

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST-3 EXAMINATION- JUNE -2018

B.Tech (IT), 6th Semester

COURSE CODE: 10B22CI622

MAX. MARKS: 35

COURSE NAME: Data Mining

COURSE CREDITS: 4

MAX. TIME: 2 HRS

Note: All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means. Assumption may taken if necessarily required.

Section A: Each part contains Two Marks [2 X 4= 8]

Q.1 (a) What is boosting? State why it may improve the accuracy of decision tree induction.

Q.1 (b) Given the following measurements for the variable age: 28; 32; 35; 52; 38; 53; 43; 45; 66; 38; standardize the variable by the following:

(i) Compute the mean absolute deviation of age. (ii) Compute the z-score for the first four measurements.

Q.1 (c) Given two objects represented by the tuples (32, 11, 52, 14) and (20, 7, 36, 8).

(i) Compute the Euclidean distance between the two objects (ii) Compute the Manhattan distance between the two objects (iii) Compute the Minkowski distance between the two objects, using $p = 3$.

Q.1 (d) Outline an efficient method that may find strong correlation rules in a large, multi-relational database.

Section B: Each part contains Four Marks [4 X 3= 12]

Q.2 Why is outlier mining important? Briefly describe the different approaches behind statistical-based outlier detection, distanced-based outlier detection, density-based local outlier detection, and deviation-based outlier detection. [4]

Q.3 Briefly outline the major steps of decision tree classification. Given a decision tree, you have the option of (a) converting the decision tree to rules and then pruning the resulting rules, or (b) pruning the decision tree and then converting the pruned tree to rules. What advantage does (a) have over (b)? [4]

Q.4 Why is the establishment of theoretical foundations important for data mining? Name and describe the main theoretical foundations that have been proposed for data mining. Comment on how they each satisfy (or fail to satisfy) the requirements of an ideal theoretical framework for data mining. [4]

Section C: [15 Marks]

Q.5. Apply the suitable algorithm to calculate the following.

[6]

Midterm Exam	Final Exam
72	84
50	63
81	77
74	78
94	90
86	75
59	49
83	79
65	77
33	52
88	74
81	90

- Plot the data. Do x and y seem to have a linear relationship?
- Use the method of least squares to find an equation for the prediction of a student's final exam grade based on the student's midterm grade in the course.
- Predict the final exam grade of a student who received an 86 on the midterm exam.

Q.6 Consider the class labeled tuples from the AllElectronics customer database as follows;

RID	Age	Income	Student	Credit-Rating	Class: buys computer
1	Youth	High	No	fair	no
2	Youth	High	No	excellent	No
3	Middle Aged	High	No	fair	Yes
4	Senior	Medium	No	fair	Yes
5	Senior	Low	Yes	fair	Yes
6	Senior	Low	Yes	excellent	No
7	Middle Aged	Low	Yes	excellent	Yes
8	Youth	Medium	No	fair	No
9	Youth	Low	Yes	fair	Yes
10	Senior	Medium	Yes	fair	Yes
11	Youth	Medium	Yes	excellent	Yes
12	Middle Aged	Medium	No	excellent	Yes
13	Middle Aged	High	Yes	fair	Yes
14	Senior	Medium	No	excellent	No
15	Middle Aged	Medium	Yes	fair	No
16	Middle Aged	Medium	Yes	excellent	Yes

Consider the tuple $X=(age=youth, income=medium, student=yes, credit\ rating=fair)$

Predict a class level using Naïve Bayesian Classification. [9 Marks]

-----END-----