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

Autonomous Institute Affiliated to VTU

April 2025 Semester End Make-Up 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)		Discuss the various definitions of probability, including the Classical, Frequentist and Axiomatic approaches. Highlight the merits and demerits of each definition with suitable examples.	CO2	PO1	**06**
	b)		Differentiate between joint, marginal, and conditional probabilities with examples.	CO1	PO1	**07**
	c)		Calculate the probability of drawing an ace followed by a king from a standard deck of 52 cards without replacement.	CO1	PO2	**07**
			OR			
2	a)		What is uncertainty in machine learning, and how is probability used to handle it?	CO1	PO1	**06**
	b)		Justify the following with suitable examples: i. Can events be mutually exclusive and exhaustive. ii. Can events be mutually exclusive and independent.	CO	PO3	**07**
	c)		A box contains 5 red, 4 blue, and 6 green balls. If one ball is drawn at random, what is the probability that it is (i) red, (ii) not blue, or (iii) red or green?	CO1	PO2	**07**
			UNIT - II			
3	a)		State and explain Bayes' theorem with an example.	CO1	PO1	**06**
	b)		What is the Maximum a Posteriori (MAP) estimation? How is it used in density estimation?	CO2	PO1	**07**
	c)		Develop a Naive Bayes classifier for classifying emails as spam or not spam using a hypothetical dataset.	CO2	PO2	**07**
			OR			
4	a)		An email filtering system classifies emails as spam or non-spam based on the presence of the word "win." Given the following probabilities:	CO3	PO2	**08**

		• $P(Spam) = 0.2$ (20% of emails are spam) • $P(Ham) = 0.8$ (80% of emails are not spam) • $P(Win \mid Spam) = 0.6$ (60% of spam emails contain the word "win") • $P(Win \mid Ham) = 0.1$ (10% of non-spam emails contain the word "win") Using Bayes' Theorem, calculate the probability that an email containing the word "win" is actually spam.			
	b)	What are prior and conditional probabilities in the context of Naive Bayes? Illustrate with an example.	CO2	PO1	**07**
	c)	Elaborate on real world applications of Bayes theorem to handle classification problems.	CO1	PO1	**05**
		UNIT - III			
5	a)	Define a discrete random variable and explain its distribution. Provide an example.	CO1	PO1	**06**
	b)	Derive the mean and variance of the Bernoulli distribution.	CO1	PO2	**07**
	c)	A random variable X follows a Binomial distribution with parameters $n = 10$ and $p = 0.6$. Find $P(X = 5)$ and $P(X \le 3)$.	CO2	PO2	**07**
		OR			
6	a)	Explain the difference between joint and marginal distributions with an example.	CO1	PO1	**06**

	b)	Let X and Y be two random variables with $Cov(X, Y) = 6$, $Var(X) = 9$, and $Var(Y) = 4$.
		  i. Calculate the correlation between X and Y.
		  ii. If $a = 2$ and $b = -3$, find the covariance of $aX + b$ and $bX - Y$.
			CO2	PO2	**07**

	c)	A random variable X follows a Poisson distribution with $\lambda = 4$. Calculate the value of 'a' such that $P(X > a) = 0.4$.	CO2	PO3	**07**
		UNIT - IV			
7	a)	Define a continuous random variable and explain the concept of probability density function (PDF) with an example.	CO1	PO2	**06**
	b)	Derive the mean and variance of the exponential distribution.	CO2	PO2	**07**
	c)	A population has a mean of 50 and a standard deviation of 10. Using the Central Limit Theorem, approximate the probability that the mean of a sample of size 25 is greater than 52.	CO2	PO2	**07**
		OR			
8	a)	A continuous random variable X follows a normal distribution with a mean of 50 and a standard deviation of 10. Calculate the probability that X lies between 40 and 60.	CO2	PO2	**06**

		b)	Discuss the Pareto distribution and its applications.	CO1	PO2	**07**
		c)	A factory produces widgets with a mean weight of 100 grams and a standard deviation of 15 grams. If a random sample of 36 widgets is taken, what is the probability that the sample mean weight is between 98 and 102 grams? Use the Central Limit Theorem to solve the problem.	CO3	PO2	**07**
			UNIT - V			
	9	a)	Differentiate between population and sample with examples.	CO2	PO1	**06**
		b)	Calculate the mean, median, and interquartile range (IQR) for the following dataset: {5, 7, 12, 15, 21, 28, 30}.	CO2	PO2	**07**
		c)	The heights (in cm) of five students are given as: 150, 155, 160, 165, 170. i. Compute the mean of the given data. ii. If each student's height is measured relative to 150 cm (i.e., subtract 150 from each value), find the mean of the transformed data. Verify whether the mean remains unchanged due to this shift in origin.	CO1	PO2	**07**
			OR			
	10	a)	Explain the significance of descriptive statistics in data analysis.	CO1	PO1	**06**
		b)	Calculate the variance and standard deviation of the following dataset: {8, 10, 10, 12, 14}.	CO3	PO2	**07**
		c)	In a dataset, the 25th percentile is 18, and the 75th percentile is 42. Determine the interquartile range (IQR) and explain its importance in detecting outliers.	CO2	PO2	**07**

******