35153

M.Sc. DEGREE EXAMINATION, NOVEMBER 2017.

THIRD SEMESTER

Statistics

Paper III — KNOWLEDGE DISCOVERY AND DATA MINING

Time : Three hours Maximum : 75 marks

(No additional sheet will be supplied)

PART A — (5 × 3 = 15 marks)

Answer any FIVE questions.

Each question carries 3 marks.

Each answer should not exceed 1 page.

1. What is a decision tree? How do you construct it?

2. Distinguish between binary classification and clustering of objects.

3. Mention various methods of clustering basing on a single continuous variable.

4. What is meant by dimension reduction? What are the advantages?

5. Write about Bayesian method for binary classification.

6. Mention how linear regression helps in discrimination between two groups.

7. What is a relational database? Give some examples.

8. Write a brief note on data warehouses with an example.

PART B — (4 × 15 = 60 marks)

Answer ALL questions.

Each question carries 15 marks.

Each answer should not exceed 6 pages.

9. Explain the method use for training a classification model. How do you determine the size of the training, validation and testing data sets?

Or

10. Mention the use of logistic regression for binary classification using k-independent features.

11. What is hierarchical clustering of cases basing on multivariate data? What is a dendrogram?

Or

12. Explain how principle component analysis helps in reducing the dimension of a data set.

13. Write about linear discriminant analysis to handle a dichotomous classification problem.

Or

14. What is a regression tree? How do you compare it with a decision tree?

15. What is a multidimensional data cube? How is the data processed in it?

Or

16. Discuss the methods used in OLAP and mention its uses.