MANUAL TO USE LinReg PYTHON LIBRARY

-Vikram Koti Mourya Vangara

Follow the steps below to understand and use the library for your use:

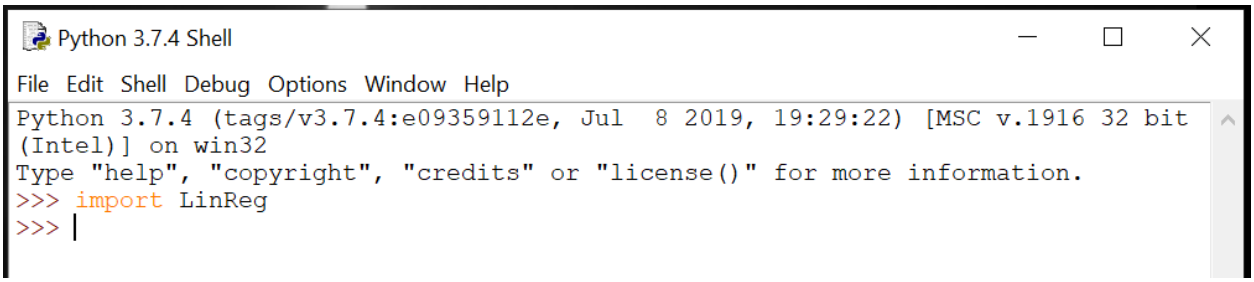
Step 1: Installation

Download the LinReg.py file and paste in the 'site-packages' folder where you have installed python in your computer. For example: C:\python37\Lib\site-packages – This is the directory of my installation folder.

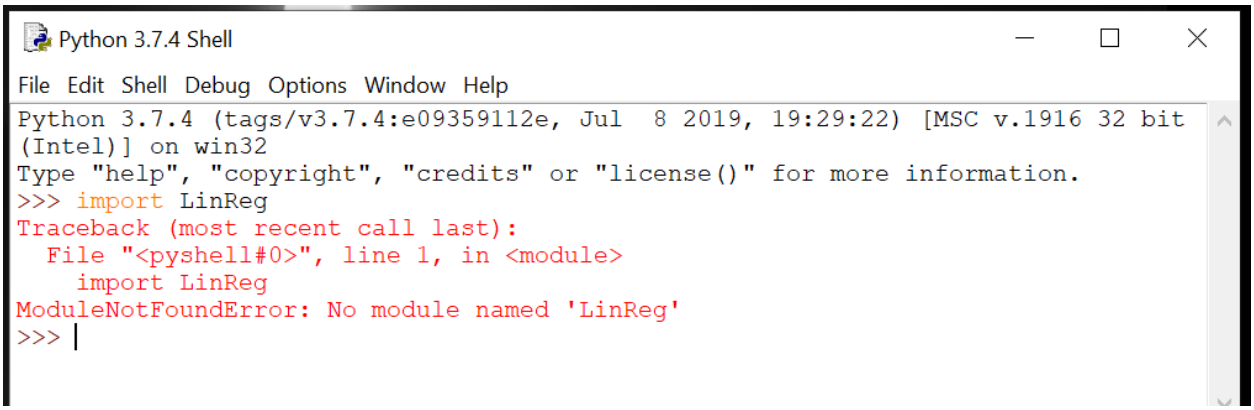
Navigate to this location and simply paste the file here. Now you are all set to move ahead.

Step 2: Verification of Installation and pre-requisites

To verify the successful installation of the library. Open the Python IDLE Shell. (see below figure)

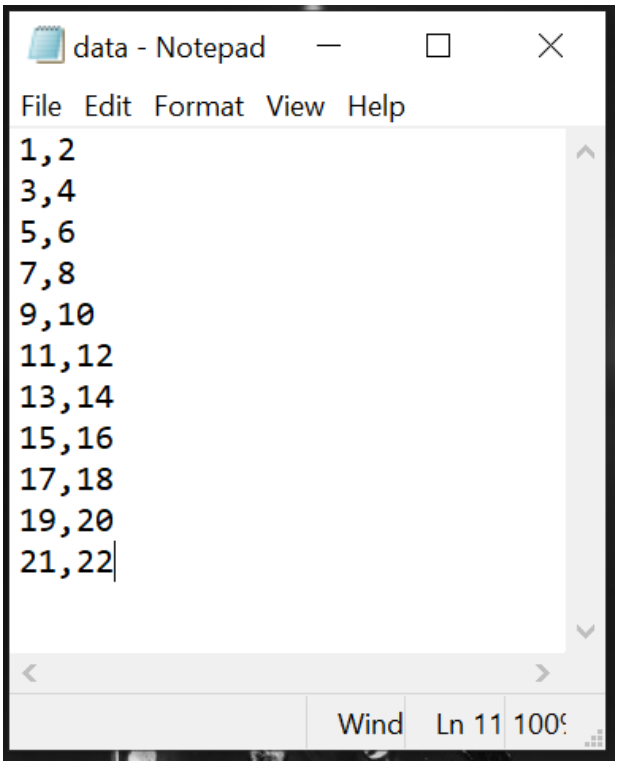


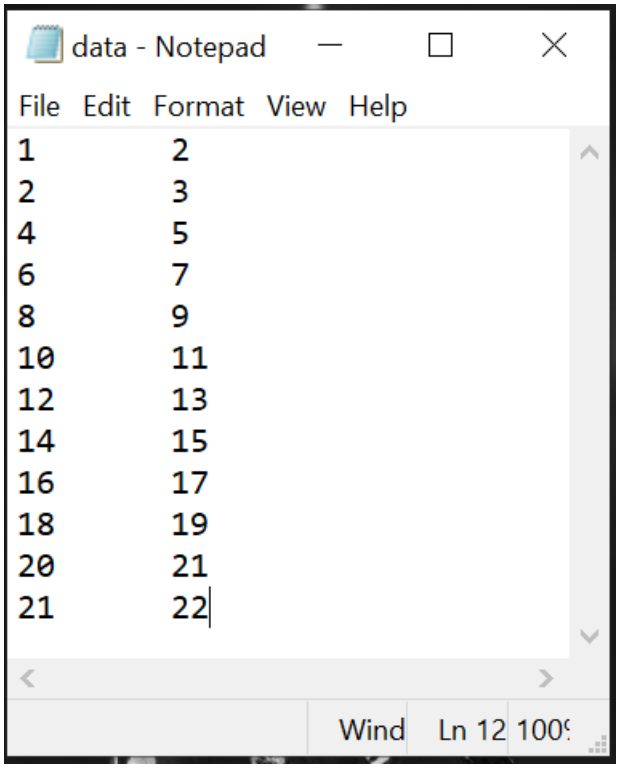
Type 'import LinReg'. It should simply pass on to next line without showing any error. If it was not copied in the desired location, an error will pop up. (see figure below)



After you have successfully finished with installation, prepare your data. As mentioned in the repository, this library only supports .txt and .csv files with only comma and tab separated values. To prepare that data, there are many methods, few of them are mentioned below.

First method is creating a .txt file with either tab separated values or comma separated values as shown below.





If you have an excel file with such data, you can simply copy the two-column data without any labels and paste it in the .txt file.

Another method is an .csv file. To create this, in the excel sheet where you have the two-column data without labels, save the excel file as .csv using 'Save As' option.

Step 3: Basic functions in the library

This library has functions to find the slope, intercept (regression coefficients) and the evaluation metrics like Mean Absolute Error, Mean Squared Error, Root Mean Squared Error.

First you have to train the model. To do this, use the train function after importing the module. It returns a list containing the values of slope and intercept.

```
import LinReg

x = LinReg.train('/Users/Vikram/Desktop/data.txt', stype='tab')

print(x)
```

Here, ‘train’ function takes 2 arguments, i.e. file location and stype. Stype is the separation type used in the file. It can take only two values, i.e. ‘tab’ and ‘csv’. Default is set to ‘csv’.

After training the model. You can predict the dependent value with given independent value and vice versa using x_predict and y_predict functions. The usage of the same is mentioned below.

```
Y_pred = LinReg.y_predict([11,23,23,54,93])

X_pred = LinReg.x_predict([23,83,35,102,34])
```

For using the above functions, you need list as an input. For this you can use the ‘get_list’ function to get the list of two-columns in the data file. The format is as below.

```
X,Y = LinReg.get_list('/Users/Vikram/Desktop/data.txt', stype='tab')
```

In this library, we have also created a class named ‘Metrics’ which contains the functions to compute the evaluation metrics like Mean Absolute Error, Mean Squared Error, Root Mean Squared Error. Let ‘x’ represent the independent values and ‘y’ represent dependent values. To compute evaluation metrics, you need two lists, one is y-predicted values and y-testing values. You can get the y-testing values list from data file using ‘get_list’ function. For y-predicted values, you can use the ‘y_predict’ function by passing the x-testing values as the argument.

Let y-testing values be named as ‘y_test’ and y-predicted values be ‘y_pred’.

```
#Mean Absolute Error
Mae = LinReg.Metrics.MAE(y_test,y_pred)


#Mean Squared Error
Mse = LinReg.Metrics.MSE(y_test,y_pred)

#Root Mean Squared Error
Rmse = LinReg.Metrics.RMSE(y_test,y_pred)
```

Thank you!