JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATION- 2024

B.Tech-VII Semester (CSE/IT/ECE)

COURSE CODE (CREDITS):19B1WCI731(2) MAX. MARKS: 15

COURSE NAME: Computational Data Analysis

COURSE INSTRUCTORS: VKS MAX. TIME: 1 Hour

Note: ***(a)*** *All questions are compulsory.*

(b) *Marks are indicated against each question in square brackets.*

(c) *The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems*

Q1: You are given the following data points for two variables, X and Y: X=(1,2,3,4,5) & Y=(2,4,6,8,10). Perform linear regression to find out the equation of the line that best fit these data points. CO1 [3]

Q2: Why sigmoid function is used in logistic regression? Find out the derivative of sigmoid. CO1 [3]

Q3: How does the choice of "k" in the k-NN algorithm affect the classification performance of the model? What are the trade-offs associated with using a small "k" versus a large "k"? CO1 [3]

Q4: Suppose we have a small dataset with two features and a binary outcome. We want to fit a logistic regression model to this data using gradient descent. CO2 [3]

X1	X2	Y
1	2	0
2	3	0
3	4	1
4	5	1

Model is of the form: $h\theta(x) = \frac{1}{1+e^{-(\theta 0+\theta 1 x1+\theta 2 x2)}}$

Loss function is:

$$J(\theta 0, \theta 1, \theta 2) = -1/m \sum_{i=1}^{m} [y^i \log(h\theta(x^i)) + (1 - y^i) \log(1 - h\theta(x^i))$$

Find out the optimal values of $\theta 0, \theta 1$ & $\theta 2$ after the first iteration if intial values of the parameters are (0,1 & 1) and learning rate is 0.1.

Q5: Derive an expression for objective function of soft SVM using a suitable example. What is the significance of hyperparameter "c" we use in this objective function? CO2 [3]