JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

MAKEUP EXAMINATION- 2016

B.Tech (BI) VI Semester

COURSE CODE: 16B11BI612   MAX. MARKS: 25

COURSE NAME: Datawarehousing and Mining for Bioinformatics

COURSE CREDITS: 04   MAX. TIME: 1Hr 30 Min

Note: *All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means.*

1. Distinguish between (12)
   (a) Hypothesis-Based vs. Exploratory-Based induction
   (b) Pixel based visualization and Geometric projection visualization
   (c) Scatter plot and q-q plot for a dataset
   (d) Binning and regression

2. Suppose we have the following 2-D 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

(a) Consider the data as 2-D 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 Euclidean distance, Manhattan distance, supremum distance and cosine similarity. (8)

(b) Normalization is very commonly used before modelling any dataset. Why is it important? Normalize the data set to make the norm of each data point equal to 1. Use Euclidean distance on the transformed data to rank the data points. (1+4)