U.S.N.										

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

Autonomous Institute Affiliated to VTU

December 2023 Supplementary 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.

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

		UNIT - I	*CO*	*PO*	Marks
1	a)	Describe a) sample space. b) probability of union for mutually exclusive events with the help of Venn diagram	*CO1*	*PO2*	**04**
	b)	Suppose that after 10 years of service, 40% of computers have problem with motherboards (MB), 30% have problem with hard drives (HD), and 15% have problem with both MB and HD. What is the probability that a 10-year-old computer still has fully functioning MB and HD?	*CO1*	*PO2*	**08**
	c)	A computer program is tested by 3 independent tests. When there is an error, these tests will discover it with probabilities 0.2, 0.3, and 0.5, respectively. Suppose that the program contains an error. What is the probability that it will be found by at least one test?	*CO1*	*PO2*	**08**
		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)	An exciting computer game is released. Sixty percent of players complete all the levels. Thirty percent of them will then buy an advanced version of the game. i) Among 15 users, what is the expected number of people who will buy the advanced version? ii) Predict the probability that at least two people will buy it?	*CO1*	*PO2*	**07**

		c)	Derive the mean and variance of Bernoulli distribution.	*CO2*	*PO4*	**07**
			UNIT - III			
	3	a)	The CPU time for n = 30 randomly chosen jobs (in seconds) is given below 70,36, 43, 69, 82, 48, 34, 62, 35, 15, 59, 139, 46, 37, 42, 30, 55, 56, 36, 82, 38, 89, 54, 25, 35, 24, 22, 9, 56, 19. Find the i) Mean ii) Median iii)1st and 3rd quartile of the CPU times	*CO2*	*PO4*	**06**
		b)	Infer the computation of a population median with respect to continuous and discrete distributions by plotting the respective graphs.	*CO2*	*PO4*	**07**
		c)	Explain outliers and interquartile range method for detection of outliers with an example.	*CO2*	*PO4*	**07**
			UNIT - IV			
	4	a)	Elaborate the acceptance and rejection regions for z-test	*CO2*	*PO4*	**04**
		b)	Derive the estimator of a geometric distribution using the method of MLE.	*CO3*	*PO3*	**08**
		c)	A candidate prepares for the local elections. During his campaign, 42 out of 70 randomly selected people in town A and 59 out of 100 randomly selected people in town B showed that they would vote for this candidate. Predict the difference in support that this candidate is getting in towns A and B with 95% confidence.	*CO2*	*PO4*	**08**
			OR			
	5	a)	Explain in detail hypothesis testing with an example	*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*	**08**
		c)	Summarize the steps involved in level-α test	*CO2*	*PO4*	**06**
			UNIT - V			
	6	a)	The following statistics were obtained from sample of size n = 75. The predictor variable X has mean 32.2, & variance 6.4. The response variable Y has mean 8.4 & variance 2.8. The sample covariance between X and Y is 3.6. i). Estimate the linear regression equation predicting Y based on X ii). Construct a 99% confidence interval for the regression slope & Verify its significance.	*CO2*	*PO4*	**07**

		b)	At a gas station, 180 drivers were asked to record the mileage of their cars and the number of miles per gallon. The results are summarized in the table. The sample correlation coefficient is r = −0.17. i). Compute the least squares regression line which describes how the number of miles per gallon depends on the mileage. What do the obtained slope and intercept mean in this situation? ii). Use R_2 to evaluate its goodness of fit. Is this a good model?	*CO2*	*PO4*	**06**
		c)	The time it takes to transmit a file always depends on the file size. Suppose you transmitted 30 files, with the average size of 126 Kbytes and the standard deviation of 35 Kbytes. The average transmittance time was 0.04 seconds with the standard deviation of 0.01 seconds. The correlation coefficient between the time and the size was 0.86. Based on this data, fit a linear regression model and predict the time it will take to transmit a 400 Kbyte file.	*CO3*	*PO3*	**07**
			OR			
	7	a)	A computer manager needs to know the efficiency of her new computer program that depends on the size of incoming data. Efficiency will be measured by the number of processed requests per hour. Applying the program to data sets of different sizes, she gets the following results, Predict the regression equation for the above data.	CO2	PO4	**07**
		b)	For a univariate linear regression, show that R-square is the squared sample correlation coefficient, $R_2 = r2$.	CO2	PO4	**07**
		c)	Consider a dataset with two features, X_1 and X_2, and a sample covariance matrix: Sample Covariance Matrix= $\begin{bmatrix} 1.5 & 0.7 \\ 0.7 & 0.9 \end{bmatrix}$ i) Calculate the eigenvalues and eigenvectors ii) Determine the principal components.	CO3	PO3	**06**

Data size (gigabytes), x	6	7	7	8	10	10	15
Processed requests, y	40	55	50	41	17	26	16

Units	Sample Mean	Standard Deviation
Mileage	24,598	14,634
Miles per Gallon	23.8	3.4

****