Peer Review 02

Reviewer: Kai Liu

Reviewee: Charles Tanguy

Date: Oct 19th, 2016

I mainly reviewed the cpp version of SGD implementation. You implemented lazy update for zero-valued xs, which can be much faster than regular update. This is awesome. However, there are several issues:

- line 58: in denominator, it should be `M.sum() + 2.0` instead of `M.sum() * 2.0`;

- line 76, `g0squared` should be initialized as 0;

- line 99: `global_iterator` initialization should be outside the first `for` loop. Otherwise, you will get into trouble when `npass` > 1;

- line 102: `obs` should be initialized as 0. Though it is just a warning when compiling the code, R will be crashed when calling the function;

- line 108: a) log likelihood of a single data point should be multiplied by `discount`. b) the equation for log likelihood of a single data point is not correct. It should be `n_LL_avg = n_LL_avg * (1.0 - discount) + discount * (M[obs] * log(1.0 + e_psi) - Y[obs] * psi)`.

Two suggestions:

- in line 156 of your code, an `epsilon` is added in denominator. I guess this is not necessary (though not a big issue) since all cumulative log likelihoods (`Gsquare`) are initialized as 1e-03 (line 87). Therefore, `Gsquare` will not be equal to 0 anyway...

- It would be better if a comments section could be included in the function. Generally, a comments section should be below the function definition line. It can help users understand meaning and format of each argument.

Overall, you did a great job! The code is readable, and can be understand easily. After correcting all above issues, it only takes around 30s to finish the whole dataset on my laptop.

In addition, I went through your exercises03/04/05 as well. All code are clear and readable. It is really great to have comments for every several lines so that others can follow your code and even help yourself understand the code a long time later. I should do something similar :). Also, I suggest you to post visualized results of these code on GitHub as well. It helps others to understand your results and figure out if your code do things correctly. I highly recommend to use `jupyter notebook` to present your results. You can find how to install and run jupyter from here (http://jupyter.readthedocs.io/en/latest/projects/content-projects.html).