## How to perform Linear regression with Linear Regression with Python Scikit Learn

## Linear regression is a statistical approach that models the relationship between input features and output. The input features are called the ****independent variables****, and the output is called ****a dependent variable****.

## In this regression task we will predict the percentage of marks that a student is expected to score based upon the number of hours they studied. This is a simple linear regression task as it involves just two variables.

## Simple linear regression mathematically can be given by

## Linear regression equation.

## Source: <https://medium.com/towards-artificial-intelligence/machine-learning-algorithms-for-beginners-with-python-code-examples-ml-19c6afd60daa>

## Tool to be use: Jupyter , Excel and python library whenever needed

## So lets start the task

## Since we will going to use various library so we need to import them, before that lets us know use of these library first

1. **Numpy**: It is a math library to work with n-dimensional arrays in Python. It enables us to do mathematical computation over given data in very efficient way.
2. **Scipy**: this library used in many ways. Scipy is a functional library for scientific and high-performance computations.
3. **Matplotlib**: It is a trendy plotting package that provides plotting of the chart for two or three dimensions.
4. **Scikit-learn**: this is most used library in machine learning as it has various function to perform classification, regression and clustering algorithm

## 

## Another necessary step to start with is reading the dataset present in the .xlsx or .csv format

## 

## Now we have to select the feature so in this feature is Scores and Hours as this is very simple dataset

## Before selecting the feature if there is any noise than we have to clean it

## After that will start with linear regression steps

## So in this we have to select x and y value which denote feature and target value

## So in general term our target ‘y’ is always stay at he last column so we can apply below method for simplicity

## 

## Now we have to split the data into test and train data

## 

## Than we have to apply linear regression function into this

## 

## After this we have to fit the data by using fir function

## 

## After this we have to check the coefficient of determination

## 

## After knowing the coefficient of determination lets find out b0 value which is intercept

## 

## Now lets check another factor which is important i.e. slope

## 

## Now let’s plot the graph

## 

## Let’s check the algorithm by predicting the some test data value

## 

## To check predicted score we can apply below method

## 

## We can compare actual and predicted value for better understanding of the accuracy

## 

## Lets check for predicted score of the student who studied for 9.25 hrs/day

## 

## Now if we check the graph we can see linear graph with most value line up there

## 

## we can check the predicted and observed error using metrics

## 

## It may happend that the data that got collected is cannot be perform using linear regression. Most of the time data follow polynomial trend where data are more ways in the non linear manner

## 

## Source: <https://medium.com/towards-artificial-intelligence/machine-learning-algorithms-for-beginners-with-python-code-examples-ml-19c6afd60daa>