

Code No: 07A70503

**R07****Set No. 2**

**IV B.Tech I Semester Examinations, December 2010**  
**DATA WAREHOUSING AND DATA MINING**  
**Computer Science And Engineering**

**Time: 3 hours****Max Marks: 80**

**Answer any FIVE Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

1. (a) Explain FP-Growth with example.  
 (b) What are the approaches to mining multilevel association rules? Explain. [8+8]
2. (a) Describe the data classification process with a neat diagram.  
 (b) Discuss about Bayesian classification. [8+8]
3. Explain the following:  
 (a) Spatial association analysis  
 (b) Sequential pattern mining  
 (c) Latent semantic indexing  
 (d) Term frequency matrix. [4+4+4+4]
4. (a) Justify the role of data cube aggregation in data reduction process with an example.  
 (b) Discuss the methods for numeric concept hierarchy generation. [16]
5. (a) What is data mining? What is data warehousing? Give their applications.  
 (b) Briefly discuss data warehouse architecture. [8+8]
6. (a) Write and explain the basic algorithm for Attribute-oriented induction.  
 (b) What are the differences between concept description in large data bases and OLAP? [8+8]
7. (a) Given two objects represented by the tuples (22,1,42,10) and (20,0,36,8):  
 i. Compute the Euclidean distance between the two objects.  
 ii. Compute the Manhattan distance between the two objects.  
 iii. Compute the Minkowski distance between the two objects, using  $q=3$ .  
 (b) Explain about Statistical-based outlier detection and Deviation-based outlier detection. [16]
8. 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]

\*\*\*\*\*

Code No: 07A70503

**R07****Set No. 4**

**IV B.Tech I Semester Examinations, December 2010**  
**DATA WAREHOUSING AND DATA MINING**  
**Computer Science And Engineering**

**Time: 3 hours****Max Marks: 80**

**Answer any FIVE Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

1. (a) Write the Backpropagation algorithm and explain.  
 (b) Describe multiple, linear and non-linear regressions. [8+8]
2. Write the syntax for the following data mining primitives:  
 (a) Task-relevant data.  
 (b) Concept hierarchies. [16]
3. (a) Define nominal, ordinal, ratio-scaled variables.  
 (b) Explain about density-based methods. [8+8]
4. (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]
5. (a) Write the algorithm for attribute-oriented induction. Explain the steps involved in it.  
 (b) How can concept description mining be performed incrementally and in a distributed manner? [8+8]
6. Write a short note on following:  
 (a) Missing values  
 (b) Noisy data  
 (c) Inconsistent data  
 (d) Data cube aggregation. [16]
7. (a) Briefly discuss the data mining functionalities.  
 (b) Briefly discuss the major issues in data mining regarding performance and diverse database types. [8+8]
8. Propose and outline a level shared mining approach to mining multilevel association rules in which each item is encoded by its level position, and initial scan of the database collects the count for each item at each concept level, identifying frequent and sub frequent items. Comment on the processing cost of mining multilevel associations with this method in comparison to mining single level associations. [16]

\*\*\*\*\*

Code No: 07A70503

**R07****Set No. 1**

**IV B.Tech I Semester Examinations, December 2010**  
**DATA WAREHOUSING AND DATA MINING**  
**Computer Science And Engineering**

**Time: 3 hours****Max Marks: 80**

**Answer any FIVE Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

1. (a) What is Concept description? Explain.  
 (b) What are the differences between concept description in large data bases and OLAP? [8+8]
2. (a) Write the FP-growth algorithm. Explain.  
 (b) Discuss about ARCS. [10+6]
3. (a) List and describe any four primitives for specifying a data mining task.  
 (b) Describe why concept hierarchies are useful in data mining. [8+8]
4. (a) Give an example of generalization-based mining of plan databases by divide-and-conquer.  
 (b) What is sequential pattern mining? Explain.  
 (c) Explain the construction of a multilayered web information base. [8+4+4]
5. (a) Write k-Means and k-Medoids algorithms.  
 (b) Explain COBWEB model. [8+8]
6. (a) Briefly discuss the role of Data compression in data reduction process.  
 (b) Briefly discuss the role of Numerosity reduction in data reduction process. [8+8]
7. The following table shows a set of paired data where X is the number of years of work experience of a college graduate and Y is the corresponding salary of the graduate.

X	Y
Years experience	Salary (in \$ 1000s)
3	30
8	57
9	64
13	72
3	36
6	43
11	59
21	90
1	20
16	83

Code No: 07A70503

**R07**

**Set No. 1**

- (a) Plot the data. Do X and Y seem to have a linear relationship?
  - (b) Use the method of least squares to find an equation for the prediction of the salary of a college graduate with some years of experience. [8+8]
8. (a) Describe the challenges to data mining regarding performance issue.
- (b) What are the differences between the three main types of data warehouse usage: Information processing, Analytical processing, and data mining? Discuss the motivation behind OLAP mining. [8+8]

\*\*\*\*\*

FIRSTRANKER

Code No: 07A70503

**R07****Set No. 3**

**IV B.Tech I Semester Examinations, December 2010**  
**DATA WAREHOUSING AND DATA MINING**  
**Computer Science And Engineering**

**Time: 3 hours****Max Marks: 80**

**Answer any FIVE Questions**  
**All Questions carry equal marks**

\*\*\*\*\*

1. (a) Briefly discuss the data smoothing techniques.  
 (b) Explain about concept hierarchy generation for categorical data. [8+8]
2. Write the syntax for the following data mining primitives:  
 (a) The kind of knowledge to be mined.  
 (b) Measures of pattern interestingness. [16]
3. (a) Categorize major clustering methods.  
 (b) Explain OPTICS algorithm.  
 (c) What is an outlier? Why is Outlier mining important? Briefly discuss about statistical-based outlier detection. [4+4+8]
4. (a) How can we perform attribute relevant analysis for concept description? Explain.  
 (b) Briefly explain about the presentation of class comparison descriptions. [8+8]
5. (a) How scalable is decision tree induction? Explain.  
 (b) Explain about prediction. [8+8]
6. (a) What is information retrieval? What methods are there for information retrieval?  
 (b) What is sequential pattern mining? Explain.  
 (c) Discuss about mining the webs link structures to identify authoritative web pages. [4+6+6]
7. Propose a method for mining hybrid-dimension association rules (multidimensional association rules with repeating predicates) and explain with an example. [16]
8. What is data mining? In your answer, address the following:  
 (a) Is it another type?  
 (b) Is it a simple transformation of technology developed from databases, statistics, and machine learning?  
 (c) Explain how the evolution of database technology led to data mining.  
 (d) Describe the steps involved in data mining when viewed as a process of knowledge discovery? [16]

\*\*\*\*\*