JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -2 EXAMINATION- 2023

B.Tech-VII Semester (CSE/IT)

COURSE CODE (CREDITS): 19B1WCI731   MAX. MARKS: 25
COURSE NAME: Computational Data Analysis

COURSE INSTRUCTORS: Dr. Nishant Sharma   MAX. TIME: 1 Hour 30 Minutes

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 working on a text classification problem using Naive Bayes, and you have a dataset with the following training data for two classes, "Spam" and "Not Spam." You want to classify a new document based on the word "offer."

Class	Number of Documents Containing "offer" (x)	Total Documents in Class (N)
Spam	120	400
Not Spam	50	600

Using Laplacian Smoothing with a smoothing factor (alpha) of 1, calculate the Naive Bayes probability of the new document belonging to the "Spam" class and the "Not Spam" class based on the presence of the word "offer." **[5 marks] [CO-2]**

Q2. What is significance of geometric margins in SVM classification? Suppose you are working with a binary classification problem using a linear SVM. Decision boundary for SVM is represented by the equation 3x−4y−7=0. The equation of the hyperplane is represented as ax+by+c=0. Based on above information, calculate the geometric margin for the hyperplane for a data point (2, 1). **[4 marks] [CO-3]**

Q3. Elaborate on the concept of Cross-Validation and its significance in model selection. How can it help avoid overfitting? **[4 marks] [CO-4]**

Q4. Describe Bayesian Regression and Bayesian Logistic Regression. How are these techniques useful in modeling and prediction? **[3 marks] [CO-4]**

Q5. What is feature selection, and why is it important in machine learning? List and briefly explain various feature selection techniques with suitable illustrations. **[4 marks] [CO-4]**

Q6. Discuss the differences between DBSCAN, k-means, and Hierarchical clustering. Discuss different scenarios where one maybe preferred over another? **[5 marks] [CO-3]**