

National Institute of Electronics and Information Technology, Calicut

Certified AI Professional

Assignment 7 (Total Marks 20)

Probability Distributions (Binomial & Poisson Distribution)

(Submit the programs from 1 to 5 as A7Part1.R, Questions 1-2 carry marks each)

1. You are conducting an experiment in which you need to examine 40 full grown potato plants. You plant 60 of the plants, knowing that 65% of the plants will reach the fully grown state.
 1. What is the probability that exactly 40 of the 60 plants will be fully grown?
 2. What is the probability that at least 40 of the 60 plants will be fully grown?
2. Calculate the probability that in 60 tosses of a fair coin the head comes up
 - a) 20,25 or 30 times
 - b) less than 20 times
 - c) between 20 and 30 times
3. Generate a series of Poisson distributions for different λ values. Keep the number of items in all the distributions as 100. Observe and explain what happens to the distribution when value of λ changes.
4. The emission of alpha particles by polonium fits a Poisson distribution. Generate 2608 numbers as per this distribution with Poisson rate parameter $\lambda = 10097/2608$. Plot a histogram of the same.
5. A random variable X has Poisson distribution with mean 7. Find the probability that
 - a. X is less than 5
 - b. X is greater than 10
 - c. X is between 4 and 16

National Institute of Electronics and Information Technology, Calicut

Certified AI Professional

Machine Learning

(Submit the programs from 6 to 8 as A7Part2.py

Question 6 & 7 carry 3 Marks each and 8 carry 4 Marks)

6. Prepare an ML model using KNN Classifier to predict the Species information for a given iris flower using Sepal Length, Sepal Width, Petal Length & Petal Width. Use the complete iris dataset for training. Use it to predict the species of an iris flower.
7. Print the Accuracy Score and Confusion matrix for KNN Classifier using iris data. (Split iris dataset to train and test sets.)
8. Using diabetes dataset in sklearn, prepare an ML model using Linear regression to predict the value of disease progression. Print Root Mean Square Error values for a test data after splitting the data into train and test set. Also print the co-efficients of the linear model. (hint: model.coef_)