Machine Learning (60050) : Assignment 1

K Rahul Dev - 15CS10021

Part d :

Experimenting with Cost Functions

- **Features** : Linear Combination
- **Error Functions** : Mean Square Error,Mean Absolute Error,Mean Cubic Error
- **Optimization Function** : Gradient Descent

Procedure :

1.Split the Dataset into Train Set and Test Set [80:20]
2.Theta Values (Learning Parameters) are initialized randomly.
3.Took Learning Rates ranging [0.01 to 0.1]
4.No of Iterations for MSE and ABE : 1000
No of Iterations for MCE : 51

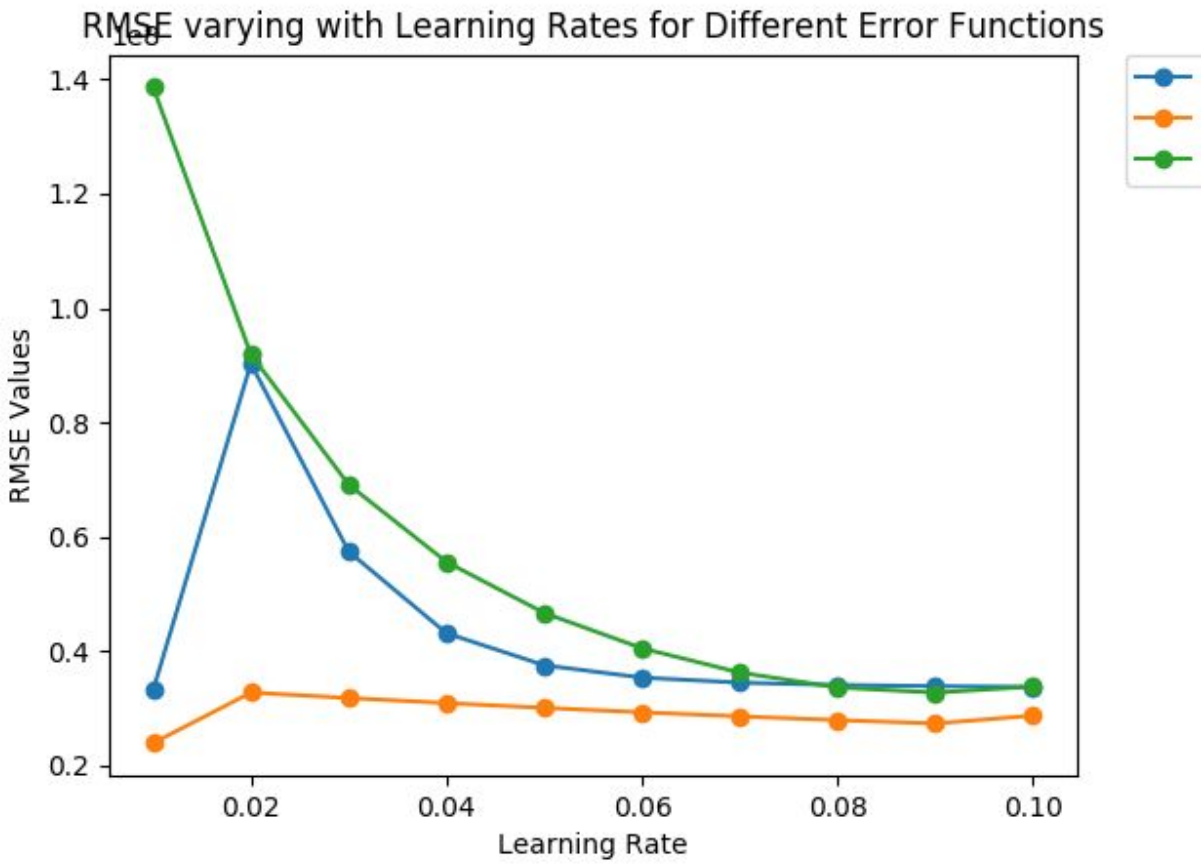
Final Learned Values (Theta Values)

Learning Rate	Mean Square Error	Mean Absolute Error	Mean Cubic Error

0.01	[0.0340093 0.26638283 0.21639018 0.41585223 -0.0698028]	[0.01000411 0.26077354 0.08343518 0.41463306 0.14510433]	[0.25125755 0.26821708 0.3127661 0.417126 -0.05065341]
0.02	[0.16676974 0.26834647 0.27998415 0.41725666 -0.07294908]	[0.02125802 0.26804364 0.2150559 0.41722878 -0.09047563]	[0.17137366 0.26782519 0.27683779 0.41682842 -0.07274878]
0.03	[0.10542089 0.26817594 0.25345406 0.41713969 -0.0882429]	[0.02017994 0.26762953 0.20471987 0.41708361 -0.07169998]	[0.13072366 0.26752568 0.25818438 0.41659562 -0.08440223]
0.04	[0.07332207 0.26803807 0.23939433 0.41704357 -0.09508035]	[0.01924529 0.26721542 0.1946576 0.41693381 -0.0535817]	[0.10498467 0.26727012 0.24612318 0.41639392 -0.09231407]
0.05	[0.05657247 0.26791727 0.23187677 0.4169583 -0.09748641]	[0.01834072 0.26680478 0.18487237 0.41678517 -0.03603714]	[0.08631911 0.26704093 0.23720484 0.41621106 -0.09859407]

0.06	[0.04780959 0.26780525 0.2277616 0.4168786 -0.09759287]	[0.01750787 0.26638719 0.17555231 0.41663422 -0.01905087]	[0.07137282 0.26682999 0.22995052 0.41604133 -0.10413121]
0.07	[0.04316693 0.2676977 0.22539995 0.41680169 -0.09651921]	[0.01670278 0.26597192 0.16660866 0.41648211 -0.0027023]	[0.05839577 0.26663301 0.22358771 0.41588171 -0.10938957]
0.08	[0.04063162 0.26759234 0.22393348 0.41672611 -0.0948482]	[0.01565478 0.26555897 0.1579933 0.41634388 0.01284687]	[0.04628987 0.26644755 0.21763171 0.41573051 -0.1146687]
0.09	[0.03916515 0.26748802 0.22291935 0.41665113 -0.09288045]	[0.01493297 0.26514833 0.14980093 0.41618136 0.02781561]	[0.0342036 0.2662722 0.2117074 0.41558674 -0.12021501]
0.10	[0.03823661 0.26738414 0.22213123 0.41657636 -0.09076961]	[0.02773344 0.26474 0.14669313 0.41604197 0.04483286]	[0.02129311 0.26610617 0.20544433 0.41544988 -0.12629019]

Plot :


RMSE varying with Learning Rates for Different Error Functions
1e8
1.4
1.2
1.0
0.8
0.6
0.4
0.2
RMSE Values
0.02
0.04
0.06
0.08
0.10
Learning Rate

NOTE :

Blue Line : MSE
Orange Line : ABE
Green Line : MCE

From the above Graph we can infer that

i) **MCE** converges for small # of Iterations,whereas **MSE** and **ABE** considerably larger # of iterations.
ii) But in the case of **MCE**,the gradients are tremendously large,so if we have a model where it has larger (predicted value - Label value),it usually overflows.
iii) In the case of **ABE**,it doesn't show fall any dip or has doesn't strive minimize the error,it justs performs linearly,independent of the errors.
iv) In the case of **MSE**,it gives considerable gradients and strives to minimize the **RMSE**,which is evident from the Graph

From the above observations,I conclude that I would prefer **Mean Square Error** for Training my data.

PS : Because of Random initialization,every execution of the trained model gives different Learned Values

The above graph is corresponding with the values submitted in part_d result file.