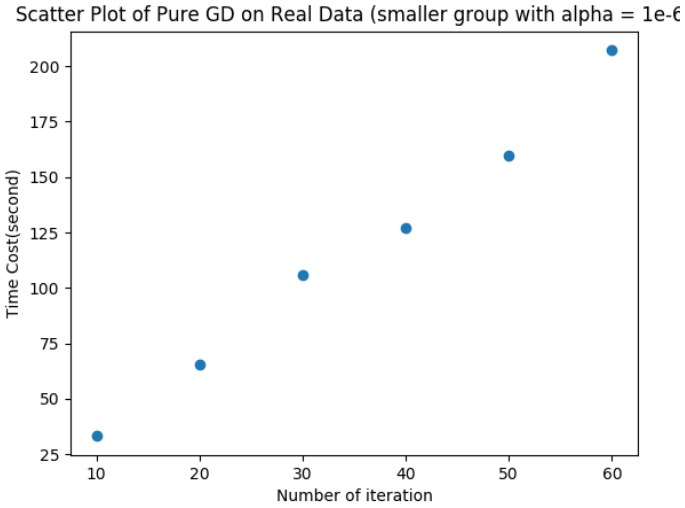
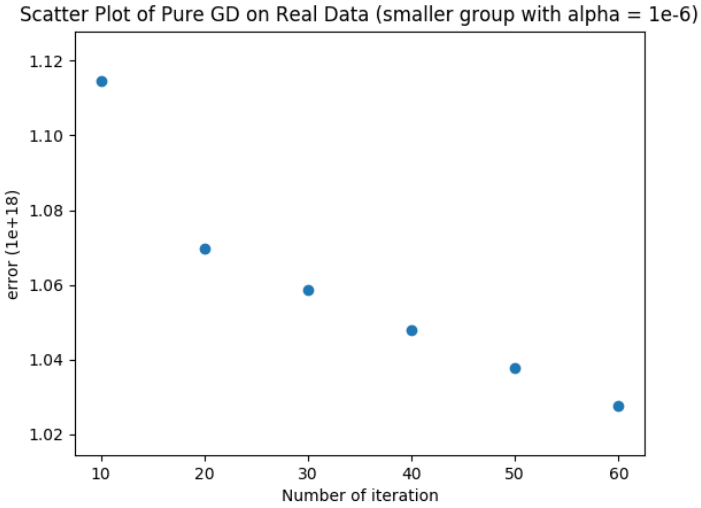
iteration	time(s)	error(e+18)	alpha=0.0000001
10	33.2033860683	1.1146544470272097	
20	65.7338712215	1.0697453341904632	
30	105.802303076	1.0586081362533221	
40	127.147273064	1.0480170826638715	
50	159.953093052	1.0376828251365121	
60	207.119021893	1.0275965858974029	





1.The time-iteration relation is linear, which performs the characteristic of GD.
2.The error should decrease in a exponential manner, but it performs like linear when iteration increases. It indicates that increment of iteration in this area can decrease the error 'significantly'.
3.When alpha = 1e-5, the GD still does not convergent. If we further decrease the alpha, we may get the proper error but the iteration could be disaster.
The error is defined as:

$$error = \sum_{i=1}^{n} = (y - \hat{y})^2 \ with \ \hat{y} = w \times X$$

4. Also the chosen of initial w is also important. In the real application the GD should be executed several times in order to get the proper result because of local minima trap.
5. The GD is somehow like local, I am wondering if there exists global algorithm to find the global minimum. As you can see, the error here is still very larger. Maybe I should use another definition of error. Do you have any idea?