**Name:- Khalfe Aynas Roll No:- 14CO12**

**Exp No:- 04 Batch No:- 01**

**Aim:-** Implementation of Linear Regression.

**Theory:-**

**Linear Regression:-**

Linear regression involves finding the “best” line to fit two attributes (or variables), so that one attribute can be used to predict the other.

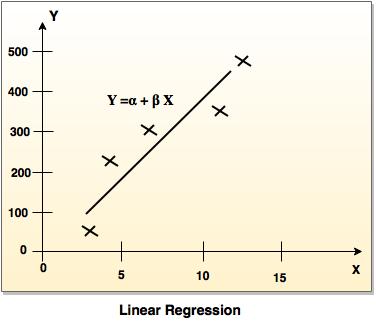
**1. Straight-line regression:**

* Straight-line regression analysis involves a response variable, y, and a single predictor variable, x.
* It is the simplest form of regression, and models y as a linear function of x.
* That is, y = b+wx;

where the variance of y is assumed to be constant,band w are regression coefficients specifying the Y-intercept and slope of the line, respectively. - These coefficients can be solved by the method of least squares, which estimates the best-fitting straight line as the one that minimizes the error between the actual data and the estimate of the line. - The regression coefficients can be estimated using this method with the following equations:

**2. Multiple linear regression:**

* Multiple linear regressionis an extension of straight-line regression so as to involve more than one predictor variable.
* It allows response variable y to be modeled as a linear function of n predictor variables or attributes.
* The equations(obtained from the method of least squares ), become long and are tedious to solve by hand.
* Multiple regression problems are instead commonly solved with the use of statistical software packages, such as SAS, SPSS, and S-Plus
* **Speed and Scalability:** Time to construct the model and also time to use the model.
* **Robustness:** This is the ability of the classifier to make correct predictions given noisy data or data with missing values
* **Scalability:** This refers to the ability to construct the classifier efficiently given large amounts of data.
* **Interpretability:** This refers to the level of understanding and insight that is provided by the classifier
* **Goodness of rules:** Decision tree size compactness of classification rules.



**Program:-**

**#lineaerregression.py**

import numpy as np

import pandas as pd

import matplotlib as mpl

#mpl.use('TkAgg')

import matplotlib.pyplot as plt

dataset=pd.read\_csv('Salary\_Data.csv')

X=dataset.iloc[:,:-1].values

y=dataset.iloc[:,1:].values

#splitting the dataset into training and test sets

from sklearn.model\_selection import train\_test\_split

#used model\_selection in place of cross\_validation since the latter is deprecated

X\_train,X\_test,y\_train,y\_test=train\_test\_split(X,y,test\_size=1/3,random\_state=0)

#fitting simple linear regression to the training set

from sklearn.linear\_model import LinearRegression

regressor=LinearRegression()

regressor.fit(X\_train,y\_train)

#predicting the test set results

y\_pred=regressor.predict(X\_test)

#visualising the set of results

plt.scatter(X\_train,y\_train,color='red')

plt.plot(X\_train,regressor.predict(X\_train),color='blue')

plt.title('Salary Vs Experience (Training set)')

plt.xlabel

**Output:-**



**Conclusion:-**

In this experiment, we have came across how linear regression algorithm works.