JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST-2 EXAMINATION - 2022

B.Tech. VII Semester (CSE, IT)

COURSE CODE (CREDITS): 19B1WCI731 (2) MAX. MARKS: 25

COURSE NAME: Computational Data Analysis

COURSE INSTRUCTORS: Dr. Ekta Gandotra MAX. TIME: 1 Hour and 30 Minutes

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

Q1. a. Using the following dataset, predict the class for the DocId=6 using Naïve Bayes with Laplace Smoothing algorithm. Take smoothing parameter as 1. [3] CO5

	DocId	Words in the document	Belongs to Chine?
Training data	1	Chinese Beijing Chinese	Yes
	2	Chinese Chinese Shanghai	Yes
	3	Chinese Macao Taiwan	Yes
	4	Tokyo Japan Chinese	No
	5	Sapporo Osaka Taiwan	No
Testing data	6	Chinese Chinese Tokyo Japan	?

b. How categorical variables are handled in Logistic Regression? [2] CO2

c. Give two scenarios where gradient descent may fail to converge. [2] CO5

Q2. a. Give the mathematical interpretation of Principal Component Analysis. [3] CO4

b. Compute the principal components for the following 2-dimensional dataset. [3] CO4
$X=(x_1, x_2)=\{(1,2), (3,3), (3,5),(5,4),(5,6), (6,5),(8,7),(9,8)\}$

Q3. a. Consider the following 10 data points (with (x, y) representing locations). Use k-Means algorithm to find the three cluster centers after the second iteration. Take the initial cluster centers as C1(1,5), C2(4,1), C3(8,4) and use L2 norm for distance computation. [3] CO3
A1(2, 4), A2(2, 6), A3(5, 6), A4(4, 7), A5(8, 3), A6(6, 6), A7(5, 2), A8(5, 7), A9(6,3), A10(4,4).

b. What is Silhouette Coefficient? How its value is interpreted to evaluate the quality of a clustering model? [3] CO3

P. T. O

Q4. a. Differentiate between wrapper and filter method of feature selection on the basis of their (i) working process (ii) computational cost (iii) examples. [3] CO4

b. A random sample of 395 people was surveyed and each person was asked to report the highest education level they obtained. Compute the value of $\chi 2$ for the following data resulted from the survey. Also discuss the relationship between the gender of individual and the level of education that they have obtained. (Note: The critical value of $\chi 2$ with 3 degrees of freedom is 7.815 at 5% level of significance). [3] CO4

	High School	**Bachelors**	**Masters**	**Ph.D.**
Female	60	54	46	41
Male	40	44	53	57