

Assignment. No.01

Assignment on Linear Regression:

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.**

Number of hours spent driving (x)	Risk score on a scale of 0-100 (y)
10	95
9	80
2	10
15	50
10	45
16	98
11	38
16	93

Linear Regression

- Linear regression is a **linear model**, e.g. a model that assumes a linear relationship between the input variables (x) and the single output variable (y).
- More specifically, that y can be calculated from a linear combination of the input variables (x).
- When there is a single input variable (x), the method is referred to as **simple linear regression**.
- When there are **multiple input variables**, literature from statistics often refers to the method as **multiple linear regression**

Linear Regression

- In Linear regression **these two variables** are related through an equation , where exponent (power) of both these variables is **1**.
- Mathematically a linear relationship represents a straight line when plotted as a graph.
- A non-linear relationship where the exponent of any variable is not equals to 1 creates a curve.
- The general mathematical equation for a LR is as,
 - **$y = ax + b$**
 - where,
 - - **y** is the response variable
 - - **x** is the predictive variable
 - **a & b** are the constants which are called the coefficient.

Steps to Establish Linear Regression

- A simple example of regression is predicting **weight** of a person when his **height** is known. To do this we need to have the relationship between **height** and **weight** of a person.
- The steps to create the relationship is,
 - - carry out the experiment of gathering a sample of observed values of height & corresponding weight.
 - - create the object of linear regression class.
 - - Train the algorithm with dataset of X and y.
 - - Get a summary of the relationship model to know the average error in prediction also called residuals.
 - - To predict the weight of new Persons, use the ***prediction() function***.

Simple Linear Regression

- Simple regression problem (a single x and a single y), the form of the model would be:

Constant

Coefficient

$$y = b0 + b1 * x1$$

**Dependent
variable (DV)**

**Independent
variable (IV)**

Simple Linear Regression

ANALYZING DATASET

IV

Number of hours spent driving (x)	Risk score on a scale of 0-100 (y)
10	95
9	80
2	10
15	50
10	45
16	98
11	38
16	93

DV

Simple Linear Regression

- LET's CODE!
- Prep your Data Preprocessing Template
 - **Import** Dataset
 - **No** need for *Missing Data*
 - Splitting into **Training** & **Testing** dataset
 - Keep **Feature Scaling** but least preferred here
- Co-relate Hours Spent with Risk Score
- Later carry out prediction
- Verify the Values of prediction
- Prediction on TEST SET

Algorithm

Algorithm:

1. Import the Required Packages
2. Read Given Dataset
3. Import the Linear Regression and Create object of it
4. Find the Accuracy of Model using Score Function
5. Predict the value using Regressor Object
6. Take input from user.
7. Calculate the value of y
8. Draw Scatter Plot

Python Packages Needed

Pandas:

- Data Analytics

Numpy:

- numeric Computing

Matplotlib.pyplot:

- Plotting graphs

Sklearn:

- Regression Classes