**Name :** Prem Vinod Bansod
**Roll No :** 41310
**Assignment No :** 01 (ML)

**Title:** To implement Linear Regression to find the equation of the best fit line for given data.

**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,

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

## Objective:

- The Basic Concepts of Linear Regression.
- Implementation logic of Linear Regression to find the equation of the best fit line for given data.

## Theory:

## Introduction:

Regression analysis is a very widely used statistical tool to establish a relationship model between two variables. One of these variable is called predictor variable whose value is gathered through experiments. The other variable is called response variable whose value is derived from the predictor variable. Linear models are the simplest parametric methods and always deserve the right attention, because many problems,

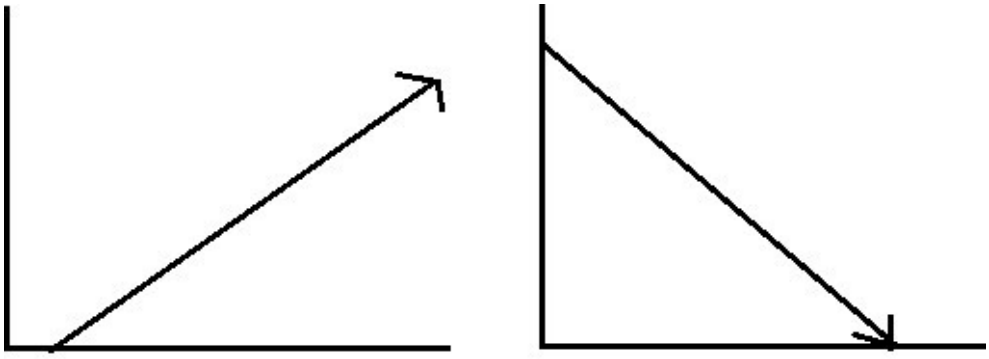
even intrinsically non-linear ones, can be easily solved with these models. A regression is a prediction where the target is continuous and its applications are several, so it's important to understand how a linear model can fit the data, what its strengths and weaknesses are, and when it's preferable to pick an alternative.

**Types of Regression**

- Linear
- Multiple Linear
- Logistic
- Polynomial

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 equal to 1 creates a curve.

The general mathematical equation for a linear regression is – $y = ax + b$
y is the response variable and x is the predictor variable.
a and b are constants which are called the coefficients.




**Applications**

- Trend lines: A trend line represents the variation in some quantitative data with passage of time (like GDP, oil prices, etc.). These trends usually follow a linear relationship. Hence, linear regression can be applied to predict future values.
- Economics: To predict consumption spending, fixed investment spending, inventory investment, purchases of a country‘s exports, spending on imports, the demand to hold liquid assets, labor demand, and labor supply.
- Finance: Capital price asset model uses linear regression to analyze and quantify the systematic risks of an investment.

- Biology: Linear regression is used to model causal relationships between parameters in biological systems.

**Python Packages needed**

- pandas
  Data Analytics
- numpy
  Numerical Computing
- matplotlib.pyplot
  Plotting graphs
- sklearn
  Regression Classes

**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 and 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 predict() function.

**Conclusion:** We have studied the Linear Regression and also implemented successfully.