

Second Semester M.Tech. Degree Examination, May/June 2010

Data Warehousing and Mining

Time: 3 hrs.

Max. Marks:100

Note: Answer any FIVE full questions.

- 1 a. Write any four definitions of a data warehouse. Explain the operation of data warehousing with a block diagram. (10 Marks)
- b. Differentiate between (with examples):
 - i) Informational data and operational data and
 - ii) Technical metadata and business metadata. (10 Marks)
- 2 a. What are the different access tools used in a data warehouse? Explain any three of them. (10 Marks)
- b. Describe :
 - i) Multidimensional data model and
 - ii) MOLAP architecture. (10 Marks)
- 3 a. Compare with examples :
 - i) KDD and DM techniques and
 - ii) Multidimensional and multirelational OLAP. (10 Marks)
- b. Explain with examples:
 - i) Chi-square test
 - ii) Contingency tables and
 - iii) Linear regression. (10 Marks)
- 4 a. Describe Baye's theorem and causality relationship. Let the prior probability of a person having a nerve problem is 0.02, probability of having cancer given that the person has nerve problem is 0.5 and the probability of cancer in the general population is 0.08. Determine the probability that a person has nerve problem given that he/she has cancer. (10 Marks)
- b. Explain the hypothesis testing, with an example. What are the three parameters used for data mining effectiveness? (10 Marks)
- 5 a. What are the characteristics of a decision tree? Explain with a numerical example, the calculation of entropy value and splitting procedure. What are the applications of decision trees? (10 Marks)
- b. Describe the prediction of wireless communication churn with CART. (10 Marks)
- 6 a. Describe the clustering techniques, with examples. When is the nearest neighbor classifier preferred? (10 Marks)
- b. Explain the image recognition technique used for human handwriting case. (10 Marks)
- 7 a. Describe the concept of genetic algorithm. How mutation and cross over occurs? Explain with an example. (10 Marks)
- b. What is the optimization criteria used in genetic algorithm? How score board is used in genetic algorithm? Explain. (10 Marks)
- 8 a. Describe the traveling salesman problem and the cost minimization procedure. (10 Marks)
- b. Describe the optimization procedure for predicting customer segments, using genetic algorithm. (10 Marks)