

COURSE CODE: 16B11BI612

MAX. MARKS: 25

COURSE NAME: Datawarehousing and Mining for Bioinformatics

COURSE CREDITS: 04

MAX. TIME: 1.5 HRS

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

1. How is data warehouse different from a database? How are they similar? (2)
2. Suppose that a hospital tested the age and body fat data for 18 randomly selected adults with the following results: (5)

age	23	23	27	27	39	41	47	49	50
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2

age	52	54	54	56	57	58	58	60	61
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

- (a) Draw the boxplots for age and %fat
 - (b) Draw a scatter plot and a q-q plot based on these two variables.
3. Briefly outline how to compute the dissimilarity between objects described by the following: (10)
 - (a) Nominal attributes
 - (b) Asymmetric binary attributes
 - (c) Numeric attributes
 - (d) term frequency vectors
 4. Given two objects represented by the tuples (22,1,42,10) and (20, 0, 36,8). Compute the Manhattan and supremum distances between the two objects, also calculate the Minkowski distance using $q=3$. (5)
 5. Data quality can be assessed in terms of several issues, including accuracy, completeness, and consistency. For each of the above three issues, discuss how data quality assessment can depend on the intended use of the data, giving examples. Propose two other dimensions of data quality. (3)