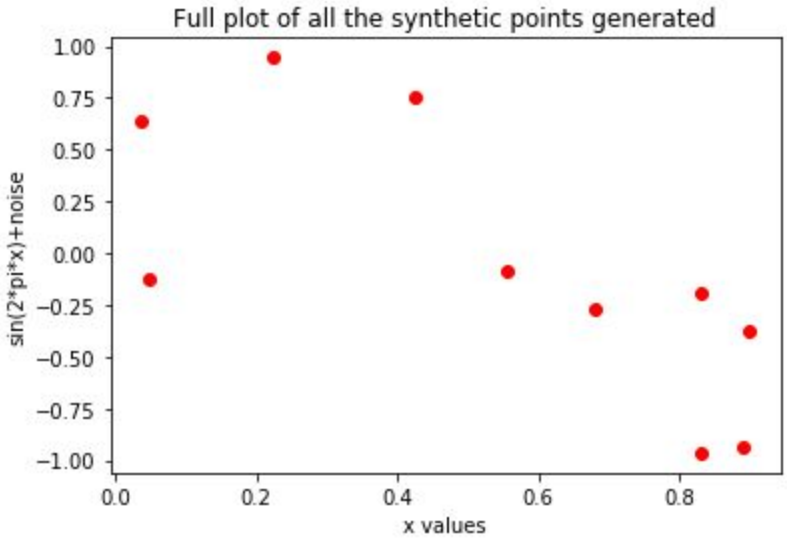
ML Assignment 1

Aakash Naik 16CS30001

Report on Linear Regression for Various Cost Functions.

Q.2 a

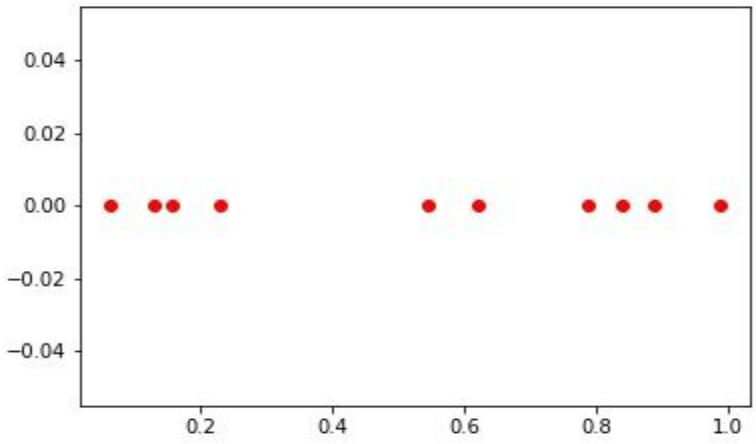
Here we have used square cost function for the analysis. Total points taken are 10
Further we have plot 9 different graphs based on different degree of x.

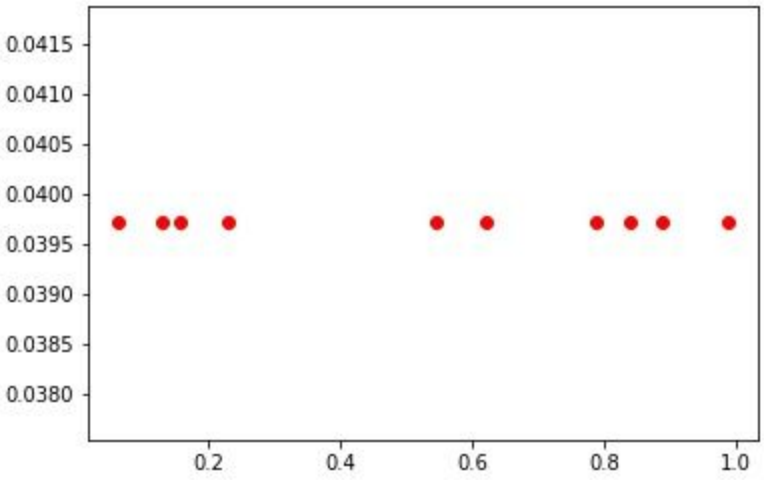


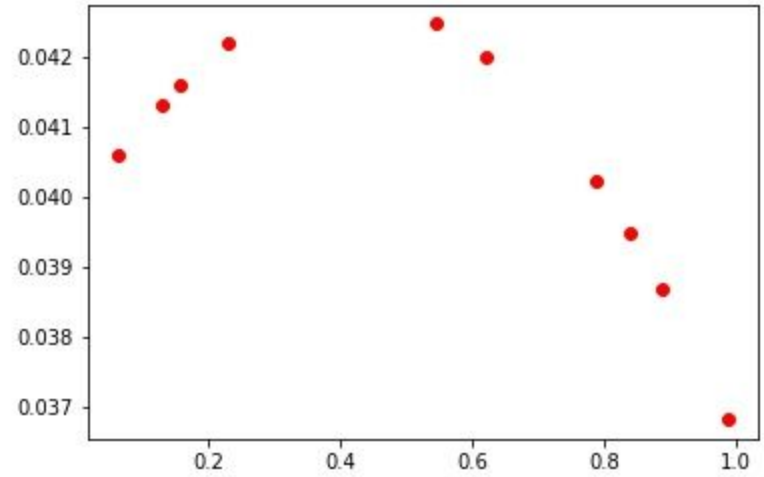
Set size = 10 data points

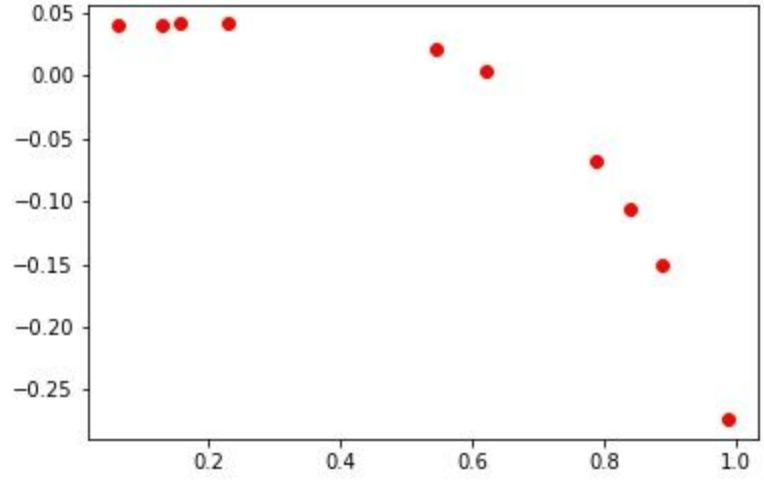
The above plot shows the randomly generated 10 data points

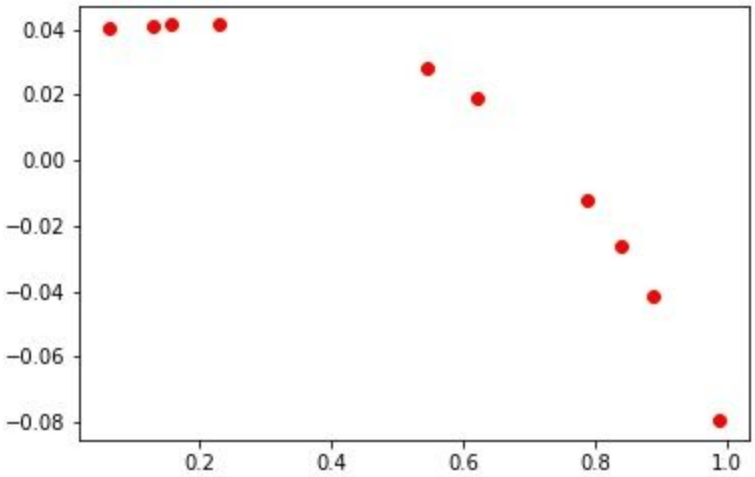
This whole set was divided into training and testing set.

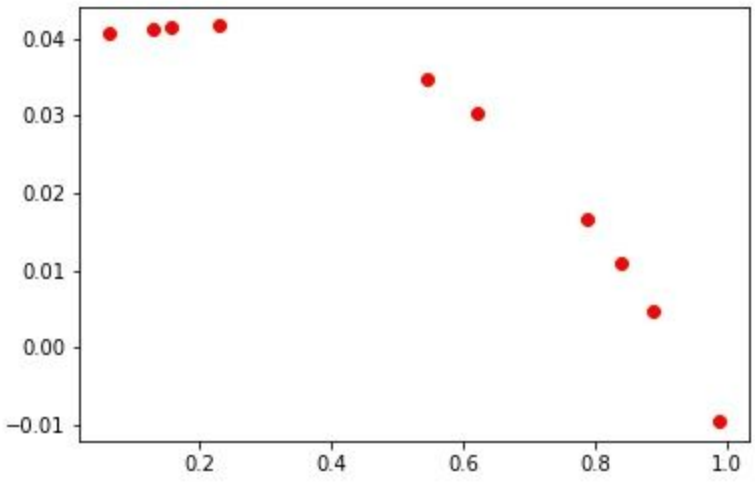

0.04
0.02
0.00
−0.02
−0.04
0.2
0.4
0.6
0.8
1.0

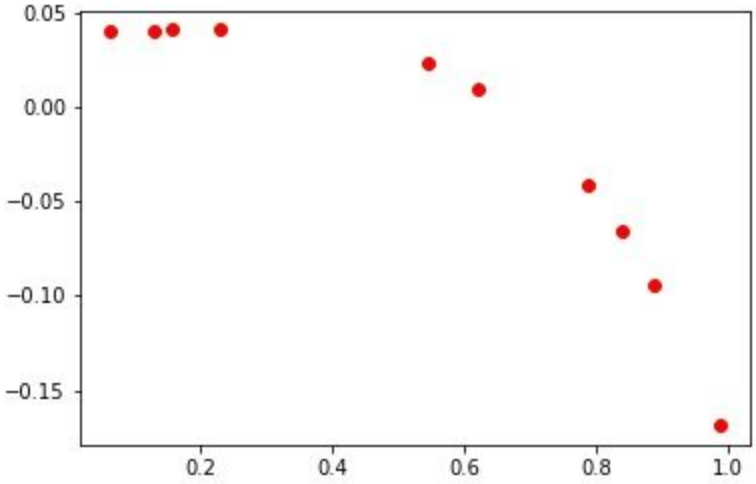

0.0415
0.0410
0.0405
0.0400
0.0395
0.0390
0.0385
0.0380
0.2
0.4
0.6
0.8
1.0

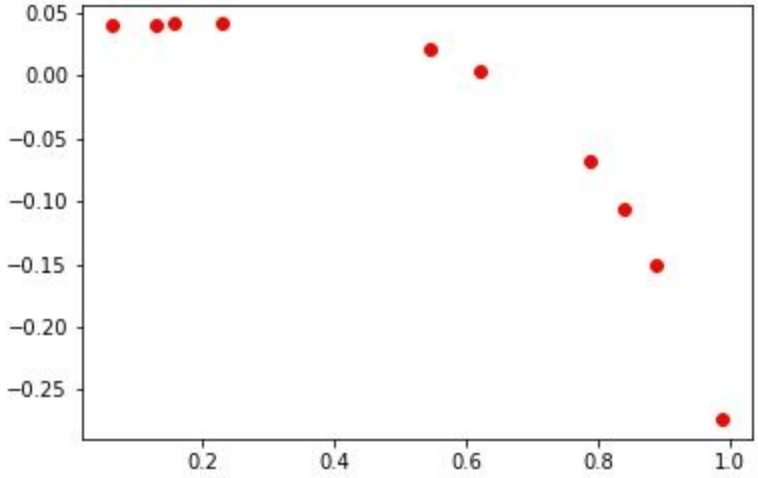

0.042
0.041
0.040
0.039
0.038
0.037
0.2
0.4
0.6
0.8
1.0

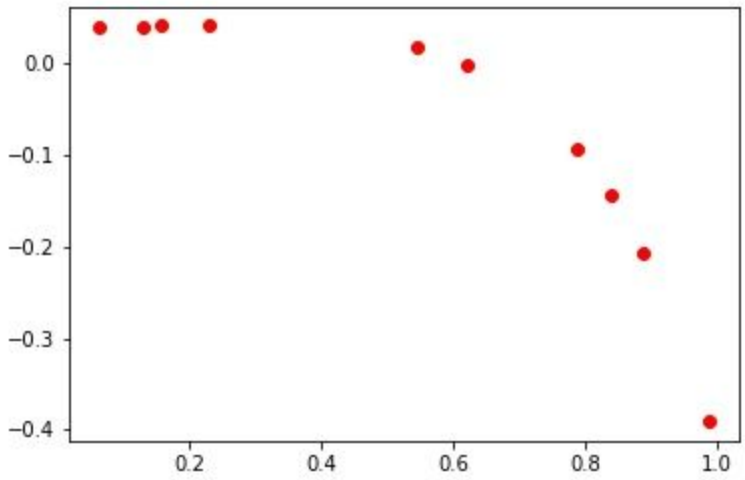

0.05
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
0.2
0.4
0.6
0.8
1.0


0.04
0.02
0.00
−0.02
−0.04
−0.06
−0.08
0.2
0.4
0.6
0.8
1.0


0.04
0.03
0.02
0.01
0.00
−0.01
0.2
0.4
0.6
0.8
1.0

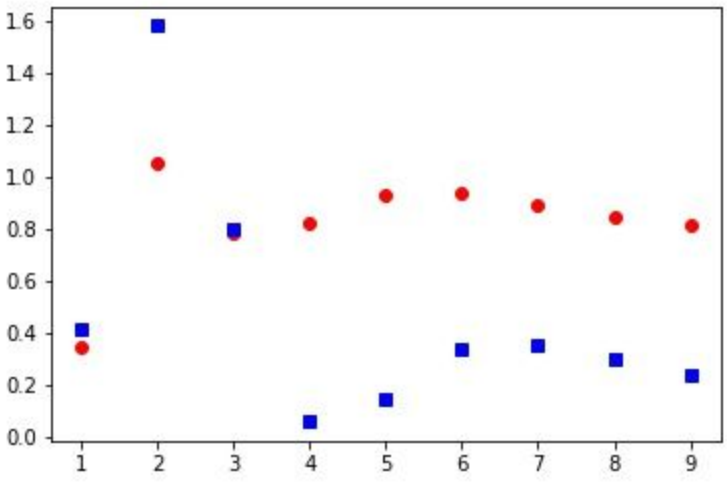

0.05
0.00
−0.05
−0.10
−0.15
0.2
0.4
0.6
0.8
1.0


0.05
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
0.2
0.4
0.6
0.8
1.0


0.0
−0.1
−0.2
−0.3
−0.4
0.2
0.4
0.6
0.8
1.0

Q.2 b)

Here sequentially we have plot error function vs degree from 1 to 9

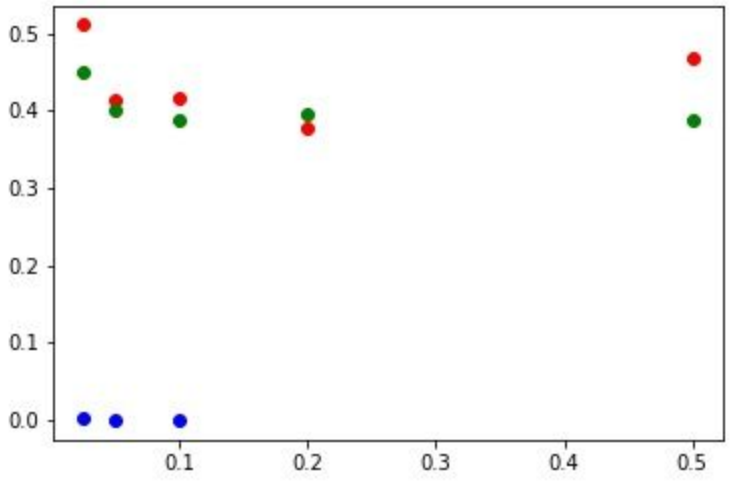


The above plot compares the error values for various degrees on 10 data points.

So, ideally we should have chosen n=3 here as the difference in error of training and test data test is minimal. Training data are marked by red whereas test by blue.

However for other data points(100,1000,10000) we have found out that n=4 works best (which can be easily generated by our code).So, n=4 works best among all possibilities.

Q.4 b



Here red one marks the squared error function ,green one for mean_average error , blue one for fourth power error. From the graph we can easily estimate that alpha=0.2 works best.

FIndings: Training Error Increases as the dataset size increases and the testing error decreases, after sufficient size of dataset, both saturate.
Also, while using the gradient descent algorithm, we must be careful while choosing the number of times the algorithm has to run for if the chosen number of iterations is very high, we will end up wasting a lot of time and if the number of iterations is very low, the result won't be satisfactory as the model would not have learnt the data properly. Same is the case while choosing the value of alpha.
Also, Note that in the code, the differential of absolute value cost function has been defined part wise.
We further observe from the graphs that for the given data set, square error cost function work well for the model.

**Please note that not all the plots have been included in the report. All the remaining plots can be generated by running the python code.
** A separate .zip file has been uploaded which contains all the plots.

Due to limitation of laptop processing power we have taken iteration=100 for each case. However results will improve as iteration increases.

FOr the graphs following naming convention is used:-
a.)'/home/aakash/document--error--'+str(error function)+'--'+ str(data_points(10,100,1000,10000)+'.jpg
b.)'/home/aakash/document--test--'+str(error function)+'--'+str(data_points)+str(i)+'.jpg'

Error function=0 ,mean squared
1,fourtherror
2, mean average