Roll No. 41163

Page No.
Date: / /

Assignment A1

Execution Date : 27/01/2021
Submission Date : 19/05/2021

Title: Linear Regression

Problem Statement:

The following table shows the results of a recently conducted study on the correlation of the number of hours spent driving with the risk of developing acute backache. Find the equation of the best fit line for this data.

No. of hours spent driving (x)	Risk score on a scale of 100 (y)
10	95
9	80
2	10
15	50
10	45
16	48
11	38
16	93

~~Outcome~~:

Objective:

To understand and implement linear regression algorithm.

Page No.
Date: / /

Outcome:
Will be able to understand how to find correlation between the variables and how to calculate accuracy of the linear model.

Software Requirements:
- Jupyter Notebook
- Python 3.8.5
- 64 bit operating system

Hardware Requirements:
- Computer with 64 bit processor.

Theory:

Regression analysis is one of the most widely used statistical techniques. It estimates relationship among a dependent variable and an independent variable.

Linear Regression:
Least square ~~method~~ regression is the method of predicting the value of dependent variable y based on the value of independent variable x.

~~The~~ least square ~~regression line~~:
Given the data

Page No.
Date: / /

Given the random sample of observations population the least square regression line is estimated by

$$y = b0 + b1x$$

where, b0 is constant
b1 is regression coefficient
x is the value of independent variable
y is the value of dependent variable.

To find the relationship between dependent and independent variable we find correlation

The formula to find correlation using pearson correlation is:

$$r = \frac{\text{covariance}(x,y)}{sd(x) \cdot sd(y)}$$

where, s.d. = standard deviation.

We know the regression model:

$$y = b0 + b1x$$

The formula for b0 and b1 are:

$$b1 = \frac{\Sigma(x_i - \bar{x}) * (y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$$

$$b0 = \bar{y} - b1\bar{x}$$

Page No.
Date: / /

Best line:

The best line is a straight line the represents the data on scatter plot.

This line may pass through all the points or none of the points or some of the points.

Algorithm:

- Import required packages
- Read given dataset
- Import the linear regression and create object of it
- Find accuracy of model using score function
- Predict the value using regressor object.
- Take input from user
- Calculate value of y
- Draw scatter plot

Conclusion:

Thus, we learned to find the trend of data by using X as independent variable and y as dependent variable by using linear regression.