JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATION- 2023

B.Tech-VI Semester (BI)

COURSE CODE(CREDITS): 18B11BI611 (3) MAX. MARKS: 35

COURSE NAME: Machine Learning for Bioinformatics MAX. TIME: 2 Hours

COURSE INSTRUCTORS: Vipul Sharma

Note: All questions are compulsory. Marks are indicated against each question in square brackets.

Q1 CO-3

a. For the given dataset and a new instance = (Green, MUV, Imported), find out whether it will be stolen or not using Naïve Bayes classifier. [5,2]

Instance	Colour	Type	Origin	Stolen
1	Green	SUV	Imported	Yes
2	Green	SUV	Imported	Yes
3	Green	SUV	Imported	No
4	White	SUV	Imported	No
5	White	SUV	Domestic	Yes
6	White	MUV	Domestic	Yes
7	White	MUV	Domestic	No
8	White	MUV	Imported	No
9	Green	MUV	Domestic	Yes
10	Green	SUV	Domestic	Yes

b. What is manhattan distance and how is it different from cosine distance?

Q2 CO-4

a. Give mathematical background of gradient descent algorithm? What will happen if learning rate is kept too high? [5,2]

b. What is oscillation problem? How will you solve it?

Q3 CO-2

a. Consider the following 8 data points with (x, y) representing locations. Use k-means clustering algorithm to group these into three clusters. A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9) Note: Consider the initial cluster centers as A1, A4 and A7. The distance function between two data points a = (x1, y1) and b = (x2, y2) is defined as: $P(a, b) = |x2 - x1| + |y2 - y1|$. [5,2]

[P.T.O]

b. What is a dendrogram? How can we measure the goodness of clusters in hierarchical clustering algorithm?

Q4 CO-5 a. Differentiate between hard and soft SVM using geometric interpretation. Derive the objective function for soft SVM and also explain how to optimize it. [5,2]

b. Differentiate between linear regression and logistic regression?

Q5 CO-1 a. How will you represent a perceptron as a logistic regression model? Write down the optimization problem to represent it and mention the steps needed to solve it. [5,2]

b. Given the following training examples from the questionnaires survey (to ask people opinion) with two attributes (acid durability and strength) to classify whether a special paper tissue is good or not. Find the class of the test sample using k-NN algorithm. Take k =3. Use L2 Norm for distance computations.

X1 = Acid Durability (seconds)	X2 = Strength (kg/squaremeter)	Y= Classification
7	7	Bad
7	4	Bad
3	4	Good
1	4	Good

Test Sample

X1 = Acid Durability (seconds)	X2 = Strength (kg/squaremeter)	Y= Classification
3	7	?