



Department of Computer Science & Engineering

QUESTION BANK

Subject: DWDM
Branch: III CSE I Semester

Faculty Name: B. Bala Krishna

UNIT-I

- Briefly discuss the data smoothing techniques.
 - 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.
 - 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.
 - How might you determine outliers in the data?
 - What other methods are there for data smoothing? [R07, Dec, 2011, 16M, set 2]
- Briefly discuss the Discretization and concept hierarchy techniques. [R07, Dec, 2011, 16M, set 4]
- Write short note on the following architectures of data mining systems:
 - No coupling
 - Loose coupling
 - Semitight coupling
 - Tight coupling. [R07, Dec, 2011, 16M, set 3]
- List and describe any four primitives for specifying a data mining task.
 - Write about Semitight coupling and Loose Coupling. Differentiate them. [R07, Dec, 2011, 16M, set 4]
- How can you go about filling in the missing values in data cleaning process?
 - Discuss the data smoothing techniques. [R07, Dec, 2011, 16M, set 1]
- Explain the architecture of a typical data mining system. [R07, Dec, 2011, 8M, set 4]
- Write the syntax for the following data mining primitives:
 - The kind of knowledge to be mined.
 - Measures of pattern interestingness. [R07, Dec, 2011, 8M, set 2]
- Explain various data reduction techniques. [R05, Dec, 2011, 16 M, set 2]
- Briefly discuss the forms of Data preprocessing with neat diagram.
 - Explain about concept hierarchy generation for categorical data. [R05, March, 2010, 8+8M, set 1]
- Draw and explain the architecture of typical data mining system.
 - Differentiate OLTP and OLAP. [R05, March, 2010, 8+8M, set 2]
- Briefly discuss the Discretization and concept hierarchy techniques. [R05, March, 2010, 16M, set 2]
- Explain data mining as a step in the process of knowledge discovery. [R05, March, 2010, 8M, set 3]
- 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.
 - What is the mean of the data?
 - What is the median?
 - What is the mode of the data? Comment on the data's modality.
 - What is the mid range of the data?
 - Can you find (roughly) the first quartile(Q1), and third quartile(Q3) of the data?
 - Give the five number summaries of the data.
 - Show a box plot of the data.
 - How is the quantile-quantile plot different from a quantile plot? [R05, March, 2010, 16M, set 3]
- Given the following measurement for the variable age: 16, 25, 28, 46, 29, 44, 38, 37, 54, 27 Standardize the variable by the following:
 - Compute the mean absolute deviation of age.
 - Compute the Z-score for the first four measurements. [R05, March, 2010, 4+4 M, set 3]
- How can we smooth out noise in data cleaning process? Explain.
 - Why preprocessing of data is needed? [R05, March, 2010, 4+4 M, set 4]

UNIT – II

- (a) Explain data mining as a step in the process of knowledge discovery.
(b) Differentiate operational database systems and data warehousing. [R07, Dec,2011, 16M, set 2]
- (a) Explain the design and construction process of data warehouses.
(b) Explain the architecture of a typical data mining system. [R07, Dec,2011, 16M, set 4]
- Briefly discuss about the following data warehouse implementation methods: (a) Indexing OLAP data
(b) Metadata Repository. [R07, Dec,2011, 16M, set 1]
- Explain the design and construction process of data warehouses. [R07, Dec,2011, 8M, set 4]
- (a) Explain data mining as a step in the process of knowledge discovery. (b) Differentiate operational database systems and data warehousing. [R05, Dec,2011, set 2; March,2010, 8 +8 M, set 1]
- (a) What is Concept description? Explain. [R05, Dec,2011, 8+8 M, set 2]
(b) What are the differences between concept description in large data bases and OLAP?
- Differentiate operational database systems and data warehousing. [R05, March,2010, 8M, set 3]
- Explain clustering using representatives algorithm with example.
(b) Write an algorithm for DBSCAN and give an example of DBSCAN [R05, March,2010, 4+4 M, set 3]

UNIT - III

- (a) Explain about multilevel Association rules from transaction databases.
(b) What are the steps involved in Association rule clustering system? Explain. [R07, Dec,2011, 16M, set 1]
- (a) Why naive Bayesian classification called 'naive'? Briefly outline the major ideas of naive Bayesian classification.
(b) Define regression. Briefly explain about linear, non-linear and multiple regressions.[R07, Dec,2011, 16M, set1]
- Write the FP-growth algorithm for discovering frequent item sets without candidate generation. Explain an example. [R07, Dec,2011, 16M, set 4]
- (a) Discuss about mining frequent item sets without candidate generation.
(b) Explain about multidimensional Association rules in detail. [R07, Dec,2011, 8+8M, set 2]
- (a) Discuss about Association rule mining. [R05, March,2010, 16M, set 2]
(b) What are the approaches for mining multilevel Association rules? Explain.
- Explain the Apriori algorithm with example. [R05, March,2010, 16M, set 3]
- (a) Write the FP-growth algorithm. Explain.
(b) Discuss about ARCS. [R05, March,2010, 16M, set 4]
- (a) Discuss about mining frequent item sets without candidate generation.
(b) Explain about multidimensional Association rules in detail. [R07, Dec,2011,16M, set 2]

UNIT – IV

- Suppose that you are to allocate a number of automatic teller machines (ATMs) in a given region so as to satisfy a number of constraints. Households or places of work may be clustered so that typically one ATM is assigned per cluster. The clustering, however, may be constrained by factors involving the location of bridges, rivers, and highways that can affect ATM accessibility. Additional constraints may involve limitations on the number of ATMs per district forming the region. Given such constraints, how can clustering algorithms be modified to allow for constraint- based clustering? [R07, Dec,2011, 16M, set 1]
- (a) Can any ideas from association rule mining be applied to classification? Explain.
(b) Explain training Bayesian belief networks.(c) How does tree pruning work? What are some enhancements to basic decision tree induction? [R07, Dec,2011, 6+5+5M, set 2]
- (a) How scalable is decision tree induction? Explain.
(b) Explain about prediction. [R07, Dec,2011, 8+8 M, set 4]
- (a) Which algorithm is an influential algorithm for mining frequent item sets for Boolean association rules? Explain.
(b) What are additional rule constraints to guide mining? Explain. [R05, Dec,2011, 8+8 M, set 2]
- (a) Give the algorithm to generate a decision tree from the given training data.
(b) Explain the concept of integrating data warehousing techniques and decision tree induction.
(c) Describe multilayer feed-forward neural network. [R05, Dec,2011, 8+4+4 M, set 2]

6. (a) Explain about iceberg queries with example.
 (b) Can we design a method that mines the complete set of frequent item sets without candidate generation? If yes, explain with example. **[R05, March,2010, 16M, set 1]**
7. The following table consists of training data from an employee database. The data have been generalized. For a given row entry, count represents the number of data tuples having the values for department, status, age, and salary given in that below:

Department	status	age	salary	count
Sales	Senior	31...35	46K...50K	30
Sales	Junior	26...30	26K...30K	40
Sales	Junior	31...35	31K...35K	40
Systems	Junior	21...25	46K...50K	20
Systems	Senior	31...35	66K...70K	5
Systems	Junior	26...30	46K...50K	3
Systems	Senior	41...45	66K...70K	3
Marketing	Senior	36...40	46K...50K	10
Marketing	Junior	31...35	41K...45K	4
Secretary	Senior	46...50	36K...40K	4
Secretary	Junior	26...30	26K...30K	6

Given a data sample with the values “systems”, “junior;”, and “26...30” for the attributes department, status, and age, respectively, what would a naive Bayesian classification of the salary for the sample be?

[R05, March,2010, 16M, set 1]

8. (a) Describe the data classification process with a neat diagram.
 (b) Discuss about Bayesian classification. **[R05, March,2010, 16M, set 2]**
9. What is Backpropagation? Explain Backpropagation classification. **[R05, March,2010, 16M, set 3]**
 (a) What is classification? What is prediction?
 (b) What is Bayes theorem? Explain about Naive Bayesian classification. **[R05, March,2010, 4+6M, set 4]**

UNIT – V

1. (a) What is Cluster Analysis? What are some typical applications of clustering? What are some typical requirements of clustering in data mining?
 (b) Discuss about model-based clustering methods. **[R07, Dec,2011, 16M, set 4]**
2. (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.
[R05, Dec,2011, 3+3+4+3+3 M, set 2]
3. (a) Define nominal, ordinal, and ratio-scaled variables.
 (b) Discuss about Classical partitioning methods. **[R05, March,2010, 2+2+2+10 M, set 1]**
4. (a) What is Cluster Analysis? What are some typical applications of clustering? What are some typical requirements of clustering in data mining? (b) Define data matrix and dissimilarity matrix. Discuss about interval scaled variables. **[R05, March,2010, 2+2+5+3+4M, set 2]**
5. Write a short note on following: (a) Missing Values. (b) Histogram analysis (c) Entropy-based discretization (d) Segmentation by natural partitioning. **[R05, March,2010, 16M, set 3]**
6. Discuss about k-Nearest neighbor classifiers and case-based reasoning. **[R05, March,2010, 6M, set 4]**

UNIT - VI

1. (a) Discuss about mining frequent item sets without candidate generation.
 (b) Explain about multidimensional Association rules in detail. **[R07, Dec,2011, 16M, set 2]**
2. Write the syntax for the following data mining primitives:
 (a) The kind of knowledge to be mined. (b) Measures of pattern interestingness. **[R07, Dec,2011, 16M, set 2]**

3. (a) Briefly discuss about Task-relevant data specification.
(b) Explain the syntax for Task-relevant data specification. [R05, Dec,2011, 3+3+4+3+3 M, set 2]

UNIT – VII

1. (a) How can we perform discrimination between different classes? Explain.
(b) Explain the analytical characterization with an example. [R07, Dec,2011, 16M, set 1]
2. (a) What is spatial data warehouse? What are the different types of dimensions in a spatial data cube? What are the different types of measures in a spatial data cube? (b) What is keyboard-based association analysis? How can automated document classification be performed?
(c) Briefly discuss about mining the World Wide Web. [R07, Dec,2011, 2+2+2+2+2+6M, set 2]
3. Explain the following: (a) Mining spatial databases (b) Mining the World Wide Web. [R07, Dec, 2011, 16M, set 1]
4. (a) How can object identifiers be generalized, if their role is to uniquely identify objects? Can inherited properties of objects be generalized. (b) What kinds of association can be mined in multimedia data? Explain.
(c) Describe similarity search in time-series analysis. [R07, Dec, 2011, 4+6+6M, set 1]
5. (a) Explain spatial data cube construction and spatial OLAP.
(b) Discuss about mining text databases. [R05, March,2010, 8+8 M, set 1]
6. (a) How to mine Multimedia databases? Explain.
(b) Define web mining. What are the observations made in mining the Web for effective resource and knowledge discovery?(c) What is web usage mining? [R05, March,2010, 10+4+2M, set 2]

UNIT – VIII

1. Write short notes for the following in detail:
(a) Attribute-oriented induction.
(b) Efficient implementation of Attribute-oriented induction. [R07, Dec, 2011, 4+6+6M, set 4]
2. (a) Discuss the importance of establishing a standardized data mining query language. What are some of the potential benefits and challenges involved in such a task?
(b) How can we standardize data mining primitives? [R07, Dec, 2011, 4+6+6M, set 1]
3. (a) What is information retrieval? What methods are there for information re-trieval?
(b) What is sequential pattern mining? Explain.(c) Discuss about mining the webs link structures to identify authoritative web pages. [R05, Dec,2011, 4+6+6 M, set 4]