1

IV B.Tech I Semester(R05) Regular/Supplementary Examinations, December 2009
DATA WAREHOUSING AND DATA MINING
(Computer Science & Engineering)

Time: 3 hours

Max Marks: 80

Answer any FIVE Questions
All Questions carry equal marks

KOTTAM
INSTITUTIONS

1. (a) Explain the architecture of a typical data mining system

   (b) Discuss the issues regarding data warehouse architecture.   [8+8]

2. Suppose that the data for analysis include the attribute age. The age values for the data tuples are (in increasing order):
   13,15,16,16,19,20,20,21,22,22,25,25,25,25,30,33,33,35,35,35,35,36,40,45,46, 52,70.

   (a) Use smoothing by bin means to smooth the above data, using a bin depth of 3. Illustrate your steps. Comment on the effect of the technique for the given data.

   (b) How might you determine outliers in the data?

   (c) What other methods are there for data smoothing?

   [16]

3. Explain the syntax for the following data mining primitives:

   (a) Task-relevant data

   (b) The kind of knowledge to be mined

   (c) Interestingness measures

   (d) Presentation and visualization of discovered patterns.   [16]

4. (a) What are the differences between concept description in large data bases and OLAP?

   (b) Explain about the graph displays of basic statistical class description.   [8+8]

5. (a) Discus about Association rule mining.

   (b) Define multidimensional Association rule. Discuss mining distance-based Association rules.   [8+8]

6. Explain Decision tree induction classification.   [16]

7. (a) Given the following measurement for the variable age:
   16, 25, 28, 46, 29, 44, 38, 37, 54, 27
   Standardize the variable by the following:

        i. Compute the mean absolute deviation of age.

        ii. Compute the Z-score for the first four measurements.

   (b) Explain clustering using representatives algorithm with example.

   (c) Write an algorithm for DBSCAN and give an example of DBSCAN.   [4+4+4+4]

8. (a) Explain the following:

        i. Generalization of structured data

        ii. Generalization of object identifiers and class/subclass hierarchies

        iii. Generalization of class composition hierarchies.

   (b) Explain mining multimedia databases.   [3+3+3+7]

IV B.Tech I Semester(R05) Regular/Supplementary Examinations, December 2009
DATA WAREHOUSING AND DATA MINING
(Computer Science & Engineering)

Time: 3 hours

Max Marks: 80

Answer any FIVE Questions
All Questions carry equal marks

KOTTAM
INSTITUTIONS

1. (a) Explain data mining as a step in the process of knowledge discovery.

   (b) Differentiate operational database systems and data warehousing. [8+8]

2. Explain various data reduction techniques. [16]

3. (a) Briefly discuss about Task-relevant data specification.

   (b) Explain the syntax for Task-relevant data specification. [8+8]

4. (a) What is Concept description? Explain.

   (b) What are the differences between concept description in large data bases and OLAP? [8+8]

5. (a) Explain about constraint-based Association mining.

   (b) Give an example for Association rule mining? Classify Association rules. [8+8]

6. (a) Why naive Bayesian classification called "naive"? Briefly outline the major ideas of naive Bayesian classi-fication.

   (b) Define regression. Briefly explain about linear, non-linear and multiple regressions. [8+8]

7. (a) Define nominal, ordinal, and ratio-scaled variables.

   (b) Discuss about Classical partitioning methods. [2+2+2+10]

8. An e-mail database is a database that stores a large number of electronic mail messages. It can be viewed as a semistructured database consisting mainly of text data. Discuss the following.

   (a) How can such an e-mail database be structured so as to facilitate multi-dimensional search, such as by sender, by receiver, by subject, by time, and so on?

   (b) What can be mined from such an e-mail database?

   (c) suppose you have roughly classified a set of your previous e-mail messages as junk, unimportant, normal, or important. Describe how a data mining system may take this as the training set to automatically classify new e-mail messages or unclassified ones.[5+5+6]

IV B.Tech I Semester(R05) Regular/Supplementary Examinations, December 2009
DATA WAREHOUSING AND DATA MINING
(Computer Science & Engineering)

Time: 3 hours

Max Marks: 80

Answer any FIVE Questions
All Questions carry equal marks

KOTTAM
INSTITUTIONS

1. (a) Discuss about Concept hierarchy.

   (b) Briefly explain about - classification of database systems.   [8+8]

2. (a) Briefly discuss about data integration.

   (b) Briefly discuss about data transformation.   [8+8]

3. The four major types of concept hierarchies are: schema hierarchies, set-grouping hierarchies, operation-derived hierarchies, and rule-based hierarchies.

   (a) Briefly define each type of hierarchy.

   (b) For each hierarchy type, provide an example.   [16]

4. (a) How can class comparison descriptions be visualized? Explain.

   (b) Explain about incremental and parallel mining of concept description.   [8+8]

5. Compare and contrast the differences between mining single dimensional Boolean Association rules and multilevel Association rules for transactional databases.[16]

6. (a) Explain the following classification methods

        i. k-Nearest neighbor classifiers

       ii. case-based reasoning

      iii. rough set approach

       iv. fuzzy set approaches.

   (b) Explain classifier accuracy.   [3+3+2+2+6]

7. (a) Explain the categorization of major clustering methods.

   (b) Write CURE algorithm.

   (c) Explain DBSCAN and OPTICS with examples.   [6+4+3+3]

8. (a) What are different approaches for similarity-based retrieval in image databases?

   (b) Define similarity search. Explain similarity search in time-series analysis.

   (c) Write a note on mining the World Wide Web.   [4+6+6]

4

IV B.Tech I Semester(R05) Regular/Supplementary Examinations, December 2009
DATA WAREHOUSING AND DATA MINING
(Computer Science & Engineering)

Time: 3 hours   Max Marks: 80

Answer any FIVE Questions
All Questions carry equal marks

KOTTAM
INSTITUTIONS

1. Briefly compare the following concepts. Use an example to explain your points.

   (a) Snowflake schema, fact constellation, starnet query model.

   (b) Data cleaning, data transformation, refresh.

   (c) Discovery driven cube, multifeature cube, and virtual warehouse.   [16]

2. Write short note on the following data reduction techniques:

   (a) Data cube aggregation.

   (b) Concept hierarchy generation for categorical data.   [16]

3. (a) Describe why is it important to have a data mining query language.

   (b) The four major types of concept hierarchies are: schema hierarchies, set-grouping hierarchies, operation-derived hierarchies, and rule-based hierarchies-Briefly define each type of hierarchy.[8+8]

4. (a) How can class comparison descriptions be visualized? Explain.

   (b) Explain the measures of central tendency in detail.   [8+8]

5. Explain the following:

   (a) Generating association rules from frequent item sets.

   (b) FP-tree

   (c) ARCS.   [4+6+6]

6. (a) Explain decision tree induction classification.

   (b) Describe backpropagation classification.   [8+8]

7. (a) What are the categories of major clustering methods? Explain.

   (b) Explain about outlier analysis.   [6+10]

8. Explain the following:

   (a) Mining the Text databases

   (b) Mining Time-series and sequence data.   [8+8]