Linear Regression Lab Task-4 _Jithin

August 7, 2018

0.1 Linear Regression Lab Task, Using a Random Numbers Dataset

a) Load and plot values on a graph. Plot it as connected dots(line) and scattered dots.

b) Plot x vs yˆ2 graph

c) Plot best fit line using stats library.

d) Compare the best fit line with another line joining the two extremes, any two points.

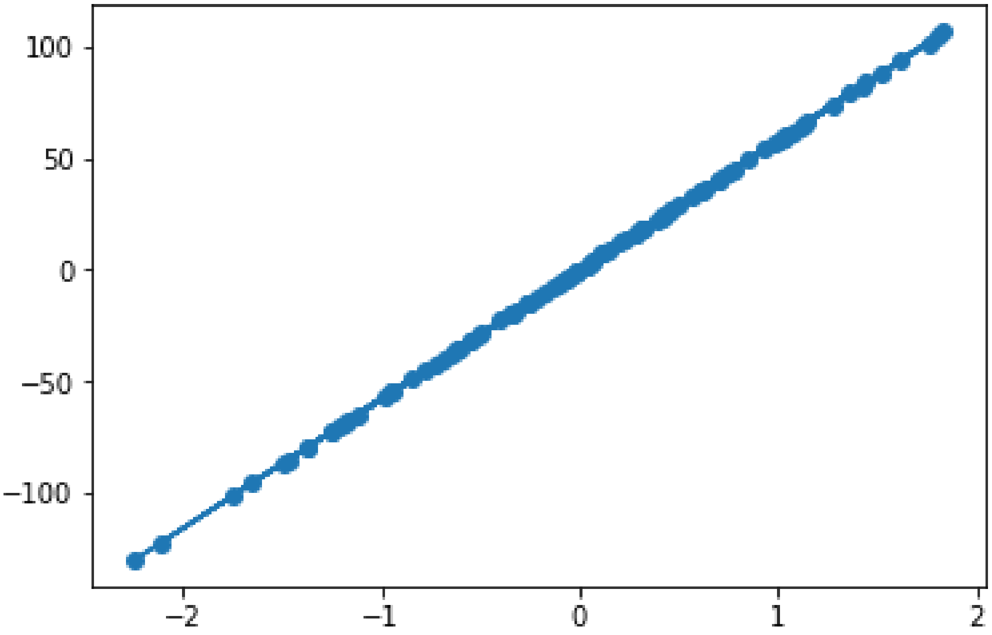
e) Change value of y and see the effect of outliers on the line.

f) Store every 10th value and find best fit curve.

```
In [56]: from sklearn.datasets import make_regression
         import matplotlib.pyplot as plt
         from scipy import stats
```

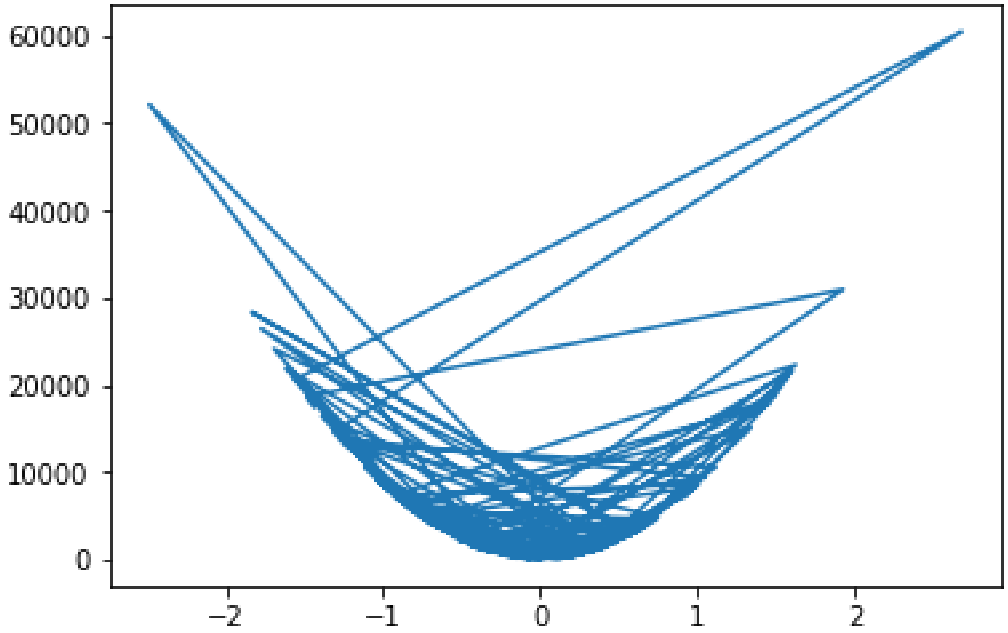
0.2 Load and plot values on a graph. Plot it as connected dots(line) and scattered dots.

```
In [57]: X, y = make_regression(n_samples=100, n_features=1, noise=.3,)
         plt.scatter(X,y)
         plt.plot(X,y)

Out[57]: [<matplotlib.lines.Line2D at 0x7f71f419cda0>]
```



0.3 Plot x vs y^2 graph

```
In [38]: import numpy as np
         y = np.asarray(y)

In [39]: y2=y*y
         pyplot.plot(X,y2)

Out[39]: [<matplotlib.lines.Line2D at 0x7f71f5318668>]
```

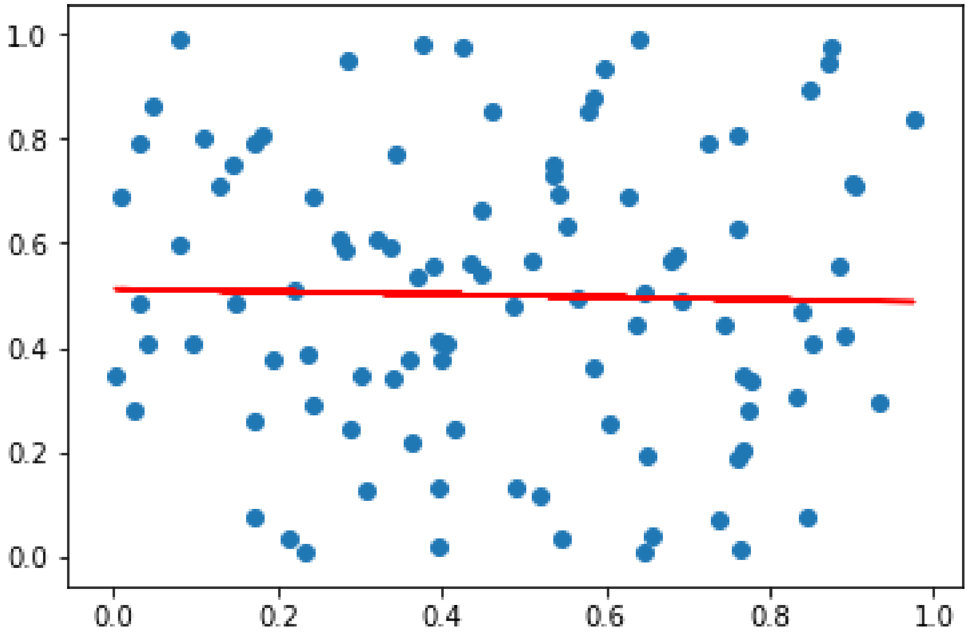


0.4 Plot best fit line using stats library.

```
In [115]: x = np.random.random(100)
          y = np.random.random(100)
          slope, intercept, r_value, p_value, std_err = stats.linregress(x, y)
```

```
In [116]: plt.scatter(x,y)
          plt.plot(x, intercept + slope*x, 'r', label='fitted line')
```

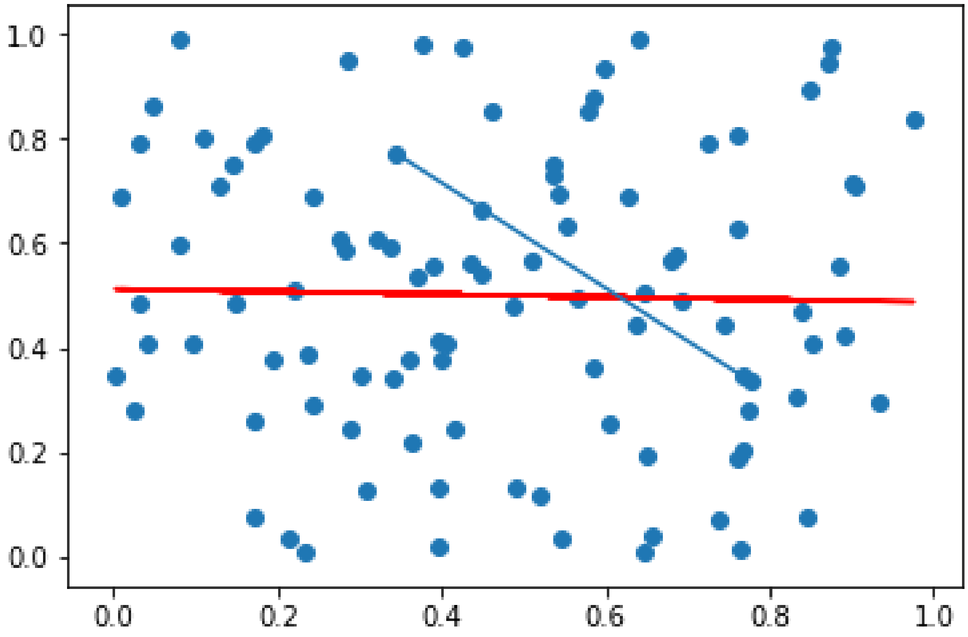
```
Out[116]: [<matplotlib.lines.Line2D at 0x7f71f2e1f0f0>]
```



0.5 Compare the best fit line with another line joining the two extremes, any two points.

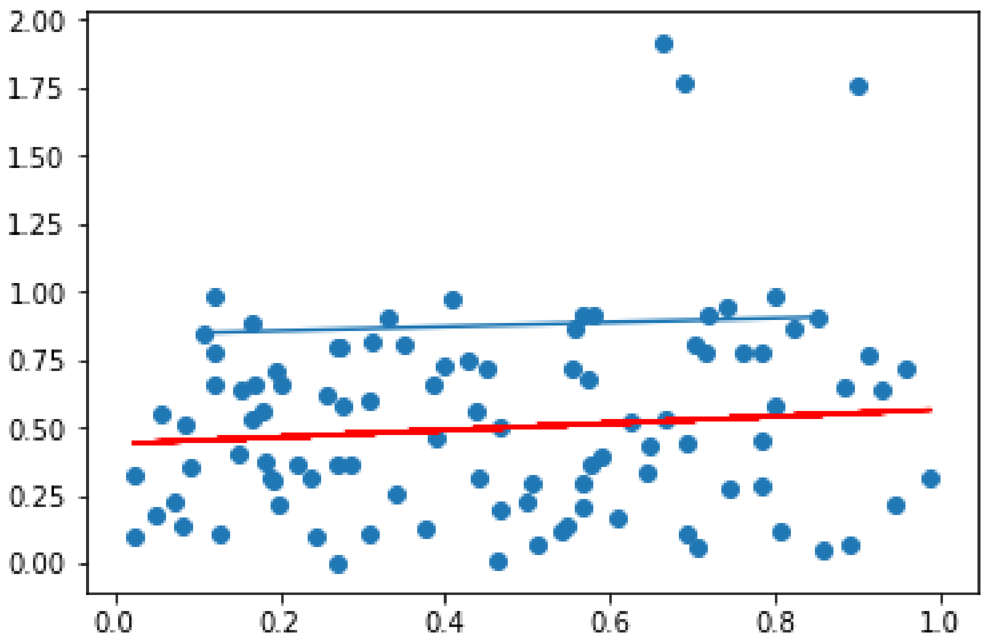
```
In [117]: plt.scatter(x,y)
          plt.plot(x, intercept + slope*x, 'r', label='fitted line')
          plt.plot([x[0],x[99]],[y[0],y[99]])
```

```
Out[117]: [<matplotlib.lines.Line2D at 0x7f71f2c799b0>]
```



0.6 Change value of y and see the effect of outliers on the line.

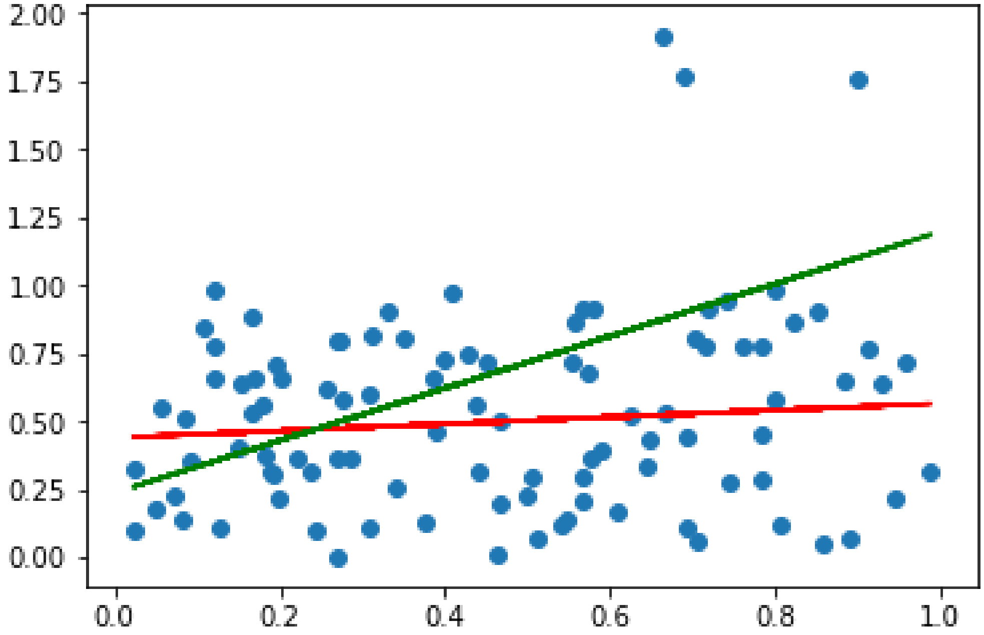
```
In [108]: y[90:95]=y[90:95]*2
          y
          plt.scatter(x,y)
          plt.plot(x, intercept + slope*x, 'r', label='fitted line')
          plt.plot([x[0],x[99]],[y[0],y[99]])

Out[108]: [<matplotlib.lines.Line2D at 0x7f71f2fa3c88>]
```



0.7 Store every 10th value and find best fit curve.

```
In [111]: x_ten=x[::10]
          y_ten=y[::10]
          slope_ten, intercept_ten, r_value_ten, p_value_ten, std_err_ten = stats.linregress(x_t
          plt.scatter(x,y)
          plt.plot(x, intercept + slope*x, 'r', label='fitted line')
          plt.plot(x, intercept_ten + slope_ten*x, 'g', label='fitted line')
```

```
Out[111]: [<matplotlib.lines.Line2D at 0x7f71f2e57e80>]
```


2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0