

CA 5314: Programming Assignment 1

Linear Regression and Multiple Regression

Aim:

Create a simple Python program for linear regression and multiple regression using the given datasets. Present the regression model coefficient and intercept, coefficient of determination (R^2), and root mean squared error (RMSE). The PA1.zip file contains the data sets that you will require for this assignment.

Linear Regression

1. Pizza Franchise (Dataset 1)

In the following data

- X = annual franchise fee (\$1000)
- Y = start-up cost (\$1000) for a pizza franchise

2. National Unemployment Male Vs. Female (Dataset 2)

In the following data pairs

- X = national unemployment rate for adult males
- Y = national unemployment rate for adult females

3. Fire and Theft in Chicago (Dataset 3)

In the following data pairs

- X = fires per 1000 housing units
- Y = thefts per 1000 population within the same Zip code in the Chicago metro area

Multiple Regression

4. Basketball (Dataset 4)

The following data (X1, X2, X3, X4, X5) are for each player.

- X1 = height in feet
- X2 = weight in pounds
- X3 = percent of successful field goals (out of 100 attempted)
- X4 = percent of successful free throws (out of 100 attempted)
- X5 = average points scored per game

5. Crime (Dataset 5)

The data (X1, X2, X3, X4, X5, X6, X7) are for each city.

- X1 = total overall reported crime rate per 1 million residents
- X2 = reported violent crime rate per 100,000 residents
- X3 = annual police funding in \$/resident
- X4 = % of people 25 years+ with 4 yrs. of high school
- X5 = % of 16 to 19 year-olds not in highschool and not highschool graduates.
- X6 = % of 18 to 24 year-olds in college
- X7 = % of people 25 years+ with at least 4 years of college