JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT
TEST -2 EXAMINATION- APRIL-2023

COURSE CODE (CREDITS): 18B11BI611 (3)

MAX. MARKS: 25

COURSE NAME: Machine Learning for Bioinformatics

COURSE INSTRUCTORS: Vipul Sharma

MAX. TIME: 1 Hour 30 Minutes

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

Q1. CO-3   a. Is the gini impurity of a node lower or greater than that of its parent? How does a decision tree handle continuous (numerical) features?   [2,3]

b. Use CART algorithm to draw the decision tree for the given dataset.

Instance	A1	A2	Class Label
1	T	T	+
2	T	T	+
3	T	F	-
4	F	F	+
5	F	T	-
6	F	T	-

Q2. CO-3   a. Why is logistic regression termed as regression and not classification?   [2,2,1]

b. Can we solve the multiclass classification problem using logistic regression? If yes then how?

c. What is the difference between hard SVM and soft SVM?

Q3. CO-4   a. What do you mean by overfitting and underfiitng of a model?   [2,3]

b. What is the difference between correlation and regression? Is it possible to apply linear regression for time series analysis?

Q4. CO-4   a. What is convex hull? How will you use it to find the hyperplane for SVM?   [1,3,1]

b. Derive the objective function for Hard SVM?

c. What do you mean polynomial kernel? Give its significance.

[P.T.O]

Q5. CO-5 a. What is K in case of KNN? How will you compute the best value of K? [2,3]

b. Given a training dataset as shown below. Find out what kind of sport Ramesh will play if k factor for KNN is 4 and his age is 5. (Use Manhattan Distance Metric)

Name	Age	Gender	Class of Sports
Ajay	32	0	Football
Mark	40	0	Hockey
Sara	16	1	Cricket
Zaira	34	1	Cricket
Sachin	55	0	Hockey
Rahul	40	0	Cricket
Pooja	20	1	Hockey
Smith	15	0	Cricket
Laxmi	55	1	Football
Michael	15	0	Football

****