

q4

October 11, 2019

0.1 Question 4: What is the advantage of using line-search to compute the step-size at each iteration instead of using constant step-sizes equal to $1/L$? Where L is the Lipschitz constant. Is gradient descent with line-search faster than gradient descent with constant step-sizes in terms of running time? Is gradient descent with line-search faster than gradient descent with constant step-sizes in terms of running time when you add computation of the Lipschitz constant in the running time? Is gradient descent with line-search faster than gradient descent with constant step-sizes in terms of number of required iterations? Marks: 10

Question breakdown (Please note that all the experimental numbers are reported on my machine and might differ from any other machine, but the main conclusion should still hold):

What is the advantage of using line-search to compute the step-size at each iteration instead of using constant step-sizes equal to $1/L$? Where L is the Lipschitz constant. Computing the Lipschitz constant is a very expensive operation especially on large number of parameters. So, the advantage is to replace that expensive operation with a less costly one that still guarantees the decrease of the objective function at each iteration.

Is gradient descent with line-search faster than gradient descent with constant step-sizes in terms of running time? Yes. Constant-step GD takes 198 seconds in total, where GD with line search takes 15.6 seconds in total. Both reaching the same tolerance as specified by the parameter epsilon.

Is gradient descent with line-search faster than gradient descent with constant step-sizes in terms of running time when you add computation of the Lipschitz constant in the running time? Yes. Constant-step GD takes 26 seconds in total iterations time, and takes 172 seconds in computing the Lipschitz constant, where GD with line search takes 15.6 seconds in total. Both reaching the same tolerance as specified by the parameter epsilon.

Is gradient descent with line-search faster than gradient descent with constant step-sizes in terms of number of required iterations? Yes. Constant-step GD takes 357 iterations, where GD with line search takes 91 iterations. Both reaching the same tolerance as specified by the parameter epsilon.

[]: