

MCA DEGREE V SEMESTER EXAMINATION NOVEMBER 2013

CAS 2503/2504/2505 DATA MINING

Time : 3 Hours

Maximum Marks : 50

PART A(Answer *ALL* questions)

(15 x 2 = 30)

- I. (a) List out some specific challenges that motivated the development of data mining.
(b) Discuss major categories of data mining tasks.
(c) Discuss the aspects of data quality that related to data measurement and collection.
- II. (a) Differentiate stem and leaf plot with standard histogram.
(b) What is OLAP?
(c) What is a multidimensional database?
- III. (a) State the characteristics of Rule-Based Classifications.
(b) Briefly explain model over fitting.
(c) Discuss different methods for comparing classifiers.
- IV. (a) State and illustrate the Apriori principle.
(b) How to evaluate association patterns.
(c) State and prove the theorem for the confidence measure.
- V. (a) What is cluster analysis and what are the different types of clustering.
(b) What is DBSCAN.
(c) Discuss different causes of Anomalies.

PART B

(5 x 4 = 20)

- VI. A. Discuss measures of similarity and dissimilarity of data.
OR
B. Discuss the main Data preprocessing methods.
- VII. A. Discuss visualization techniques.
OR
B. Explain summary statistics in detail.
- VIII. A. What is a Rule based classifier and how it works.
OR
B. Explain decision free induction algorithm with an example.
- IX. A. Discuss alternate methods for generating Frequent item sets.
OR
B. Explain FP-Growth algorithm with an example.
- X. A. Discuss Bisecting K-means algorithm with example.
OR
B. Explain Basic Agglomerative hierarchical clustering algorithm.
