RV College of Engineering

Autonomous
Institution Affiliated — Approved by AICTE
to Visvesvaraya — New Delhi
Technological
University, Belagavi

USN: _____ 055

Department of Artificial Intelligence and Machine Learning

Course Code: 21AI41
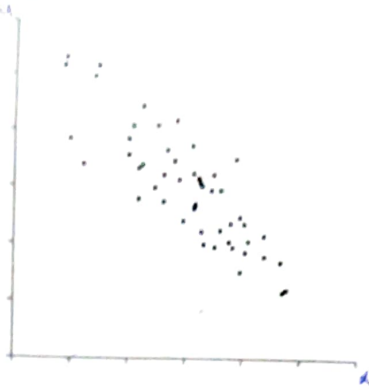**Sem: IV**

Date: 05-Sept-2023
Duration: 90 Minutes

MAKEUP TEST
Statistics for Data Science
Answer all the Questions

SL. No	Questions	M	BT	CO
1	Given that X and Y are two Random variables, is it always true that Var(X +Y) = Var(X) + Var(Y)? Justify your choice.	05	02	01
2	Identify as to when each of the following cases can occur. A and B are events defined on the sample space S. (a) $P(A\|B) = P(A)$ (b) $P(A \cap B) = 0$ (c) $P(A\|B) = 1$	05	02	01
3	A professor tries to count the number of students attending his lecture. For each student in the audience, the professor either counts the student properly with probability p or overlooks with probability $1-p$. The exact number of students in the course is 70. (a) The number of students counted by the professor is a random variable N. What is the PMF of N? (b) Let $U = 70 - N$ denote the number of uncounted students. What is the PMF of U?	05	03	01
4	The daily sales total (excepting Saturday) at a small restaurant has a probability distribution that is approximately normal, with a mean $\mu = \$1230$ per day and a standard deviation σ equal to $\$120$. a) What is the probability that the sales will exceed $\$1400$ for a given day? b) The restaurant must have at least $\$1000$ in sales per day to break even. What is the probability that on a given day the restaurant will not break even?	05	03	02
5	Compute c and $E[X]$ for the following continuous random variable X. $$f_X(x) = cx(1-x^2) \text{ for } 0 \leqslant x \leqslant 1 \text{ and } 0 \text{ otherwise.}$$	05	02	01

USN: _____

Department of Artificial Intelligence and Machine Learning

6	Consider the scatter plot below. The x-axis corresponds to some variable and y-axis another variable (*Do not worry what exact variables they are.*). You figured out in your CIE 2, their correlation was close to -0.85. *Assume that the data was collected by one of your classmates who is known for sleeping, sitting at one of the corners of your classroom.* He wakes up to tell you all that instead of plotting (x, y) he had plotted (y, x). Now, identify among the following, -0.85, -0.1, 0.1, 0.85, 0, which is the correct value of correlation for the correct plot, and justify your choice.	05	04	04
7	List out 5 important requirements for modeling a random variable as a binomial random variable.	05	02	03
8	There is a new movie of a very famous actor that just got released. You visit the theater and conduct a survey among the people on the first day, first show of the release, to review the movie. Explain what could go wrong in your review about the movie with this survey technique. How would you overcome the problem you just identified? *Please restrict your answer to not more than 3 or 4 sentences together for the two questions.*	05	04	04
9	Suppose that 70% of the families in your (very large) city have no dogs, 22% have 1 dog and 8% have 2 dogs. (a) Let X be the number of dogs that a randomly chosen family has. a) Compute $E(X)$ and $Var(X)$. b) Assuming your 200 family neighborhood constitutes a random sample and that families make their choices about dog ownership independently approximate the probability that there are more than 90 dogs in your neighborhood. *Use the central limit theorem.*	05	03	02
10	Is it possible to plot 4 dimensional data in 2 dimensions? Justify with not more than 3 reasons, your choice.	05	03	02

Marks Distribution	Particulars	CO1	CO2	CO3	CO4	CO5	L1	L2	L3	L4	L5
	Max Marks	20	15	05	10	--	--	20	20	10	--