U.S.N.										

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

Autonomous Institute Affiliated to VTU

September / October 2024 Supplementary Examinations

Programme: B.E. **Semester: IV**
Branch: Artificial Intelligence & 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.

		UNIT - I	*CO*	*PO*	Marks
1	a)	Suppose that after 10 years of service, 40% of computers have problems with motherboards (MB), 30% have problems with hard drives (HD), and 15% have problems with both MB and HD. What is the probability that a 10-year old computer still has fully functioning MB and HD?	*CO1*	*PO2*	**06**
	b)	Three computer viruses arrived as an e-mail attachment. Virus A damages the system with probability 0.4. Independently of it, virus B damages the system with probability 0.5. Independently of A and B, virus C damages the system with probability 0.2. What is the probability that the system gets damaged?	*CO1*	*PO2*	**06**
	c)	A quiz consists of 6 multiple-choice questions. Each question has 4 possible answers. A student is unprepared, and he has no choice but guessing answers completely at random. He passes the quiz if he gets at least 3 questions correctly. What is the probability that he will pass?	*CO1*	*PO2*	**08**
		UNIT - II			
2	a)	Summarize the properties of variances and covariance	*CO1*	*PO2*	**06**
	b)	A service organization in a large town organizes a raffle each month. One thousand raffle tickets are sold for $1 each. Each has an equal chances of winning. First prize is $300, second prize is $200 and third prize is $100. Let X denote the net gain from the purchase of one ticket. i) Construct the probability distribution of X. ii) Find the probability of winning any money in the purchase of one ticket. iii) Find the expected value of X, and interpret its meaning.	*CO1*	*PO2*	**08**
	c)	Derive the mathematical representation of expectation and variance of an exponential distribution.	*CO1*	*PO2*	**06**

			UNIT - III			
	3	a)	What is the limitation of sample mean? With an example and necessary diagram, analyze whether it is reliable or not. What is the alternate measure that can be used for estimation?	*CO2*	*PO4*	**08**
		b)	A computer maker sells extended warranty on the produced computers. It agrees to issue a warranty for x years if it knows that only 10% of computers will fail before the warranty expires. It is known from past experience that lifetimes of these computers have Gamma distribution with $\alpha = 60$ and $\lambda = 5$ years^{-1}. Compute x and advise the company on the important decision under uncertainty about possible warranties.	*CO2*	*PO4*	**06**
		c)	The following data set represents the number of new computer accounts registered during ten consecutive days. The data is 43, 37, 50, 51, 58, 105, 52, 45, 45, 10. Compute the mean, median, quartiles, and standard deviation.	*CO2*	*PO4*	**06**
			UNIT - IV			
	4	a)	A manager evaluates effectiveness of a major hardware upgrade by running a certain process 50 times before the upgrade and 50 times after it. Based on these data, the average running time is 8.5 minutes before the upgrade, 7.2 minutes after it. Historically, the standard deviation has been 1.8 minutes, and presumably it has not changed. Construct a 90% confidence interval showing how much the mean running time is reduced due to the hardware upgrade.	*CO2*	*PO4*	**08**
		b)	Consider the sample 3, 3, 3, 3, 3, 7, 7, 7 drawn from discrete distribution with the probability mass function $\begin{cases} P(3) = \theta \\ P(7) = 1 - \theta \end{cases}$ Compute: i) The method of moment estimator ii) The maximum likelihood estimator iii) estimate standard error of each estimator of θ	*CO2*	*PO4*	**08**
		c)	Differentiate between the confidence interval for the population mean and difference between 2 means.	*CO2*	*PO4*	**04**
			OR			
	5	a)	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 they would vote for this candidate. Estimate the difference in support that this candidate is getting in towns A and B with 95% confidence. Can we state affirmatively that the candidate gets a stronger support in town A?	*CO2*	*PO4*	**10**
		b)	In order to ensure efficient usage of a server, it is necessary to estimate the mean number of concurrent users. According to records, the average number of concurrent users at 100 randomly selected times is 37.7, with a standard deviation $\sigma = 9.2$. (a) Construct a 90% confidence interval for the expectation of the number of concurrent users.	*CO2*	*PO4*	**10**

			(b) At the 1% significance level, do these data provide significant evidence that the mean number of concurrent users is greater than 35?			
			UNIT - V			
	6	a)	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*	**10**
		b)	Summarize the matrix approach of multivariate regression	*CO3*	*PO3*	**10**
			OR			
	7	a)	The following statistics were obtained from a 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; and – The sample covariance between X and Y is 3.6. a) Estimate the linear regression equation predicting Y based on X. b) Complete the ANOVA table. What portion of the total variation of Y is explained by variable X? c) Construct a 99% confidence interval for the regression slope. Is the slope significant?	*CO3*	*PO3*	**10**
		b)	For a univariate linear regression, show that $SS_{TOT} = SS_{REG} + SS_{ERR}$.	*CO3*	*PO3*	**10**