Peer Review of CA 2 – Assignment by Group 2, Review by Group 4 (Henrik)

Intro

Apologies for the tardiness in writing this peer review! We completely missed that the peer reviews were supposed to be written continuously as the course progressed until we got the email for CA7. Anyways, we are writing all our peer reviews now, so expect to receive a bunch.

Calculation of L

I really liked that you chose to calculate L in order to find the maximum learning rate! This seems like a good approach when the problem is simple enough to find it. However, I did not understand your derivation of L! The objective function in our case is:

$$f(w) = \frac{1}{N} * \sum \log(1 + \exp\{-yw^T x\}) + \lambda ||w||_2^2$$

and the condition for Lipschitz smoothness is:

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||$$

In your solution you immediately get to the step:

$L \geq \frac{1}{4N} \sum y_i^2 x_i x_i^T + 2\lambda$ (1)

and to me it's not clear how you got there, what happened to the weights for instance? I did a quick calculation and for me it looks a lot more complicated than (1). The answer might be obvious but I'm not sure how I would go about calculating L for this problem. If you have the time, please let me know how you reached equation (1), I would appreciate it! (hhells@kth.se)

Choice of stepsize

Nice that you chose to go with a smaller value than 1/L for the stochastic methods, it makes sense especially considering the heavy fluctuations we saw in Fig. 1. Just a quick question though, why did you choose 1/(16L)? 16 seems like an arbitrary number. Was it determined empirically by trying a few different numbers?

Summary

I think the report looks good, all three solvers seem to work well and most of the choices are motivated. My only gripe is that I didn't understand how L was calculated, mostly because I have never done it myself!