Birla Institute of Technology & Science-Pilani, Hyderabad Campus
1st Semester 2010-2011
Data Mining (CS C415)
Test I (Regular)

Type: Closed Book **Time: 50 mins** **Wt.age: 20%** **Date: 09/9/2010**

Answer all the following questions.

a. In a star schema, usually [5]
(i) the fact table is normalized
(ii) the number of rows in dimension tables is usually much less compared to that in fact table(s)
(iii) the fact table is de-normalized
(iv) the dimension tables contain less number of columns compared to fact tables(s)

b. In OLAP, dimension reduction can occur due to
(i) Roll-up
(ii) Drill-down
(iii) Slicing
(iv) Dicing

c. It is beneficial and practical to materialize all the views in a data cube when
(i) The number of levels in dimensional hierarchies are very large and there are too many dimensions
(ii) The speed of retrieval is the primary objective
(iii) The cardinality of the dimension is high
(iv) We can implement a greedy algorithm for selecting the views to be materialized

d. Sorting a student database based on student identification numbers.
Is this a function of data mining?
A) Yes
B) No
C) Cannot decide
D) None

e. In multi-dimensional analysis, roll-up can be achieved by:
(i) Moving up a dimension hierarchy like city -> state, etc.
(ii) Moving down a dimension hierarchy like state -> city, etc.
(iii) Adding a new dimension
(iv) Removing one or more dimensions

2. Classify the following attributes as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation, so briefly indicate your reasoning if you think there may be some ambiguity. [5]

Example: Age in years. Answer: Discrete, quantitative, ratio

(a) Time in terms of AM or PM.
(b) Brightness as measured by a light meter.
(c) Brightness as measured by people's judgments.
(d) Angles as measured in degrees between 0°and 360°.
(e) Bronze, Silver, and Gold medals as awarded at the Olympics.
(f) Height above sea level.
(g) Number of patients in a hospital.
(h) ISBN numbers for books.
(j) Military rank.
(k) Distance from the center of campus.

3. Distinguish between noise and outliers. Be sure to consider the following questions and answer whether YES/NO. [5]

(a) Is noise ever interesting or desirable? Outliers?
(b) Can noise objects be outliers?
(c) Are noise objects always outliers?
(d) Are outliers always noise objects?
(e) Can noise make a typical value into an unusual one, or vice versa?

4. Below are the exam scores of a subject make a frequency histogram in R for the exam scores using bins of width 10 beginning at 120 and ending at 200. [5]

192	**160**	**183**	**136**	**162**	**165**	**181**	**188**	**150**	**163**	**192**	**164**
184	**189**	**183**	**181**	**188**	**191**	**190**	**184**	**171**	**177**	**125**	**192**
149	**188**	**154**	**151**	**159**	**141**	**171**	**153**	**169**	**168**	**168**	**157**
160	**190**	**166**	**150**								

5. Discuss why a document-term matrix is an example of a data set that has asymmetric discrete or asymmetric continuous features. [7]

6. Consider the relation given below containing the manufacturer's sales data: Arrange the data so that the automobile manufacturer can increase sale volumes by examining sales data collected throughout the organization. The evaluation would require viewing historical sales volume figures from multiple dimensions such as [8]

(Hint: generate 2D, 3D and 4D data)

- Sales volume by model
- Sales volume by color
- Sales volume by dealership
- Sales volume over time

SALES VOLUMES FOR ALL DEALERSHIPS

MODEL	COLOR	DEALERSHIP	VOLUME
MINI VAN	BLUE	CLYDE	6
MINI VAN	BLUE	GLEASON	6
MINI VAN	BLUE	CARR	2
MINI VAN	RED	CLYDE	3
MINI VAN	RED	GLEASON	5
MINI VAN	RED	CARR	5
MINI VAN	WHITE	CLYDE	2
MINI VAN	WHITE	GLEASON	4
MINI VAN	WHITE	CARR	3
SPORTS COUPE	BLUE	CLYDE	2
SPORTS COUPE	BLUE	GLEASON	3
SPORTS COUPE	BLUE	CARR	2
SPORTS COUPE	RED	CLYDE	7
SPORTS COUPE	RED	GLEASON	5
SPORTS COUPE	RED	CARR	2
SPORTS COUPE	WHITE	CLYDE	4
SPORTS COUPE	WHITE	GLEASON	5
SPORTS COUPE	WHITE	CARR	1
SEDAN	BLUE	CLYDE	6
SEDAN	BLUE	GLEASON	4
SEDAN	BLUE	CARR	2
SEDAN	RED	CLYDE	1
SEDAN	RED	GLEASON	3
SEDAN	RED	CARR	4
SEDAN	WHITE	CLYDE	2
SEDAN	WHITE	GLEASON	2
SEDAN	WHITE	CARR	3