Assessment 1

Shubham Gopale
18MCS0066

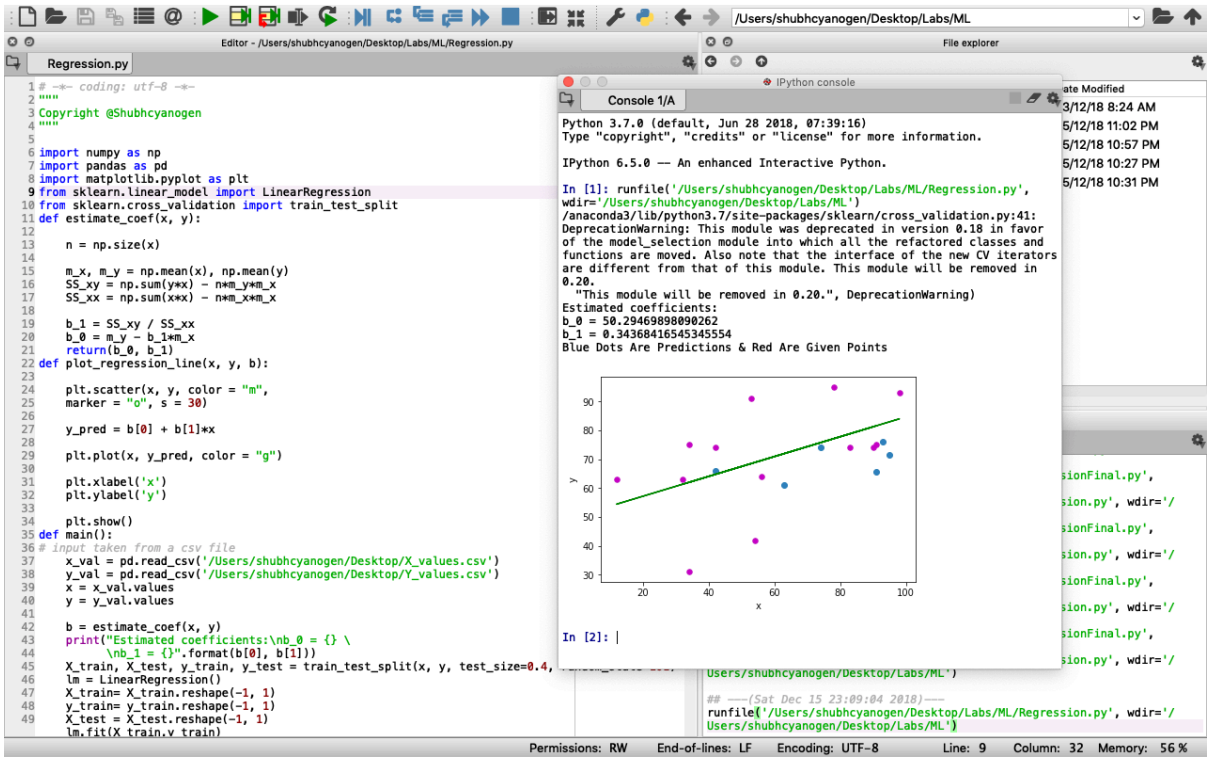
Que: Read the data from the excel file with two coordinates x and y. Divide the dataset into training and testing. Apply linear regression and logistic regression for the training data. For the test data, given x value predict the y value.

Code :

```
# -*- coding: utf-8 -*-
"""
Copyright @Shubhcyanogen
"""

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.cross_validation import train_test_split
def estimate_coef(x, y):

    n = np.size(x)

    m_x, m_y = np.mean(x), np.mean(y)
    SS_xy = np.sum(y*x) - n*m_y*m_x
    SS_xx = np.sum(x*x) - n*m_x*m_x

    b_1 = SS_xy / SS_xx
    b_0 = m_y - b_1*m_x
    return(b_0, b_1)
def plot_regression_line(x, y, b):

    plt.scatter(x, y, color = "m",
    marker = "o", s = 30)

    y_pred = b[0] + b[1]*x

    plt.plot(x, y_pred, color = "g")

    plt.xlabel('x')
    plt.ylabel('y')

    plt.show()
def main():
# input taken from a csv file
    x_val = pd.read_csv('/Users/shubhcyanogen/Desktop/X_values.csv')
    y_val = pd.read_csv('/Users/shubhcyanogen/Desktop/Y_values.csv')
    x = x_val.values
    y = y_val.values

    b = estimate_coef(x, y)
```

```
    print("Estimated coefficients:\nb_0 = {} \
        \nb_1 = {}".format(b[0], b[1]))
    X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.4, random_state=101)
    lm = LinearRegression()
    X_train= X_train.reshape(-1, 1)
    y_train= y_train.reshape(-1, 1)
    X_test = X_test.reshape(-1, 1)
    lm.fit(X_train,y_train)
    predictions = lm.predict(X_test)
    plt.scatter(y_test,predictions)
    print("Blue Dots Are Predictions & Red Are Given Points")
# plotting regression line
    plot_regression_line(x, y, b)
if __name__ == "__main__":
    main()
```

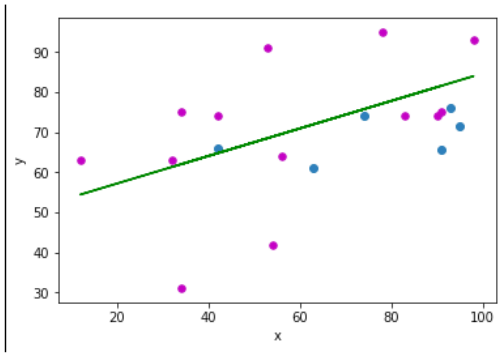
Output :



IPython console

Console 1/A

```
Python 3.7.0 (default, Jun 28 2018, 07:39:16)
Type "copyright", "credits" or "license" for more information.

IPython 6.5.0 -- An enhanced Interactive Python.

In [1]: runfile('/Users/shubhcyanogen/Desktop/Labs/ML/Regression.py',
wdir='/Users/shubhcyanogen/Desktop/Labs/ML')
/anaconda3/lib/python3.7/site-packages/sklearn/cross_validation.py:41:
DeprecationWarning: This module was deprecated in version 0.18 in favor
of the model_selection module into which all the refactored classes and
functions are moved. Also note that the interface of the new CV iterators
are different from that of this module. This module will be removed in
0.20.
  "This module will be removed in 0.20.", DeprecationWarning)
Estimated coefficients:
b_0 = 50.29469898090262
b_1 = 0.34368416545345554
Blue Dots Are Predictions & Red Are Given Points
```




```
In [2]:
```

	A
1	12
2	34
3	56
4	78
5	90
6	12
7	54
8	83
9	91
10	34
11	98
12	53
13	42
14	32

	A
1	54
2	75
3	64
4	95
5	74
6	63
7	42
8	74
9	75
10	31
11	93
12	91
13	74
14	63
15	