Machine Learning (60050) : Assignment 1

K Rahul Dev - 15CS10021

Part b :

Experimenting with Optimization Algorithms

- **Features** : Linear Combination
- **Error Function** : Mean Square Error
- **Optimization Function** : Gradient Descent and Iterative Re-weighted Least Square
- **Learning Rate** : 0.05

Procedure :

1.Split the Dataset into Train Set and Test Set [80:20]
2.Theta Values (Learning Parameters) are initialized randomly.
3. No of Iterations : [0,50]

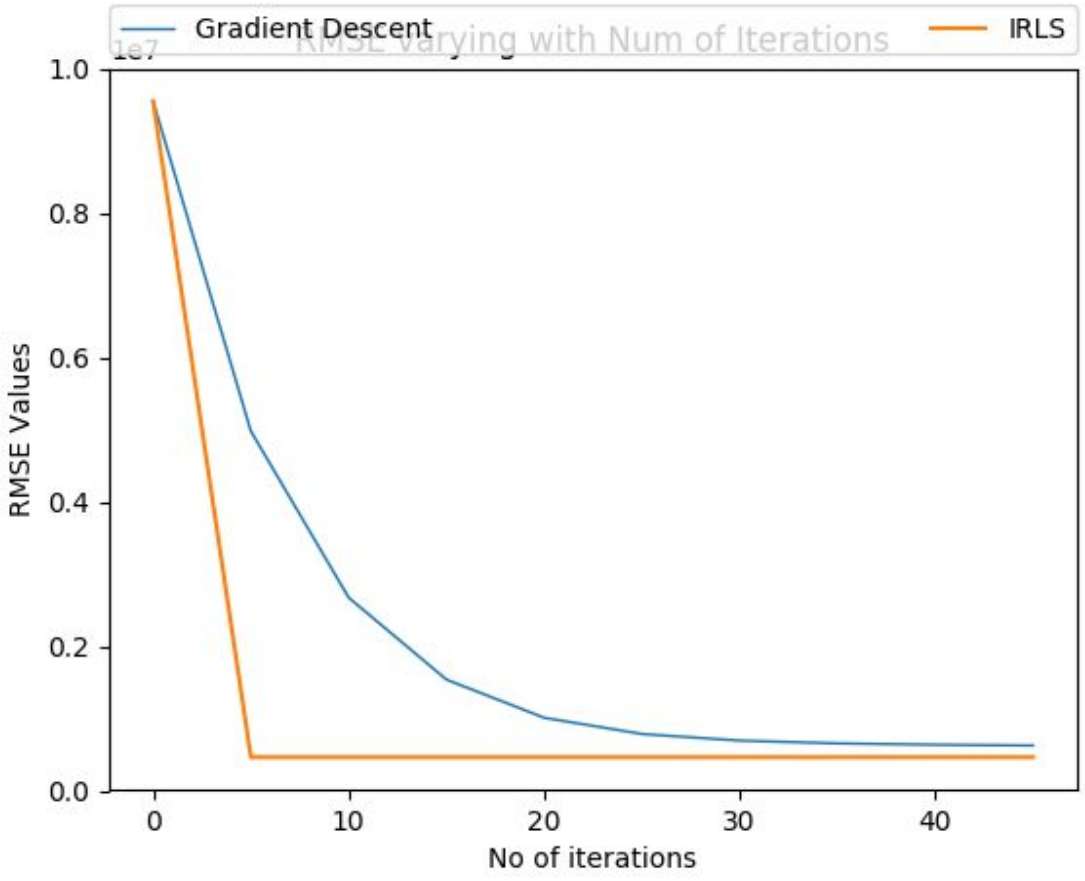
Final Learned Values (Theta Values) :

No of Iterations	Gradient Descent	ILRS
0	[0.84560145 0.34247479 0.64116082 0.99658391 0.91700321]	[0.84560145 0.34247479 0.64116082 0.99658391 0.91700321]

5	[0.35218136 0.34163681 0.42736249 0.99588954 0.77166862]	[-0.01138928 0.04809478 -0.00224824 0.13795661 0.41052934]
10	[0.10134339 0.3410716 0.31705578 0.99537007 0.69660002]	[-0.01138928 0.04809478 -0.00224824 0.13795661 0.41052934]
15	[-0.02578274 0.34064576 0.25954599 0.99494003 0.6573774]	[-0.01138928 0.04809478 -0.00224824 0.13795661 0.41052934]
20	[-0.08982267 0.34029124 0.22897788 0.99455582 0.63644592]	[-0.01138928 0.04809478 -0.00224824 0.13795661 0.41052934]
25	[-0.12169631 0.33997332 0.21216468 0.99419521 0.62485224]	[-0.01138928 0.04809478 -0.00224824 0.13795661 0.41052934]
30	[-0.13717323 0.33967431 0.20238079 0.99384685 0.6180299]	[-0.01138928 0.04809478 -0.00224824 0.13795661 0.41052934]
35	[-0.14429605 0.33938518 0.19619588 0.99350497 0.61364985]	[-0.01138928 0.04809478 -0.00224824 0.13795661 0.41052934]
40	[-0.14716657 0.33910132 0.19186028 0.99316659 0.61052417]	[-0.01138928 0.04809478 -0.00224824 0.13795661 0.41052934]
45	[-0.14787665 0.33882037 0.18848146 0.99283023 0.60804687]	[-0.01138928 0.04809478 -0.00224824 0.13795661 0.41052934]

Plot :

Test RMSE vs No of Iterations



From the above graph we can infer that :

I. **Gradient Descent** takes more no of iterations to converge or **minimize RMSE.**

Ii. **ILRS** takes very less no of iterations to converge and **minimize RMSE.**

Considering fixed Learning Rate (0.05) and less no of features and also as ILRS computes approximating L1 Norm,given L1 Norm Solutions are known to be more robust L2 Norm
,given Gradient Descent uses L2 Norm,**ILRS** is more preferable than Gradient Descent.

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_b result file.