_\_\_\_\_\_is a graph clustering algorithm that searches graph to identify well –connected components as clusters

**SECTION – B**

**Answer all five questions 5×2M= 10M**

1. Explain Z-Score normalization with an example.
2. Identify the measure used to find the number of times an item set is appeared in transactional databases.
3. Write about information gain and gain ratio.
4. Describe the features of partition based clustering algorithms.
5. What is dimensionality reduction?

**SECTION – C**

**Answer all four questions 4×5M = 20M**

1. What is data mining? Briefly explain the Knowledge discovery process.

**(OR)**

1. Explain about classification of Data Mining Systems.
2. Describe briefly about data cube computation.

**(OR)**

19. Explain briefly constraint based association rules.

20. Explain Decision tree induction algorithm with an example.

**(OR)**

1. Explain Multidimensional Analysis and Descriptive Mining of Complex Data object.
2. Distinguish Operational Data base systems and data warehouse.

**(OR)**

1. Describe different schemas used in design of Data warehouse system.

**SECTION – D**

**Answer all two questions 2×10M= 20M**

1. a) What is data cleaning? Describe the approaches to fill missing values.

b) What is the use of data Discritization? Explain entropy based data Discritization.

**(OR)**

1. a) What are the approaches to mining multilevel association rules? Explain.

b) Describe briefly about an association rule mining.

1. a) Define the term cluster analysis along with its requirements.

b) Explain different type of data in cluster analysis.

**(OR)**

27. Explain briefly about K-means and K-medoids partitioning methods.