Question Paper

Exam Date & Time: 07-Jan-2023 (02:30 PM - 05:30 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

VII SEMESTER B.TECH. (IT/CCE)
MAKE-UP EXAMINATIONS, JANUARY 2023

MACHINE LEARNING FOR DATA ANALYTICS [ICT 4056]

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

A

Answer all the questions.

Instructions to Candidates: Answer ALL questions Missing data may be suitably assumed

1) A) Consider the dataset given in table to build the univariate model using logistic regression, given the values β_0= -3.3 and β_1=0.1. (5)

a. Plot the s curve for the given data set using the grouping method of 10 (age 1-10,11-20.) in stepwise.

b. Compute the log likelihood using last four columns of Table 1.1.

c. Calculate the end points of 100(1-α)% Walt based confidence interval for the logit function for Age =37, Z value = 1.96 and Log-Likelihood = -52.89.

Table 1.1

Age	5	10	25	42	61	65	69	85	75	28	89	72
CHD	No	No	No	No	No	Yes	No	Yes	Yes	No	Yes	No

Table: 1.2

	Age	**Height**	**Constant**
Height	0.85		
Age	0.7	0.6	
Constant	0.5	-0.7	0.555

B) How do you interpret the estimated coefficients in logistic regression model? Explain the fitted logistic regression model with a suitable example. (3)

C) Derive an equation to find the variance of a random variable X and also state its properties. (2)

2) A) A machine learning model to predict Pneumonia is trained with 1000 patients was able to do right prediction for 480 patients with Pneumonia and 180 patients without Pneumonia. If the actual number of patients with Pneumonia is 600 and without Pneumonia is 400. Answer the following questions (5)

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

b. Calculate the F1-score of the model.

c. Compute the receiver operating characteristic curve, and area under curve

B) Describe different machine learning techniques using appropriate examples. (3)

C) Differentiate linear discriminative learning algorithm and generative learning algorithm with suitable examples. (2)

3) Consider the dataset given in below table and Calculate the ideal value for Θ_1 using linear regression method? (Assume $\Theta_0 = 0.5$, $\Theta_2 = 0$, $\Theta_3 = 0.1$, and use gradient descent) (5)

A)

Size(feet)	Number of rooms	Age of home	Price
2104	5	45	460
1416	3	40	232
1534	3	30	315
852	2	36	178

B) How do you select the important features in your data? (3)

C) Write the two advantages and disadvantages of AdaBoost. (2)

4) Suppose $X \sim Binomial(500, 0.2)$ and let $\mu = E[x] = 100$. What is the best bound we can get on $P(X \geq 150)$? Explain the moment generating function of binomial distribution. (5)

A)

B) Consider that X is a discrete random variable with PMF pX(x) and g: R -> R is an arbitrary function and let g(x) be the random variable. Define the expected value for g(x) along with its properties. (3)

C) Show that PCA is variance maximizing problem. (2)

5) Derive the E and M steps of the standard *Expectation Maximization* (EM) algorithm for Gaussian Mixture Model (GMM). (5)

A)

B) Consider a data set with 52 subjects with values (x=1, y=1), 73 with (x=0, y=1), 74 with (x=1, y=0), and 301 with (x=0, y=0), where x being independent variable and y being dependent variable. Compute odds ratio for the given data and also verify this with the fitted value in Table Q5B. (3)

Table: Q5B

Variable	Coeff.
X	1.064
Constant	−1.417

C) "Locally weighted linear - regression is non- parametric algorithm". Justify the statement. (2)

-----End-----