1)

Without regularization:

Hypothesis function - p0 + p1 * sqft + p2 * floors + p3 * bedrooms + p4 * bathrooms

Learning rate = 0.05

Parameters Learned - p0 = 533335

p1 = 15707.3

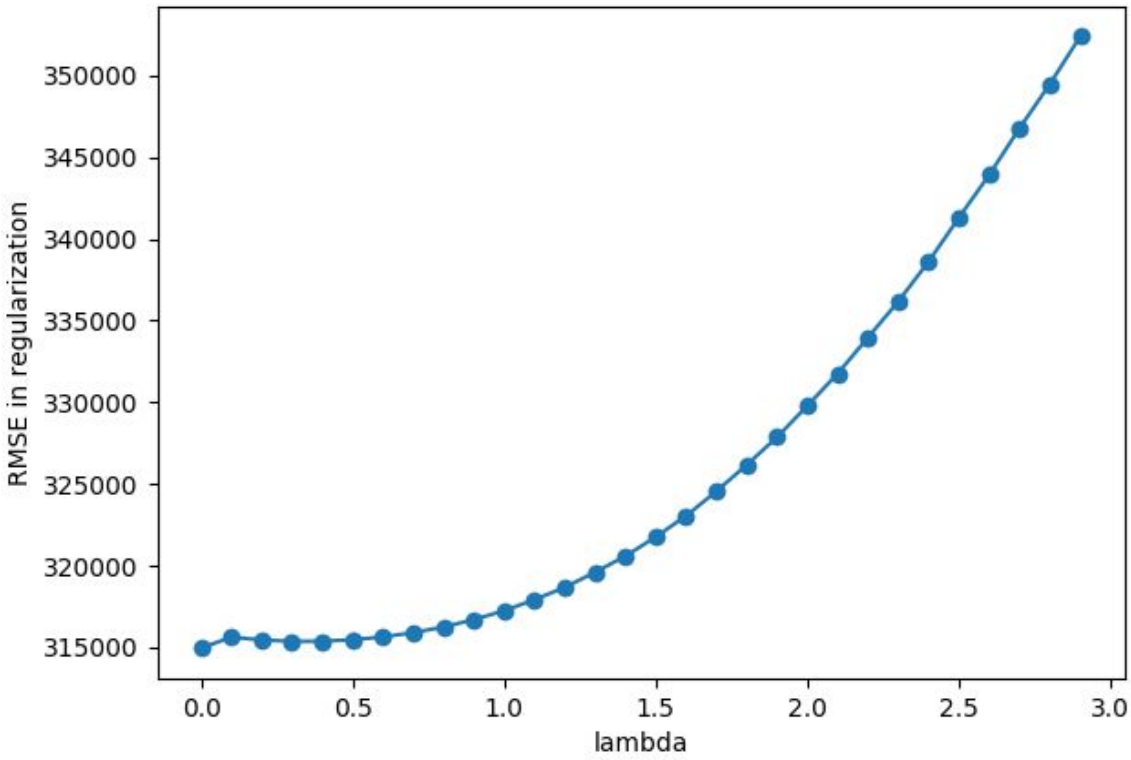
p2 = 13513.8

p3 = 17566.6

P4 = 176950

With regularization:

Parameters with reg param 0 are 533335 15707.3 13513.8 17566.6 176950

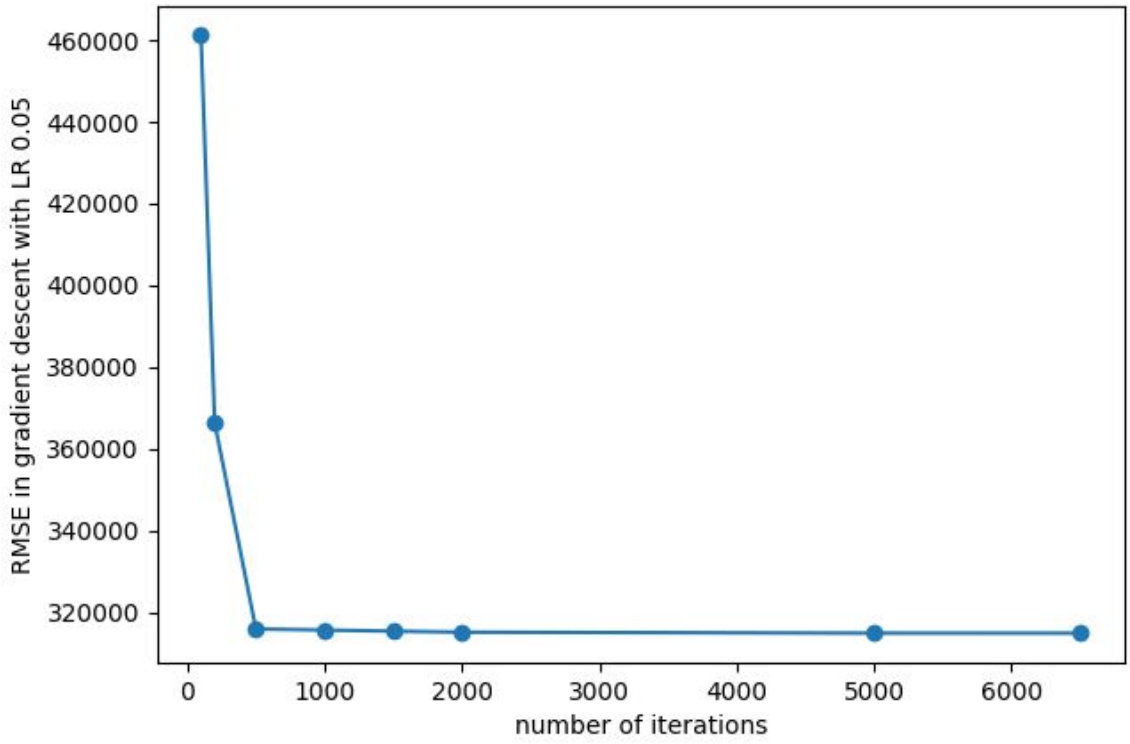


2)

Parameters in IRLS are 533336 15707.5 13511.4 17565.7 176954

Parameters in gradient descent are 533335 15707.3 13513.8 17566.6 176950

Plot of test RMSE vs number of iterations



Conclusion:

Since IRLS can be ran in a single iteration and moreover gives a better solution, IRLS can be a better fit in this problem

3) **Linear hypothesis function** :

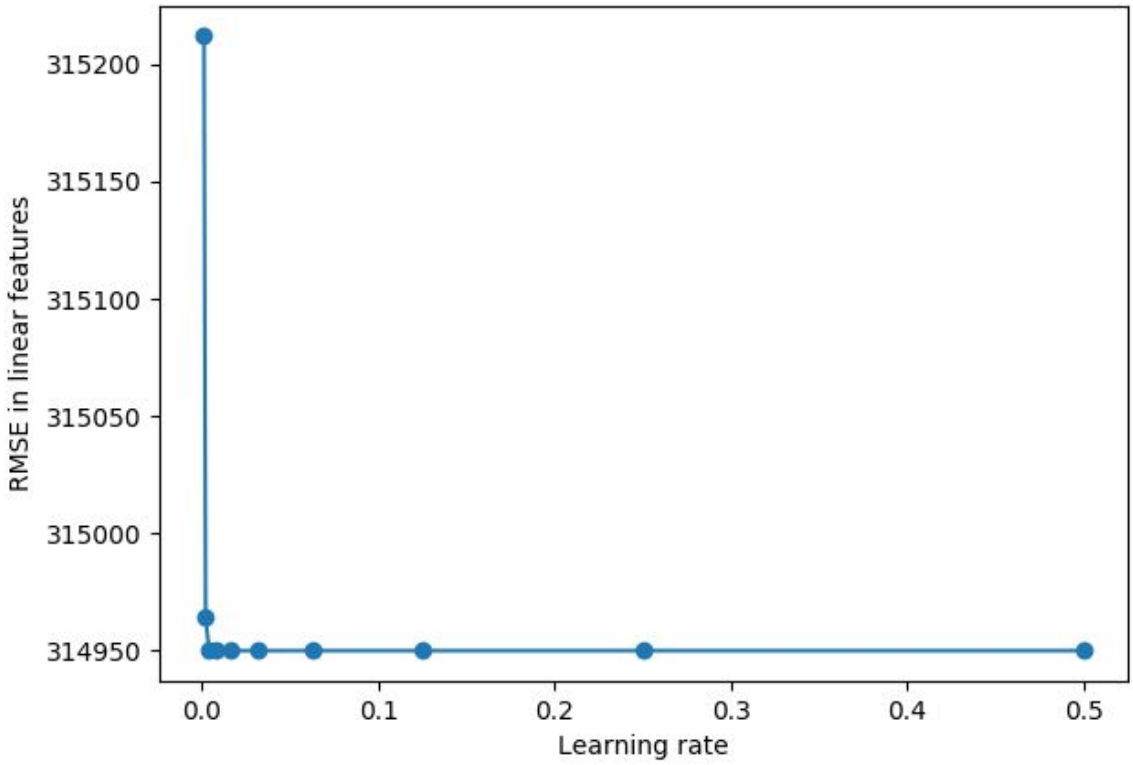
Hypothesis function - p0 + p1 * sqft + p2 * floors + p3 * bedrooms + p4 * bathrooms

Gradient Descent was ran for 10000 epochs in all the cases

Learning rate (Alpha)	Learned Parameters
0.5	533335 15707.1 13510.6 17564.9 176953
0.25	533335 15707.1 13510.6 17564.9 176953
0.125	533335 15707.1 13510.6 17564.9 176953
0.0625	533335 15707.1 13510.6 17564.9 176953
0.03125	533335 15707.1 13510.6 17564.9 176953
0.015625	533335 15707.1 13510.6 17564.9 176953

0.0078125	533335 15707.1 13510.6 17564.9 176953
0.00390625	533335 15707.1 13511.1 17565.1 176952
0.00195312	533335 15721.5 13688.7 17660 176786
0.000976562	533305 15989 16826.8 19493.9 173717

Plot of test RMSE vs Learning rate

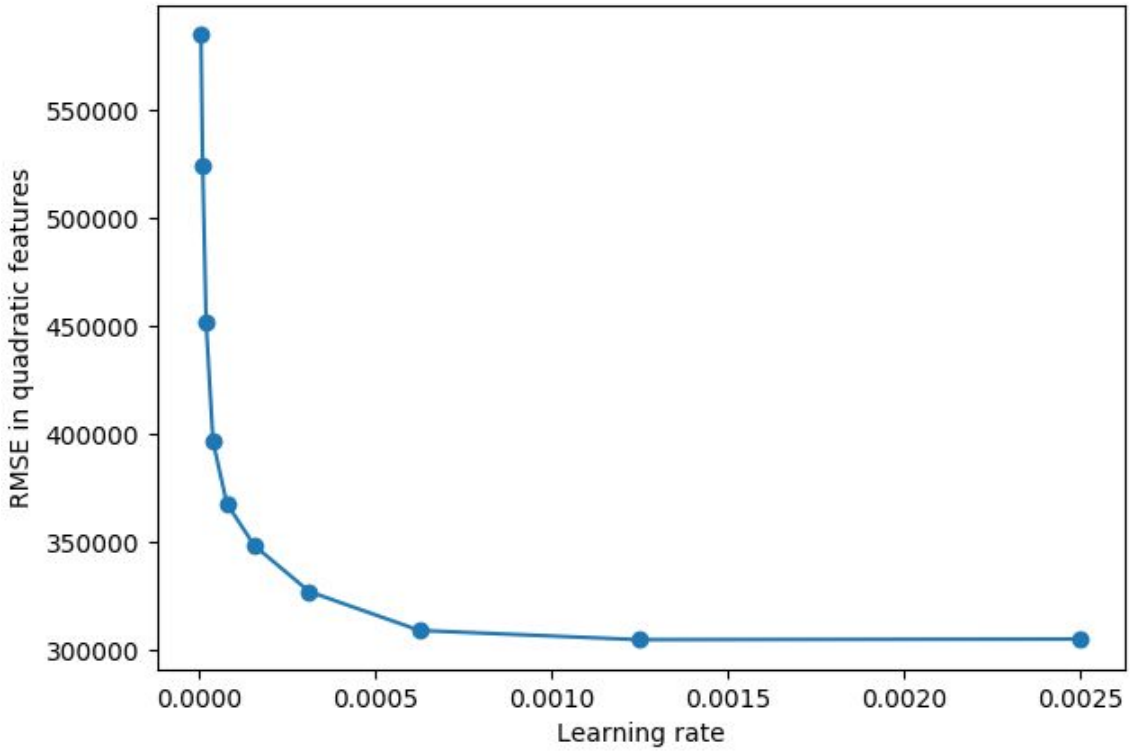


Quadratic hypothesis function :
Hypothesis function - p0 + p1 * sqft + p2 * floors + p3 * bedrooms + p4 * bathrooms + p5 * sqft^2 + p6 * floors^2 + p7 * bedrooms^2 + p8 * bathrooms^2
Gradient Descent was ran for 10000 epochs in all the cases

Learning rate	Learned Parameters
0.0025	488880 19118.8 32735.6 24480.4 138263 -521.431 -14518.1 -1011.72 54370.3
0.00125	475682 19631.7 19087.4 27642.5 138337 -533.127 3095.55 -1120.21 55009.5

0.000625	439287 18989.2 112.33 39391.6 126958 -466.191 44694.2 -1502.65 58843.8
0.0003125	376835 12548.6 -687.889 47150.4 101961 -104.217 92676.9 -1540.09 69988.3
0.00015625	292939 4943.66 10579.5 41024.9 76316 362.313 114153 -741.782 92646.1
7.8125e-05	206428 1232.92 16353.2 29870 56508.2 716.951 102223 587.762 117914
3.90625e-05	134191 301.956 14880 19980.7 40436.3 1077.64 75287.9 2733.39 119432
1.95313e-05	80483.7 176.363 10430.9 12389.4 26244.1 1514.56 48087.7 5594.5 92416.6
9.76563e-06	45000.4 154.617 6267.81 7209.67 15379.1 1905 27729.4 8051.51 58940.8
4.88281e-06	23996.9 137.194 3451.46 4041.33 8400.63 2176.22 15010.1 8723.92 33599.6

Plot of test RMSE vs Learning rate

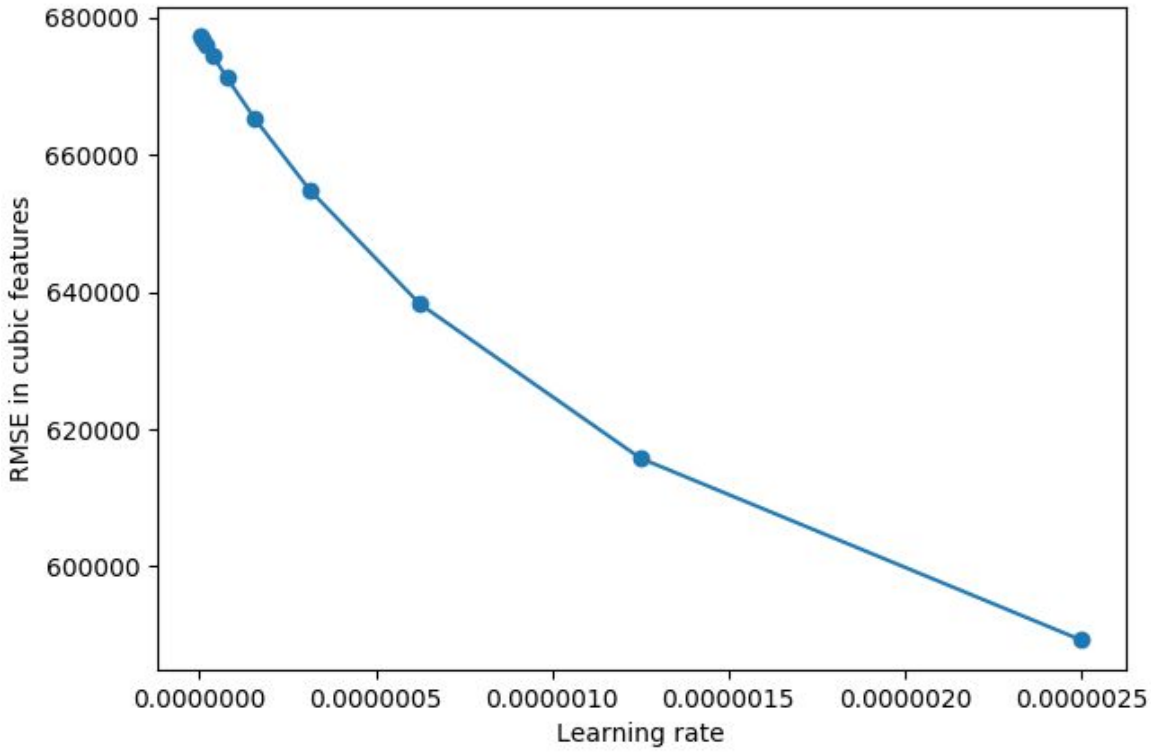


Cubic hypothesis function :

Hypothesis function - p0 + p1 * sqft + p2 * floors + p3 * bedrooms + p4 * bathrooms + p5 * sqft^2 + p6 * floors^2 + p7 * bedrooms^2 + p8 * bathrooms^2 + p9 * sqft^3 + p10 * floors^3 + p11 * bedrooms^3 + p12 * bathrooms^3
Gradient Descent was ran for 10000 epochs in all the cases

Learning rate	Parameters learnt
2.5e-06	12570.5 6.09191 1498.86 1820.1 3275.92 6163.5 7684.5 10028 14631.7 -154.452 4941.15 -313.591 16160.5
1.25e-06	6448.82 110.428 835 1032.44 1896.29 3960.11 4002.67 5371.37 8283.93 -85.726 2683.01 -151.45 12673.8
6.25e-07	3273.81 93.6691 449.341 564.168 1049.85 2276.39 2051.94 2805.12 4505.7 -33.1501 1411.26 -60.9559 8162.23
3.125e-07	1650.59 58.1119 234.851 297.837 558.223 1225.83 1040.83 1438.14 2366.23 -0.29767 725.942 -12.2673 4666.35
1.5625e-07	829.143 32.3678 119.857 153.475 288.625 637.052 524.643 729.784 1215.77 18.1306 368.428 13.1213 2499.51
7.8125e-08	415.382 17.021 60.9137 77.8906 147.339 326.099 263.66 368.671 616.769 27.8692 185.904 26.1094 1294.52
3.90625e-08	208.48 8.93281 30.6122 39.1755 73.9477 165.41 132.144 185.5 310.498 32.9024 93.5824 32.7081 658.989
1.95313e-08	104.263 5.0685 16.0388 20.1381 37.2394 84.2259 66.5824 93.9455 155.85 35.4456 47.0976 35.9871 332.523
9.76563e-09	52.2598 2.5648 8.37943 10.5059 19.5674 42.7004 33.8935 47.5191 78.3752 36.7465 24.1429 36.7447 167.102
4.88281e-09	26.1278 2.00715 4.54893 5.00541 9.75129 22.7446 17.211 24.4011 39.494 37.3718 11.7802 32.3905 83.9792

Plot of test RMSE vs Learning rate



Conclusion:

For the situation, quadratic hypothesis function suits better as it converges pretty quick and also gives better results compared to cubic and linear hypothesis functions in low epoch as well when a proper learning rate is set.

4)

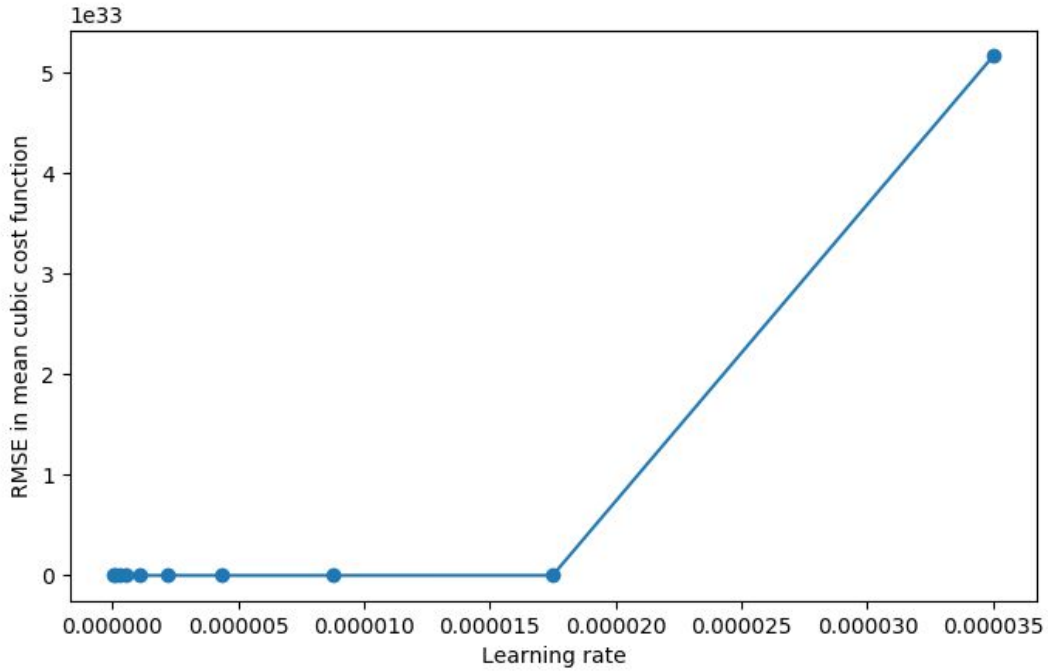
Mean cubic error function

Gradient Descent was ran for 5 epochs in all the cases

Learning rate	Learned parameters
3.5e-05	-6.91924e+32 -1.9903e+33 -9.14914e+32 -2.44216e+33 -2.36743e+33
1.75e-05	1.75e-05 is -1.92033e+28 -5.57611e+28 -2.55411e+28 -6.83633e+28 -6.61435e+28
8.75e-06	-4.81688e+23 -1.42727e+24 -6.49036e+23 -1.74737e+24 -1.68369e+24
4.375e-06	-9.68117e+18 -3.00435e+19 -1.34639e+19 -3.67266e+19 -3.50618e+19
2.1875e-06	-1.14528e+14 -4.00061e+14 -1.74564e+14 -4.91578e+14 -4.58839e+14

1.09375e-06	-2.85915e+08 -1.38459e+09 -5.95535e+08 -1.80895e+09 -1.60189e+09
5.46875e-07	5.46875e-07 is 542946 16113.7 51539.8 30509.5 281444
2.73437e-07	2.73437e-07 is 386680 22069.8 61015.1 72016.6 243822
1.36719e-07	241225 19996.5 52593.3 72945.8 185691
6.83594e-08	139784 14686.9 37232.6 53344.1 122156

Plot of test RMSE vs Learning rate

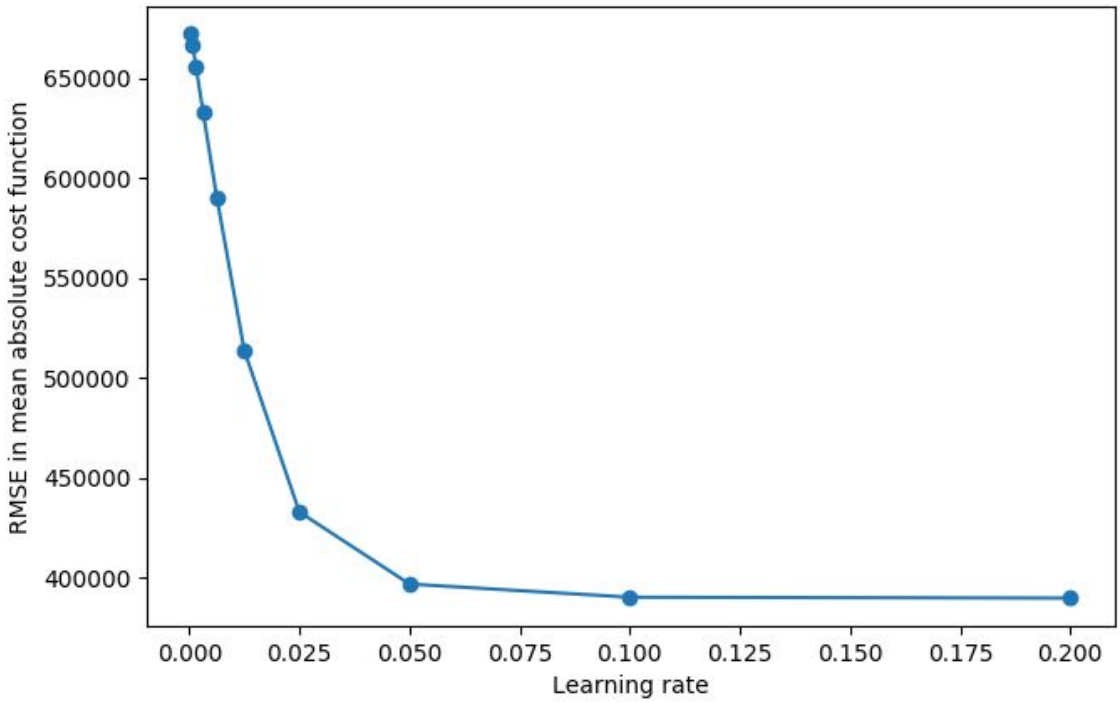


Mean Absolute Error Function
Gradient Descent was ran for 1000 epochs in all the cases

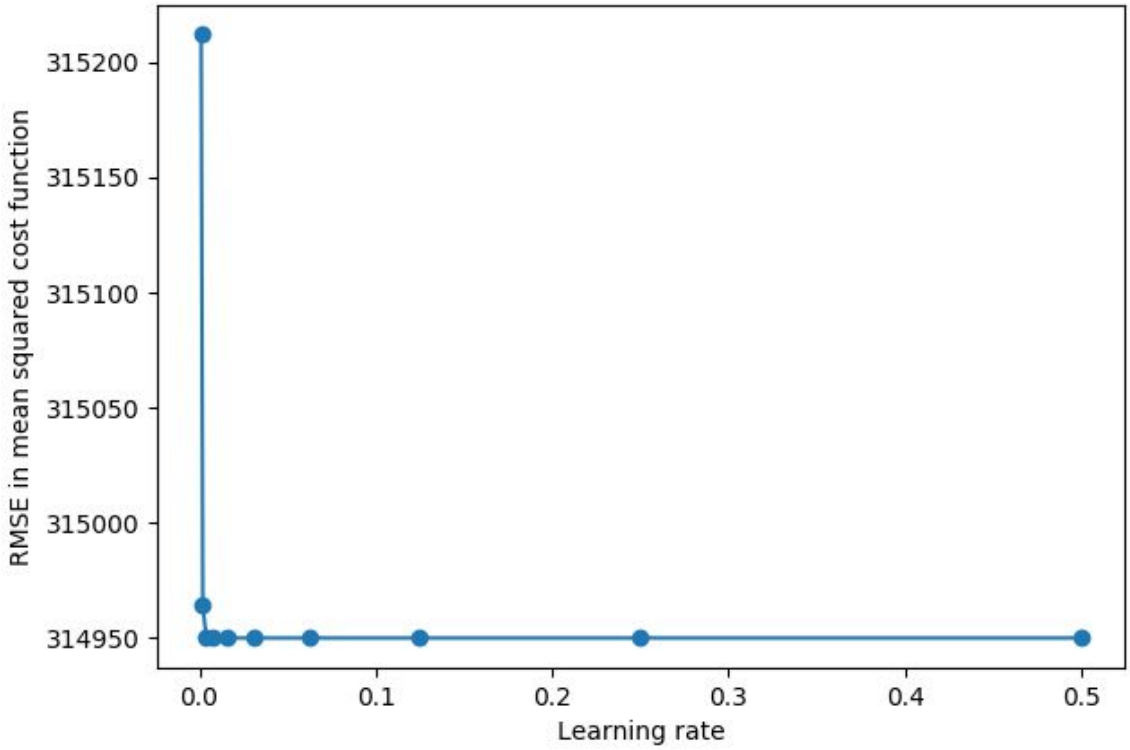
Learning rate	Parameters learned
0.2	447027 8.42633 17.6812 21.7998 36.347
0.1	445598 3.76672 7.88393 9.85608 15.9156
0.05	425908 2.37108 3.97118 4.44422 6.37948
0.025	343963 0.700184 1.63558 1.50976 2.33915
0.00125	211833 0.0549223 0.499184 1.0262 1.06889

0.00625	107903 0.28444 0.739589 0.642375 0.356606
0.003125	53981 0.165974 0.440105 0.880075 0.829201
0.0015625	26990.5 0.228968 0.893372 0.35036 0.68667
0.00078125	13496 0.58864 0.657304 0.858676 0.43956
0.000390625	6748.47 0.398437 0.814767 0.684219 0.910972

Plot of test RMSE vs Learning rate



Plot of test RMSE vs Learning rate in Mean squared error



Conclusion

For the situation mean squared error function works better as cubic function and absolute error functions diverge and moreover don't have a single points of convergence. In order to have a proper and faster convergence at a good learning rate, mean square error function is better.