Peer Review 2

Reviewer: Yanxin Li
Reviewing: Hyunsang Son
Topic: EX01 & EX02 & EX03: Linear Regression & Generalized Linear Models & Online Learning

In this review I will provide my overall impressions of comments on both your written report and the R code you provided on your GitHub.

1. Comments on written report (file exercise1_solution.pdf & bigdata exercise2.pdf & big data exercise3.pdf)

    It is a good way to answer each question in the LaTex file Prof. James Scott has provided, which makes it readable and efficiently understandable. Also, you made the font blue to highlight the answers. But I would recommend to summarize each question and then answer it, because it helps you to better understand the problem and keep in mind the key ideas of the algorithm. Or you can add some comments and supplemental materials to the algorithms discussed in each exercise. It would be helpful for us to learn from you if you could share more on your GitHub.

    For exercise01 Part 1(B), you have showed 8 matrix decomposition or matrix factorization methods, but you did not describe them in detail. If there were more details as Mauricio had shown in class, it woud be very useful for me to learn more. For exercise 01 part 1(C), I would recommend to include plots to illustrate the performance of benchmarking results, which can be interpretable. For other exercises, you did a good job.

    What I have learned from your written report is that I will try my best to use LaTex to finish each exercise instead of using Word + MathType. The format and layout is not as beautiful as LaTex.

2. Comments on R code (file Exercise1_Rcode)

    I tried to run your code in RStudio, but there were some error messages in the output. It seems certain arguments are missing in the code. But the code stills looks logical for benchmarking the inverse method and Cholesky method.

    Here are some recommendations for the code:
    (a) Try to use crossprod instead of %*% to perform matrix multiplication;
    (b) Include more explanatory comments, like proc.time( );
    (c) Keep consistent when using "<-" and "=";
    (d) For w <- 1 / (1 + exp(-X %% beta)) in log_l and gradient functions, I think it may be

more computationally efficient if we write a separate function for w. Also, this can make your subsequent functions less clustered.

Overall, your R code looks very clean and organized. The spacing is done well so that it is easy to read. You include certain comments so that a reader can get through your code more efficiently. I also learned a new way to compare the performance of the two methods, that is, inverse method and Cholesky decomposition method.

3. Conclusion

I hope this review has been helpful. Since I am not familiar with LaTex (I usually use R Markdown and Word, although I know R Markdown is almost the same as Latex) and the format that you used for putting problems and answers together within one file is a much better way to complete the written report. Meanwhile, it is more readable when someone out of our class to learn from the GitHub. As for algorithm and implementation, basically, they look good. I hope we can discuss after class to finish each exercise with a better understanding of every topic.