Dr. Sakshi Babbar

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT
MID SEMESTER EXAMINSATION-2014
B.Tech VI Semester

Course Name: Data Mining **Max. Marks : 30**
Course Code: 10B22C1622 **Time: 2hrs**

Section A **(6 marks)**
All questions are compulsory.

1. In real world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem.
2. How many association rules are possible for 4 items?
3. What is the role of Eigen values and Eigenvectors in PCA?
4. How will you compute support count and support of following rule?

   $X \rightarrow Y$
5. What do you understand by supervised and unsupervised learning in data mining?
6. What is the value ranges of the following normalization methods:
   1. min-max
   2. z-score

Section B **(9 marks)**
All questions are compulsory.

1. Summarize proximity measures for binary, nominal, ordinal, and numerical data types. Explain each with the help of an example.
2. Given one-dimensional data set

   X =(-5.0, 23.0, 17.6, 7.23, 1.11)

   Normalize the data set using: (1) Min-max in interval [-1,1] and (2) Z-score
3. Using data in Table 1:
   1. Draw box plot for dimension A2. Show mild and extreme outliers.
   2. Interpret spread in dimension A2.
   3. Draw scatter plot of 2D data set. Do you see any correlation?

Section C **(15 marks)**

All questions are compulsory.

1. Compute Euclidian and correlation matrix for following 2D data set. Analyse correlation matrix.

	A1	A2
X1	4.7	1.7
X2	1.9	1.9
X3	1.8	1.8
X4	3.5	1.5
X5	2.0	9.0

Table 1

2. Consider the database shown in the following table over the set of items

Items : { beer, chips, pizza, wine}

TID	Set of items
1	Beer, chips, wine
2	Beer, chips
3	Pizza, wine
4	Chips, wine

Compute frequent itemsets with support count=2.

3. Consider the 2D database given in Table 2

	A1	A2
X1	2	6
X2	3	9
X3	4	12
X4	5	15

Apply PCA to convert 2D data in 1D data set. Show step by step approach of PCA on this data set.