Reviewer’s report

Title: Sparse Generalized Linear Model with L0 Approximation for Feature Selection and Prediction with Big Omics Data

Version: 0 Date: 11 Jun 2017

Reviewer: Jun Chen

Reviewer’s report:

In the manuscript, the authors proposed an interesting algorithm to approximate the ideal L0 penalized regression. Simulations and real data applications demonstrated a clear advantage of the proposed method for big data application. Overall, I believe that the method is novel and computationally efficient, and is a valuable contribution to the arsenal of methods for omics data. However, some details should be clarified and additional analyses performed before publication.

Major:

1. Algorithm: the initialization of eta is not clear. If eta = beta =0 as the author stated, the solution will be stuck in beta=0.

2. Algorithm: based on the reweighting algorithm, is it possible that the algorithm will be stuck in a local minimum? Will it be sensitive to a different starting point of eta?

3. Algorithm: it is not obvious to the reviewer how the sparsity is achieved using a ridge-type regression in each step. It seems that a final step of thresholding is needed to set these very small coefficients to 0's. Please state it explicitly.

4. Lemma: the lemma has not been rigorously proved in the reviewer's opinion: the proof is more of an intuitive justification of the proposed iterative algorithm.

5. Simulations: the selection of lambda should be based on the same criteria (AIC and BIC, or CV) when comparing different methods (table 2);

6. Simulations: I would suggest comparing to L1 regression using glmnet, since most applications are still using L1 penalization due to the super efficiency of glmnet.

7. Simulations: ANSF and PTM criteria are good. The evaluation will be strengthened if some measure of false positive control (e.g. empirical FDR) is also reported since one of the aims is to select the gene signature.

8. Application: the authors may consider reporting the coefficients (magnitude) for the selected features. This will also guide the design of the simulations.
9. Application: the evaluation (based on AUCs) may be too optimistic since the filtering (P < 0.01) seems to have been conducted before the model training. If the AUCs are reported, I will suggest that screening/filtering and tuning parameter selection are both performed on the training data sets and evaluation on the test data set (which has not been used for feature screening and tuning parameter selection).

10. Application: it is interesting to compare to other methods. For example, compare to the simple univariate testing procedure (coupled by FDR control), and the L1 penalization procedure to demonstrate the benefit of the proposed procedure.

Minor:

1. P3 L21-28. It is not clear how these applications are related to the proposed method. Need to rewrite.

2. P4 L25. Please explicitly state that it is the "canonical" link function. I believe the method can be applied to a general link function with some modifications.

3. "Debulking" should be defined since readers may not be familiar with it.

Level of interest
Please indicate how interesting you found the manuscript:

An article of importance in its field

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?
5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal

Do you want to get recognition for reviewing this manuscript? Add a record of this review to Publons to track and showcase your reviewing expertise across the world’s journals. Signing up is quick, easy and free!

No