Question Paper

MANIPAL ACADEMY OF HIGHER EDUCATION

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

VII SEMESTER B.TECH. COMPUTER & COMMUNICATION ENGINEERING

END SEMESTER EXAMINATIONS, NOVEMBER 2022

MACHINE LEARNING FOR DATA ANALYTICS [ICT 4056]

Marks: 50   **Duration: 180 mins.**

Descriptive Questions

Answer all the questions.   Section Duration: 180 mins

Missing data, if any, may be suitably assumed.

1) A) Consider a univariate model build with logistic regression, given the values β_0= -2.9 and β_1=0.7. (5)
The covariance matrix is shown below.

	x	Constant
x	0.000283	
Constant	-0.0377	3.1

   a. Compute the end points of $100(1-\alpha)$ % walt based confidence interval for the logit function for x =40. [Assume SE=0.2549 and Z value = 1.82].

   b. Calculate the estimated variance for x = 60.

   c. Find the estimated logistic probability value for x = 60.

   d. Given n_1 = 38, n_0 = 63, evaluate the value for G. [Log likelihood = -45.677]

   B) Give an appropriate example to illustrate the steps involved in the machine learning process. (3)

   C) Write the steps to eliminate redundant variables from our dataset using correlation analysis? (2)

2) A) Illustrate the term hypothesis function with a suitable example. Derive normal equation for parameter as per the LMS algorithm. (5)

   B) Consider a data set given in Table 1, where x being independent variable and y being dependent variable. Assume that the univariate model is build using logistic regression. Answer the following questions: (3)

   a. Compute odds ratio for the given data and also verify that with the following fitted value (Table 2).

   b. Estimate the end-points for the dichotomized variable PriorFrac [z = 1.96] using logistic regression model.

Table 1			
Fracture during follows-up (y)	**Prior Fracture (x)**		**Total**
	Yes	No	
Present	73	63	136
Absent	74	201	275
Total	147	264	411

Table 2					
Variable	Coeff.	Std.Err	z	p	95%CI
PriorFrac	1.82	0.2231	4.77	<0.001	0.627, 1.501
Constant	-1.417	0.1305	-10.86	<0.001	-1.672, -1.161
Log-Likelihood = -270.033					

C)   Describe the process that can minimize the empirical risk of the machine learning model   (2)

3)   Outline the Steps 1, 3, and 4 of *purposeful variable selection of covariates* in machine learning model.   (5)

A)

B)   A machine learning model to predict Cancer is trained with 2386 patients was able to do right prediction for 900 patients with Cancer and 400 patients without Cancer. If the actual number of patients with Cancer is 1238 and without Cancer is 762. Answer the following questions below: (3)

    a. What will be the model's ability to correctly identify the patients with disease and without disease?

    b. Draw the matrix which clearly represents the above data.

C)   Differentiate between Gradient descent and Normal equation (minimum 4 features)   (2)

4)   Consider the two dimensional patterns (2, 1), (3, 5), (4, 3), (5, 6), (6, 7), (7, 8). Compute the principal component using PCA Algorithm.   (5)

A)

B)   Illustrate the working principle of bagging algorithm in machine learning.   (3)

C)   Write any four properties of multivariate normal distribution.   (2)

5)   Elaborate the computation of correlation matrix, and factor extraction process in factor analysis.   (5)

A)

B)   Identify the issues that occur from the various combination of *Bias - Variance*.   (3)

C)   Suppose we have a model to diagnose heart disease with 75% accuracy. 30% of the people are having positive test for heart disease. What is the probability of a patient being tested positive is actually having that disease?   (2)

-----End-----