Question Paper Code : 41145

B.E./B.Tech. DEGREE EXAMINATION, NOVEMBER/DECEMBER 2015.

Seventh Semester

Computer Science and Engineering

CS 1011 — DATA WAREHOUSING AND MINING

(Common to information Technology)

(Regulation 2008)

Time : Three hours   Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. State why pattern evaluation is included as a step in the process of knowledge discovery from databases.

2. What does a time-series database store?

3. Are all patterns generated are interesting and useful? Give reasons to justify.

4. Consider the following set of data X = {15, 27, 62, 35, 39, 50, 44, 44, 22, 98}

    Do preprocessing using smoothing by bin means and bin boundary to smooth the data, using a bin of depth 3.

5. What is association rule mining?

6. List the two interesting measures of an association rule.

7. State the difference between classification and clustering.

8. What is outlier analysis?

9. Write the difference between time-series database and sequence database.

10. List out various commercial data mining tools.

PART B — (5 × 16 = 80 marks)

11. (a) (i) Diagrammatically illustrate and discuss the steps in the process of knowledge discovery. (10)

    (ii) Explain why a data warehouse is well equipped for providing the data for data mining. (6)

Or

    (b) Diagrammatically illustrate and explain the architecture of a data warehouse.

12. (a) (i) Explain how sampling can be used for data reduction with an example. (8)

    (ii) Explain smoothing by · bin means, smoothing by bin medians and smoothing by bin boundaries with examples. (8)

Or

    (b) Explain the different forms of data preprocessing with examples and diagrammatic illustrations.

13. (a) Write and explain the algorithm for minig frequent itemsets without candidate generation.

Or

    (b) With an example, explain the approaches to mine multi-level association rules.

14. (a) State Baye's theorem of posterior probability and explain the working of a Bayesian classifier with an example.

Or

    (b) What is clustering? Explain K-means partitioning algorithm with an example.

15. (a) Explain the various types of mining the text databases and worldwide web. What kind of preprocessing are necessary in mining the text and. web? State the importance of preprocessing the text and web before mining.

Or

    (b) (i) Discuss multimedia database and temporal database. (5+5)

    (ii) What are the significance of spatial data mining? What kind of applications need this type of mining and why? (6)