

--	--	--	--	--	--	--	--

MCA DEGREE V SEMESTER EXAMINATION DECEMBER 2014

CAS 2503/2504 DATA MINING (Regular & Supplementary)

Time: 3 Hours

Maximum Marks: 50

PART A (Answer *ALL* questions)

(15 × 2 = 30)

- I. (a) "Predicting the outcomes of tossing a fair pair of dice". Is it a data mining task? Why or why not?
(b) What is the type of data required for association analysis? Give example.
(c) Daily temperature show more auto correlation than daily rainfall. Why?
- II. (a) Is random sampling a good approach to sampling? Why or Why not?
(b) What is a stem and leaf plot? Compare it with histogram.
(c) What is multi dimensional data analysis? Explain.
- III. (a) What is Gini Index? Why is it used in data mining?
(b) State and explain principle of parsimony.
(c) How a Naïve Bayes classifier works? Explain.
- IV. (a) Why support and confidence is used for association analysis?
(b) What is the computational complexity of Apriori algorithm?
(c) Explain the relationship among frequent, maximal frequent and closed frequent itemsets.
- V. (a) Give example four data sets in which cosine measure is an appropriate/inappropriate similarity measure. Give reasons.
(b) What are the key issues in hierarchical clustering?
(c) State and explain F-measure.

PART B

(5 × 4 = 20)

- VI. Discuss various proximity measures used in data mining techniques.
OR
- VII. What is data mining and knowledge discovery? Illustrate data mining tasks with examples.
- VIII. What are summary statistics? Explain with examples various summary statistics.
OR
- IX. Discuss visualization techniques for higher dimensional data.
- X. What is a classifier? Explain various methods to compare classifiers.
OR
- XI. Outline methods to build a rule-based classifier.
- XII. Outline Apriori algorithm for frequent itemset generation.
OR
- XIII. Discuss FP-Growth algorithm to discover frequent itemsets.
- XIV. Outline basic K-means algorithm. Why this algorithm is extended? Explain any one extension.
OR
- XV. Discuss various cluster validation schemes.