B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

September / October 2023 Semester End Main Examinations

Programme: B.E.   Semester: IV
Branch: Machine Learning   Duration: 3 hrs.
Course Code: 22AM4PCPSM   Max Marks: 100
Course: Probability and Statistics for Machine Learning

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
                   2. Missing data, if any, may be suitably assumed.

				CO	PO	Marks
			UNIT - I	*CO*	*PO*	**Marks**
1	a)		Define sample space and event. Prove that a sample space of N possible outcomes yields 2^N possible events.	*CO1*	*PO2*	**05**
	b)		In a group of 100 persons,72 people can speak English and 23 can speak French. Find i) How many can speak English only? ii) How many can speak French only? iii) How many can speak both English and French?	*CO1*	*PO2*	**08**
	c)		Construct equally likely outcomes for the probability. Consider a card is drawn from a scaffold 52 card deck indiscriminately. Using the relation derived identify the likelihood that the chosen card is a spade.	*CO1*	*PO2*	**07**
			UNIT - II			
2	a)		A program consists of two modules. The number of errors, X_1, in the 1st module and the number of errors, X_2, in the 2nd module have the joint distributions P(0,0)=P(0,1)=P(1,0)=0.2, P(1,1)=P(1,2)=P(1,3)=0.1,P(0,2)=P(0,3)=0.05. The maximum number errors in X_1 is 1 and X_2 is 3 Find: i) Marginal Distributions of $Y = X_1+X_2$ ii) The Probability of number of errors in the first module. iii) The distribution of the total number of errors in the program.	*CO2*	*PO4*	**06**
	b)		For the Bernoulli distribution derive mean, variance & Establish relationship among them	*CO1*	*PO2*	**07**
	c)		The 8% of population in a small-town work as a teacher. i) Relate the suitable discrete distribution. ii) What is the probability that the 10th person in the town is a teacher? iii) Calculate Mean, Variance and Standard Deviation.	*CO2*	*PO4*	**07**

			UNIT - III			
3	a)		Outline the computational procedure of a population median with respect to continuous and discrete distributions along with necessary plots.	CO2	PO4	**06**
	b)		Given data for the continuous series	CO2	PO4	**07**

Class Interval	F(frequency)
0-5	5
5-10	8
10-15	12
15-20	16
20-25	20
25-30	10
30-35	4
35-40	3

Calculate
   i)   quartile deviation
   ii)   Coefficient of quartile deviation

	c)		Using simple descriptive statistics, find Mean, Median, Mode, the quartiles Q1, Q2, Q3 and the interquartile range of the following data 20,30,25,23,22,32,36,18. Plot the graph to represent the quartiles identified.	CO2	PO4	**07**
			UNIT - IV			
4	a)		If X is a continuous random variable with the following pdf $f(x)= \alpha(2x-x^2)$ where $0<x<2$; 0 Otherwise. Find i) α ii) $P(x>1)$	CO2	PO4	**06**
	b)		Derive the maximum likelihood estimator of an geometric distribution function $f(x,p)=p(1-p)^{x-1}$.	CO3	PO3	**08**
	c)		Define the terminologies i). Population Moment ii). Population Central Moment iii). Method of Moments Estimator	CO2	PO4	**06**
			OR			
5	a)		Differentiate between z-Test and t-Test.	CO2	PO4	**06**
	b)		The number of concurrent users for some internet service provider has always averaged 5000 with a standard deviation of 800.After an equipment upgrade, the average number of users at 100 randomly selected, moment of time is 5200.Does it indicate, at 5% level of significance, that the mean number of concurrent users has increased? Assume that standard deviation of the number of concurrent users has not changed.	CO3	PO3	**07**

		c)	Construct the confidence interval for a general method where the level of significance is 95% interms of quartiles.	CO2	PO4	**07**

<div align="center">

UNIT - V

</div>

6	a)	The values of x and their corresponding y values are	CO2	PO4	**07**

x	0	1	2	3	4
y	2	3	5	3	4

   i) Derive the linear regression model
   ii) Estimate the value of y when x = 10

	b)	In the study of performance of 3 detergents and 3 different water temperature, the following readings were obtained with specially designed equipment.	CO2	PO4	**08**

Water temperature	A	B	C
Cold Water	47	45	50
Warm Water	39	42	52
Hot Water	44	36	48

using two-way ANNOVA.
     i) Find grand total correction factor
     ii) Calculate the ratio of F
     iii) Write the conclusion

	c)	For a univariate linear regression, show that $SS_{TOT} = SS_{REG} + SS_{ERR}$	CO3	PO3	**05**

<div align="center">

OR

</div>

7	a)	Differentiate between ANOVA and R-Square tests with an example.	CO2	PO4	**06**
	b)	Apply matrix method for the given database and determine the Least Squares Estimation.	CO2	PO4	**07**

Data Size(gigabytes), x1	6	7	7	8	10	10	15
Number of tables, x2	4	20	20	10	10	2	1
Processed requests, y	40	55	50	41	17	26	16

	c)	Given the following data	CO3	PO3	**07**

     i) With the help of PCA, reduce its dimensionality.
     ii) Populate the Covariance Matrix
     iii) Find Eigen Value

Feature	Example-1	Example-2	Example-3	Example-4
x	4	8	13	7
y	11	4	5	14

<div align="center">

******

</div>