U.S.N.										

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

Autonomous Institute Affiliated to VTU

October 2024 Supplementary Examinations

Programme: B.E. **Semester: III**
Branch: Artificial Intelligence and Machine Learning **Duration: 3 hrs.**
Course Code: 23AM3PCPSM **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)		Can quantification of uncertainty through probabilistic models enhance the decision-making process in Machine Learning applications? Illustrate with an example.	*CO1*	*PO1*	**05**
	b)		The Joint probability distribution of two random variables X and Y are: P (X=0, Y=1) = 1/3, P (X=1, Y=−1)= 1/3 and P (X=1, Y=1) = 1/3. i. Find marginal distribution of X and Y ii. Are X and Y independent?	*CO2*	*PO1*	**08**
	c)		In a multinational corporation, 60% of the employees are college graduates. Among these graduates, 10% work in the sales department. Among the employees who did not graduate from college, 80% are in sales. Calculate the probability that a randomly selected employee is neither in sales nor a college graduate.	*CO2*	*PO2*	**07**
			UNIT - II			
2	a)		Explain how Maximum a Posteriori (MAP) probability can be used to estimate the unknown parameter from data.	*CO1*	*PO1*	**05**
	b)		State Bayes' Theorem and elaborate on its applications.	*CO1*	*PO1*	**05**
	c)		The chances that doctor 'A' will Diagnose disease 'X' correctly is 60%. The chances that a patient will die by his treatment after a correct diagnosis is 40% and the chance of death by wrong diagnosis is 70%. If a patient of doctor 'A', who had disease 'X' dies, compute i. The probability that his disease was diagnosed correctly? ii. The probability that his disease was not diagnosed correctly?	*CO2*	*PO3*	**10**
			OR			

	3	a)	Consider the Customer loan dataset. <table><tr><th>Applicant ID</th><th>Income (>30 K)</th><th>Employment Status</th><th>Loan amount (>5Lakh)</th><th>Was loan given</th></tr><tr><td>1.</td><td>Yes</td><td>Employed</td><td>Yes</td><td>No</td></tr><tr><td>2.</td><td>No</td><td>Unemployed</td><td>No</td><td>Yes</td></tr><tr><td>3.</td><td>Yes</td><td>Employed</td><td>Yes</td><td>No</td></tr><tr><td>4.</td><td>Yes</td><td>Employed</td><td>Yes</td><td>No</td></tr><tr><td>5.</td><td>No</td><td>Employed</td><td>No</td><td>Yes</td></tr><tr><td>6.</td><td>No</td><td>Employed</td><td>Yes</td><td>No</td></tr><tr><td>7.</td><td>No</td><td>Unemployed</td><td>No</td><td>Yes</td></tr><tr><td>8.</td><td>Yes</td><td>Employed</td><td>Yes</td><td>Yes</td></tr></table>If an applicant, whose income is less than 30K and is unemployed, has applied for a loan of 3 Lakhs, will the bank sanction the loan?	*CO2*	*PO3*	**10**
		b)	A diagnostic test is used to identify a rare disease in patients. The test has a sensitivity of 95% and a specificity of 90%. The prevalence of the disease in the population is 0.1%. i. What is the probability that a person who tests positive actually has the disease? ii. If the test result is negative, what is the probability that the person does not have the disease?	*CO2*	*PO2*	**10**
			UNIT - III			
	4	a)	Differentiate between Discrete and Continuous random variables.	*CO1*	*PO1*	**05**
		b)	Using the definition of variance show that $V(X) = E(X^2) - E(X)^2$	*CO1*	*PO1*	**05**
		c)	i. Obtain the expression for the mean and variance of binomial distribution with parameters n and p. ii. Show that the variance derived in (i) is less than mean.	*CO1*	*PO2*	**10**
			OR			
	5	a)	A software company is testing its new antivirus program's effectiveness against different types of malwares. The probability that the antivirus program successfully detects and removes virus of type A and type B which are independent of each other is 0.9 and 0.8 respectively. i. Find the probability mass function (PMF) of the virus that are successfully detected and removed. ii. Draw a graph of its cumulative distribution function.	*CO2*	*PO2*	**10**
		b)	In a video game, a player attempts to complete a challenging level with a 10% success rate. The player will keep trying until the level is completed successfully for the first time. i. Calculate the probability that the player clears the level in his first attempt. ii. Determine the probability that the player needs exactly 3 attempts to clear the level. iii. If the player has not cleared the level after 5 attempts, what is the probability that he will clear it on the 6^{th} attempt?	*CO2*	*PO2*	**06**
		c)	Find λ for a Poisson random variable x if 9P (X = 4) = P (X = 2).	*CO2*	*PO2*	**04**

			UNIT - IV			
	6	a)	An internet service provider has two connection lines for its customers. Eighty percent of customers are connected through Line 1, and twenty percent are connected through Line 2. Line 1 has an exponential connection time with parameter $\lambda_1 = 2$ units per minute. Line 2 has an exponential connection time with parameter $\lambda_2 = 3$ units per minute. If a customer takes more than 30 seconds to connect. Compute the probability that the delay was due to Line 1?	*CO3*	*PO3*	**08**
		b)	Waiting time for a metro is uniformly distributed between 0 and 6 minutes, inclusive. i. What is the probability that a person waits for more than 2 minutes? ii. On an average, how long must a person wait?	*CO3*	*PO2*	**05**
		c)	In a university library, there is a photocopier machine that occasionally jams while making copies. The probability of the machine jamming is 35% for each copy made, and these events are independent of each other. A student is assigned to make 2400 copies using this machine. Determine the probability that the photocopier machine will jam between 800 and 850 times while making the copies.	*CO3*	*PO3*	**07**
			UNIT - V			
	7	a)	Explain the terms: population, sample, parameters and statistics with an example	*CO1*	*PO1*	**05**
		b)	The daily closing price of a company stock is: 12.25, 13.50, 14.80, 15.10, 15.10, 16.00, 16.50, 16.90, 17.50, 17.90, 32.00, 18.00. i. Using 1.5(IQR) rule, identify the potential outliers ii. Remove the outliers if any present and compute Median, D_7, P_{32}	*CO3*	*PO3*	**10**
		c)	The score of students in a Machine Learning course is: 85, 78, 92, 88, 79, 92, 75, 81, 90 and 86 i. Compute variance and standard deviation ii. What is the significance of Standard Deviation, when variance is available to measure dispersion in random variables?	*CO3*	*PO2*	**05**

****