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

COURSE NAME: Computational Data Analysis

COURSE INSTRUCTORS: Ekta Gandotra   MAX. TIME: 1 Hour 30 Min

Note: *(a) All questions are compulsory.*
   (b) The candidate is allowed to make Suitable numeric assumptions wherever required
      for solving problems
   (c) Calculator is allowed.

Q. No.	Question	CO	Marks
Q1.	Show that applying Maximum Likelihood Estimation, under the assumption that the error terms in linear regression model follow a normal distribution, leads to the Ordinary Least Squares objective.	1	5
Q2.	Using the following dataset of movie reviews, predict the class for the Review ID=6 using Naïve Bayes with Laplace Smoothing algorithm. Take smoothing parameter as 1.	2	5
Q3.	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 χ^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 χ^2 with 3 degrees of freedom is 7.815 at 5% level of significance).	4	5

Q2 table:

	Review ID	Words in the review	Class
Training data	1	I loved the movie	+
	2	I hated the movie	-
	3	A great movie. Good movie	+
	4	Poor acting	-
	5	Great acting. A good movie	+
Testing data	6	I hated the poor acting	?

Q3 table:

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

Q4.	Compute the first principal component for the following 2-dimensional dataset. Also, find the proportion of total variance explained by the first and second principal components separately. $X = (x_1, x_2) = (1,2), (3,3), (3,5), (5,4)$	4	5

Q5.	a. Discuss the steps (including mathematics) of Gradient Descent algorithm. Can Gradient Descent be applied to Non-Convex Functions? Explain.
	b. Differentiate between the following with the help of an example:
	   i. Parameters and hyper-parameters
	   ii. Ordinal and nominal variables
		1	3
			
			2