

**NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI-15**  
**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**  
**B. Tech DEGREE VI SEMESTER, I CYCLE TEST, MARCH-2021**  
**CSPE14 – Data Warehousing and Data Mining**

**DATE: 02-03-2021    TIME: 03.30 p.m. - 04.30 p.m.    MAX.MARKS:20 marks**

**Answer all Questions**

**5 x 4 =20 marks**

1. Explain the difference and similarity between discrimination and classification, between characterization and clustering, and between classification and regression.
2. Suppose that a data warehouse consists of the four dimensions, *date*, *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 date. Spectators may be students, adults, or seniors, with each category having its own charge rate.
  - (a) Draw a *snow flake schema* diagram for the data warehouse.
  - (b) Write the DMQL for the corresponding schema.
3. Suppose that the data for analysis includes 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 and bin boundaries to smooth the data, using a bin depth of 3.
  - b) Normalize the data using min –max normalization for the first six data points.
4. Suppose we have the following two-dimensional data set:

|    | A1  | A2  |
|----|-----|-----|
| X1 | 1.5 | 1.7 |
| X2 | 2   | 1.9 |
| X3 | 1.6 | 1.8 |
| X4 | 1.2 | 1.5 |
| X5 | 1.5 | 1.0 |
|    |     |     |

Consider the data as two-dimensional data points. Given a new data point,  $x = (1.4, 1.6)$  as a query, rank the database points based on similarity with the query using (1) Euclidean distance and (2) cosine similarity.

5. Explain the integrated OLAM and OLAP architecture for banking and financial institutions

-----Best Wishes-----