



M.E DEGREE EXAMINATIONS: MAY 2018

(Regulation 2015)

Second Semester

COMPUTER SCIENCE AND ENGINEERING

P15CST201: Data Warehousing and Data Mining

COURSE OUTCOMES

CO1: Explain the concepts of Data Warehousing architecture and implementation.

CO2: Apply the association rules for mining applications.

CO3: Discuss on appropriate Classification/ Clustering techniques for various problems with high dimensional data.

CO4: Discover the knowledge imbibed in the high dimensional system.

CO5: Illustrate various data mining techniques on complex data objects.

Time: Three Hours

Maximum Marks: 100

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

1. The various aspects of data mining methodologies is/are _____ CO1 [K₂]
 - i) Mining various and new kinds of knowledge
 - ii) Mining knowledge in multidimensional space
 - iii) Pattern evaluation and pattern or constraint-guided mining.
 - iv) Handling uncertainty, noise, or incompleteness of data

a) i, ii and iv only	b) ii, iii and iv only
c) i, ii and iii only	d) i, ii, iii and iv

2. _____ is used to map a data item to a real valued prediction variable. CO1 [K₁]

a) Regression	b) Prediction
c) Time series analysis	d) Classification

3. Data that are not of interest to the data mining task is called as _____ CO2 [K₁]

a) Relevant data	b) Irrelevant data
c) Frequent data	d) Outlier data

4. Match list I with list II and select the correct answer from the table given below.

CO2 [K₂]

List I	List II
A. Frequent set	i. The attributes within a given entity are fully dependent on the entire primary key of the entity.
B. Sampling	ii. In web mining, this is used to know which URLs tend to be requested together.
C. Association rule mining	iii. All set of items whose support is greater than the user-specified minimum support
D. Functional dependency of data	iv. Used to enable rapid election polling

- | | A | B | C | D |
|----|----|-----|-----|----|
| a) | i | Ii | iii | iv |
| b) | ii | Iii | iv | i |
| c) | iv | Iii | ii | i |
| d) | ii | Iv | iii | i |

5. Assertion (A): True positives refer to the positive tuples that were correctly labeled by the classifier

CO3 [K₂]

Reason (R): False negatives are the negative tuples that were incorrectly labeled by the classifier

- | | |
|---|---|
| a) both A and R are individually true and R is the correct explanation of A | b) both A and R are individually true but R is not the correct explanation of A |
| c) A is true but R is false | d) A is false but R is true. |

6. _____ clustering techniques starts with all records in one cluster and then try to split that cluster into small pieces.

CO3 [K₂]

- | | |
|----------------|------------------|
| a) Divisive | b) Agglomerative |
| c) Partitional | d) Regression |

7. Suppose while performing DBSCAN we randomly choose a point which has less than MinPts (i.e., number of points in its neighborhood). which among the following is true for such a point?

CO4 [K₂]

- | | |
|--|---|
| a) It is treated as noise, and not considered further in the algorithm | b) It becomes part of its own cluster |
| c) Depending upon other points, it may later turn out to be a core point | d) Depending upon other points, it may be density connected to other points |

8. Assertion (A) : Unsupervised training methods are called as learning with a teacher
Reason (R) : Unsupervised training may be used when a clear link between input data sets and target output values do not exist. CO4 [K₂]
- a) both A and R are individually true and R is the correct explanation of A
b) both A and R are individually true but R is not the correct explanation of A
c) A is true but R is false
d) A is false but R is true.
9. _____ stores sequence of values or events obtained over repeated measurements of time. CO5 [K₂]
- a) Temporal Database
b) Time-series database
c) Sequence database
d) Spatial database
10. Sequence the k-means algorithm steps: CO5 [K₂]
- i. Take any random object as the initial centroid
ii. Determine the distance of each object to the centroid points
iii. Group the object based on minimum distance
iv. Determine k number of clusters
- a) i, iii, ii and iv
b) i, ii, iii and iv
c) vi, ii, i and iii
d) iv, i, ii and iii

PART B (10 x 2 = 20 Marks)

11. Define data cube. CO1 [K₃]
12. Distinguish between Data Warehouse and Data Mart. CO1 [K₃]
13. List the contents of dimension table. CO2 [K₃]
14. What is meant by concept hierarchy? CO2 [K₂]
15. Using Equi-depth binning method, partition the data given below into 4 bins and perform smoothing by bin boundaries. CO3 [K₃]
- 24,25,26,27,28,56,67,70,70,75,78,89,89,90,91,94,95,96,100,102,103,107,109,112
16. What is data discretization? Give an example. CO3 [K₃]
17. Define gain ratio. CO4 [K₂]
18. Write any four data mining applications. CO4 [K₂]
19. List out some social impact of Data mining. CO5 [K₃]
20. State the goals of time series analysis. CO5 [K₂]

PART C (6 x 5 = 30 Marks)

21. How are concept hierarchies useful in OLAP? Explain OLAP operations in the multidimensional data model. CO1 [K₃]
22. How data warehouses systems different from operational database systems and how are they similar? CO2 [K₃]

- | | | |
|-----|---|-----------------------|
| 23. | What are the essential difference between association rules and decision rules? | CO3 [K ₂] |
| 24. | Give an account of Data Mining Query Language (DMQL). | CO4 [K ₃] |
| 25. | Write the Back Propagation Neural Network training algorithm | CO4 [K ₂] |
| 26. | Briefly explain the text data analysis and information retrieval | CO5 [K ₂] |

**Answer any FOUR Question:
PART D (4 x 10 = 40 Marks)**

- | | | |
|-----|--|-----------------------|
| 27. | Suppose that a data warehouse consists of the three dimensions data, spectator, location, and game, and the two measures count and charge, where charge is the fare that a spectator pays when watching a game on a given data. Spectators may be students, adults, or seniors, with each category having its own charge rate.
(1) Draw a star schema diagram for the data warehouse. (5)
(2) Starting with the base cuboid [data, spectator, location, game], what specific OLAP operations should one perform in order to list the total charge paid by student spectators GM_Place in 2000. (5) | CO1 [K ₄] |
| 28. | Explain the three tier data warehousing architecture with a neat diagram. | CO2 [K ₃] |
| 29. | What is web mining? Differentiate between web content mining, web structure mining, and web usage mining. | CO5 [K ₃] |
| 30. | Discuss in detail about the following classifiers:
(i) Naïve Bayesian Classifier Algorithm. (5)
(ii) Genetic Algorithm. (5) | CO4 [K ₃] |
| 31. | Explain the algorithm for constructing a decision tree from training samples. | CO3 [K ₃] |
