Ve 485

Optimization in Machine Learning

Report of Homework 6

2020 Summer

Name:

Junjie Zhang

Student Number:

517021911128

Instructor:

Dr. Xiaolin Huang

July 8, 2020

1 Objectives

In this homework, I am supposed to implement gradient descent method for the following LS problem:

$$\min_a ||Ax - b||$$

2 Implementation

To implement, I choose L2 norm as the norm, then the object becomes:

$$\min_a ||Ax - b||_2$$

And the equivalent convex problem is:

$$\min_a ||Ax - b||_2^2$$

Then, to implement gradient descent method, I choose backtracking linear search to find the step t, where α is 0.001 and β is 0.6. And the algorithm will ends when $||\nabla ||Ax - b||_2^2||_2 < 1 \times 10^{-11}$.

Moreover, for the b with noise, ϵ $N(0, 0.001)$ here. As for the starting points, for both b and b with noise, four different starting points are randomly chosen with each element ranges from -5 to 5.

3 Results

We define the solution obtained by gradient descent method as $\hat{x}_{gd}$. For b without noise, the result can be seen in Table 1.

And for b with noise, the result can be seen in Table 2.

Starting Points	$\lvert\lvert A\hat{x}_{gd} - b\rvert\rvert_2$	$abs(\lvert\lvert(A\hat{x}_{gd} - b)\rvert\rvert_2 - \lvert\lvert(A\hat{x} - b)\rvert\rvert_2)$	$\lvert\lvert\hat{x}_{gd} - \hat{x}\rvert\rvert_2$
1	9.29261×10^{-14}	1.16508×10^{-14}	49.6388
2	9.04308×10^{-14}	9.15541×10^{-15}	51.1179
3	9.57401×10^{-14}	1.44647×10^{-14}	50.8856
4	6.19851×10^{-14}	1.92903×10^{-14}	50.5067

Table 1: Results for b without noise condition

Starting Points	$\lvert\lvert A\hat{x}_{gd} - b\rvert\rvert_2$	$abs(\lvert\lvert(A\hat{x}_{gd} - b)\rvert\rvert_2 - \lvert\lvert(A\hat{x} - b)\rvert\rvert_2)$	$\lvert\lvert\hat{x}_{gd} - \hat{x}\rvert\rvert_2$
1	8.85536×10^{-14}	1.34454×10^{-14}	50.4452
2	7.87114×10^{-14}	3.60316×10^{-15}	52.1786
3	9.92997×10^{-14}	2.41915×10^{-14}	51.0567
4	7.42045×10^{-14}	9.03688×10^{-16}	53.4887

Table 2: Results for b with noise condition

From results, it can be concluded that the gradient descent algorithm has nearly the same performance with and without noise. And the algorithm execute effectively to approach the optimal for both conditions. Moreover, we notice that although the difference of the value of object function at $\hat{x}_{gd}$ and $\hat{x}$ is small, the different between $\hat{x}_{gd}$ and $\hat{x}$ is relatively large. I think it due to the high dimension of x. Given that high dimension, the distance is actually small.