Practical no: 5

Aim: Implement Simple and Multiple Linear Regression Models.

- Regression The term regression is used when you try to find the relationship between variables.
- In Machine Learning, and in statistical modeling, that relationship is used to predict the outcome of future events.
- Linear Regression Linear regression uses the relationship between the data points to draw a straight line through all of them.
- This line can be used to predict future values.
- Regression The term regression is used when you try to find the relationship between variables.
- In Machine Learning, and in statistical modeling, that relationship is used to predict the outcome of future events.
- Linear Regression Linear regression uses the relationship between the data-points to draw a straight line through all of them.
- This line can be used to predict future values.

Why do we use Regression Analysis?

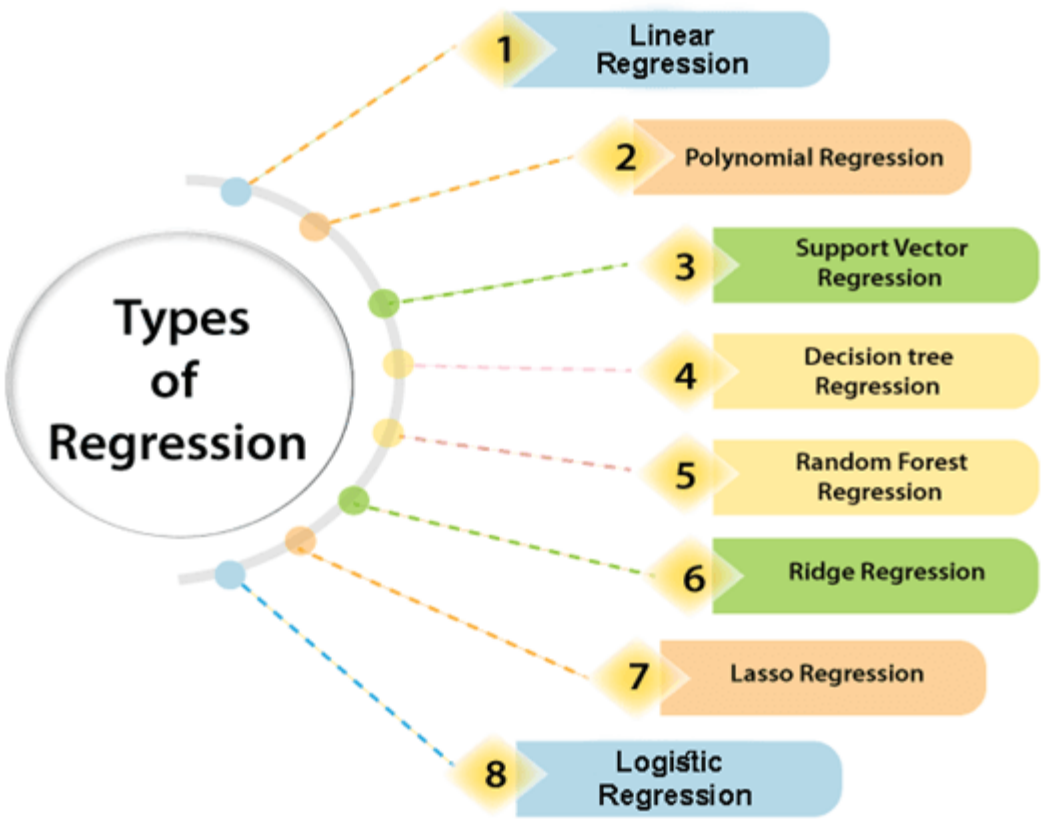
As mentioned above, Regression analysis helps in the prediction of a continuous variable. There are various scenarios in the real world where we need some future predictions such as weather condition, sales prediction, marketing trends, etc., for such case we need some technology which can make predictions more accurately. So for such case we need Regression analysis which is a statistical method and used in machine learning and data science. Below are some other reasons for using Regression analysis:

- Regression estimates the relationship between the target and the independent variable.
- It is used to find the trends in data.
- It helps to predict real/continuous values.
- By performing the regression, we can confidently determine the **most important factor, the least important factor, and how each factor is affecting the other factors**.

Types of Regression

There are various types of regressions which are used in data science and machine learning. Each type has its own importance on different scenarios, but at the core, all the regression methods analyze the effect of the independent variable on dependent variables. Here we are discussing some important types of regression which are given below:

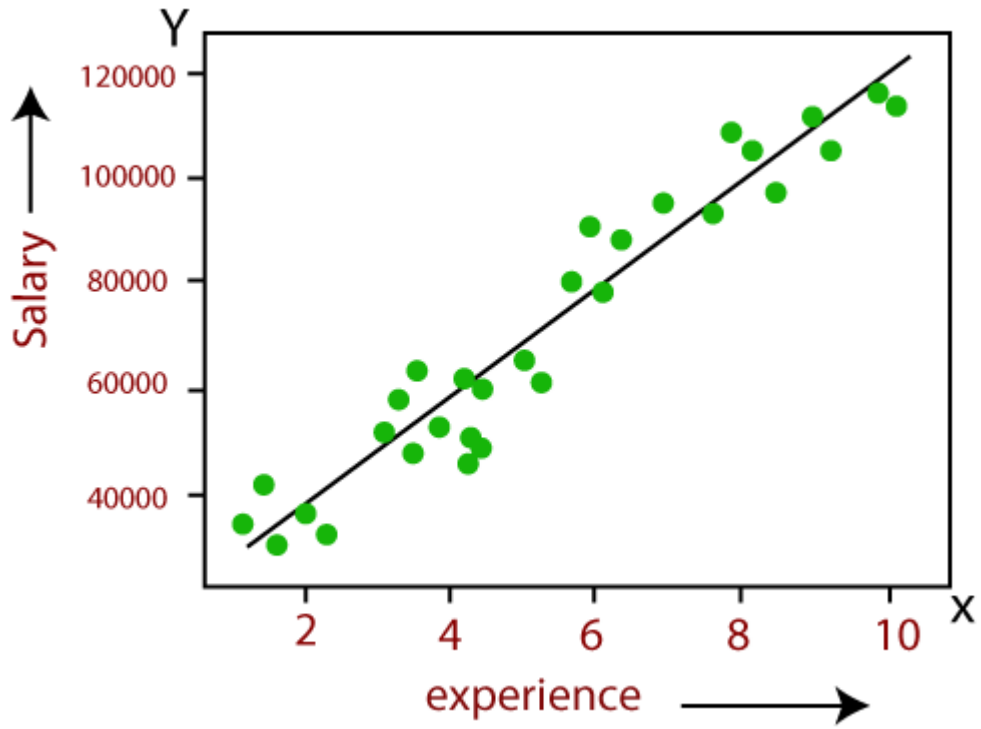
- **Linear Regression**
- **Logistic Regression**
- **Polynomial Regression**
- **Support Vector Regression**
- **Decision Tree Regression**
- **Random Forest Regression**
- **Ridge Regression**
- **Lasso Regression:**



Linear Regression:

- Linear regression is a statistical regression method which is used for predictive analysis.
- It is one of the very simple and easy algorithms which works on regression and shows the relationship between the continuous variables.
- It is used for solving the regression problem in machine learning.
- Linear regression shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis), hence called linear regression.

- If there is only one input variable (x), then such linear regression is called **simple linear regression**. And if there is more than one input variable, then such linear regression is called **multiple linear regression**.
- The relationship between variables in the linear regression model can be explained using the below image. Here we are predicting the salary of an employee on the basis of **the year of experience**.



- Below is the mathematical equation for Linear regression:

1. Y= aX+b

Here, Y = dependent variables (target variables),
X= Independent variables (predictor variables),
a and b are the linear coefficients

Some popular applications of linear regression are:

- **Analyzing trends and sales estimates**
- **Salary forecasting**
- **Real estate prediction**
- **Arriving at ETAs in traffic.**

Multiple Linear Regression

In the previous topic, we have learned about Simple Linear Regression, where a single Independent/Predictor(X) variable is used to model the response variable (Y). But there may be various cases in which the response variable is affected by more than one predictor variable; for such cases, the Multiple Linear Regression algorithm is used.

Moreover, Multiple Linear Regression is an extension of Simple Linear regression as it takes more than one predictor variable to predict the response variable. We can define it as:

Multiple Linear Regression is one of the important regression algorithms which models the linear relationship between a single dependent continuous variable and more than one independent variable.

Some key points about MLR:

- For MLR, the dependent or target variable(Y) must be the continuous/real, but the predictor or independent variable may be of continuous or categorical form.
- Each feature variable must model the linear relationship with the dependent variable.
- MLR tries to fit a regression line through a multidimensional space of data-points.

MLR equation:

In Multiple Linear Regression, the target variable(Y) is a linear combination of multiple predictor variables x_1, x_2, x_3, ...,x_n. Since it is an enhancement of Simple Linear Regression, so the same is applied for the multiple linear regression equation, the equation becomes:

1.

Y= b₀+b₁x₁+ b₂x₂ + b₃x₃+..... bnxn ..... (a)

Where,

Y= Output/Response variable

b_0, b_1, b_2, b_3 , b_n....= Coefficients of the model.

x_1, x_2, x_3, x_4,...= Various Independent/feature variable

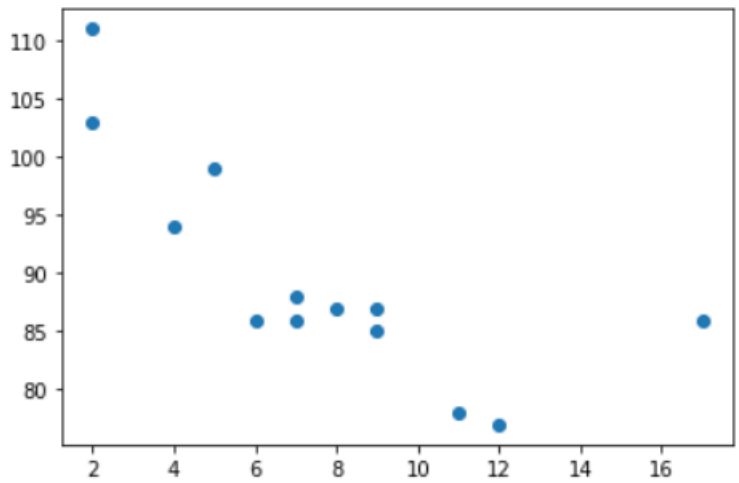
Assumptions for Multiple Linear Regression:

- A **linear relationship** should exist between the Target and predictor variables.
- The regression residuals must be **normally distributed**.

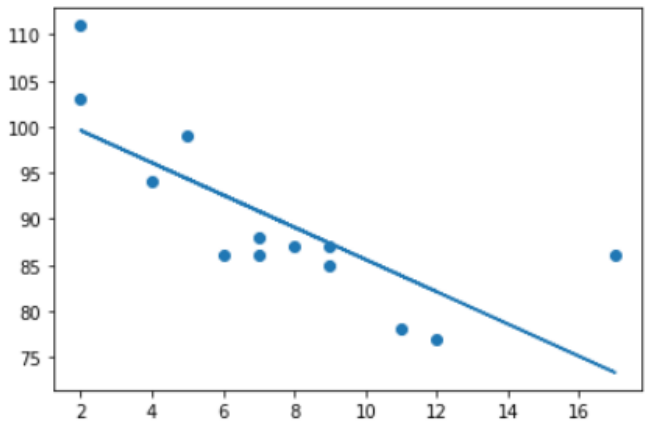
- MLR assumes little or **no multicollinearity** (correlation between the independent variable) in data.

Program

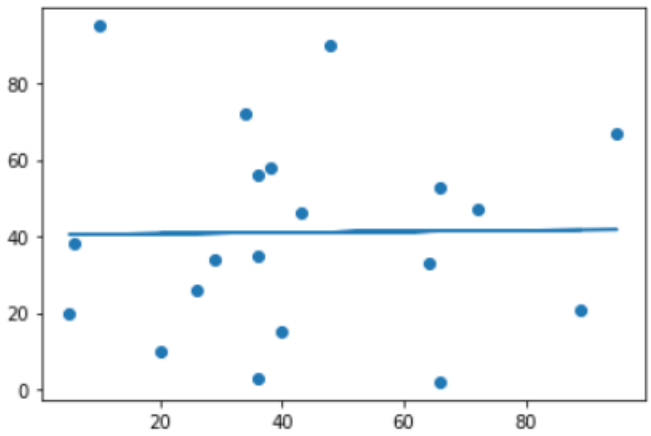
```
import matplotlib.pyplot as plt

x = [5,7,8,7,2,17,2,9,4,11,12,9,6]
y = [99,86,87,88,111,86,103,87,94,78,77,85,86]

plt.scatter(x, y)
plt.show()
```



```
[ ]
import matplotlib.pyplot as plt
#SciPy is a scientific computation library that uses NumPy underneath.
#SciPy stands for Scientific Python.
#It provides more utility functions for optimization, stats and signal processing.
#Like NumPy, SciPy is open source so we can use it freely.
#SciPy was created by NumPy's creator Travis Olliphan
from scipy import stats

x = [5,7,8,7,2,17,2,9,4,11,12,9,6]
y = [99,86,87,88,111,86,103,87,94,78,77,85,86]

slope, intercept, r, p, std_err = stats.linregress(x, y)

def myfunc(x):
  return slope * x + intercept

mymodel = list(map(myfunc, x))

plt.scatter(x, y)
plt.plot(x, mymodel)
plt.show()
```



```
[ ]  import matplotlib.pyplot as plt
     from scipy import stats

     x = [89,43,36,36,95,10,66,34,38,20,26,29,48,64,6,5,36,66,72,40]
     y = [21,46,3,35,67,95,53,72,58,10,26,34,90,33,38,20,56,2,47,15]

     slope, intercept, r, p, std_err = stats.linregress(x, y)

     def myfunc(x):
       return slope * x + intercept

     mymodel = list(map(myfunc, x))

     plt.scatter(x, y)
     plt.plot(x, mymodel)
     plt.show()
```




```
import pandas
from sklearn import linear_model

df = pandas.read_csv("/content/drive/MyDrive/Colab Notebooks/dataset /data.csv")

X = df[['Weight', 'Volume']]
y = df['CO2']

regr = linear_model.LinearRegression()
regr.fit(X, y)

#predict the CO2 emission of a car where the weight is 2300kg, and the volume is 1300
predictedCO2 = regr.predict([[2300, 1300]])

print(predictedCO2)
```

Coefficient

The coefficient is a factor that describes the relationship with an unknown variable.

Example: if x is a variable, then 2x is x two times. x is the unknown variable, and the number 2 is the coefficient.

In this case, we can ask for the coefficient value of weight against CO2, and for volume again st CO2. The answer(s) we get tells us what would happen if we increase, or decrease, one of t he independent values.

```
[ ]  import pandas
     from sklearn import linear_model

     df = pandas.read_csv("/content/drive/MyDrive/Colab Notebooks/dataset /data.csv")

     X = df[['Weight', 'Volume']]
     y = df['CO2']

     regr = linear_model.LinearRegression()
     regr.fit(X, y)



     print(regr.coef_)
```

```
[0.00755095 0.00780526]
```

Result Explained The resulting array represents the coefficient values of weight and volume.

Weight: 0.00755095 Volume: 0.00780526

These values tell us that if the weight increase by 1kg, the CO2 emission increases by 0.00755095g.

And if the engine size (Volume) increases by 1 cm3, the CO2 emission increases by 0.00780526 g.

I think that is a fair guess, but let's test it!

We have already predicted that if a car with a 1300cm3 engine weighs 2300kg, the CO2 emission will be approximately 107g.

Conclusion: Hence we have Implement Simple and Multiple Linear Regression Models.