Math 199 Report

Bo Zhang Xuanyu Wu
b4zhang@ucsd.edu xuw057 @ucsd.edu

In our work, we studied the sparse penalized quantile regression question where the dimension outweighs the data size. Especially, we implemented the work of Gu et al. [1] in R code and compared the result with glmnet method (Lasso and Elastic-Net Regularized Generalized Linear Models). In [1], Gu et al. introduced a Fast Alternating Direction Method of Multipliers (ADMM) to computing the sparse penalized quantile regression. We performed experiments on both proximal ADMM (pADMM) and descent ADMM (scdADMM) and compared our results with the output by the glmnet method.

We generated X from standard normal distribution with error term from normal distribution with standard deviation 1.5. True beta is $(2, 2, 2, -1.5, -1.5, -1.5, 2, 2, 2, 2, 0, 0, \ldots)^T$. In addition, we take dimension as 256 and sample size as 100. Therefore, true beta is a vector of length 256 with sparsity of 10.

We use beta estimated by glmnet with lasso penalty as our benchmark. After repeating the simulations 200 times, the average l2 norm error between true beta and estimated beta is 1.3478590; the average number of false positives is 35.63; and the average number of false negatives is 0. Therefore, glmnet tends to give high false positives while few false negatives, and has a good control of l2 norm error.

Using pADMM algorithm under the same setting, we repeat the simulations 200 times. The average l2 norm error is 15.77184; the average number of false positives is 19.455; and the average number of false negatives is 6.745. Therefore, pADMM gives high l2 norm error and many false negatives while fewer false positives than glmnet. The reason why there are so many false negatives is that for 93 out 200 times, the algorithm penalizes the beta to all zeros.

We conduct the experiment with scdADMM algorithm under the same setting by solving only one sub problem (that is, running the scdADMM for only one round), and repeat the simulations 200 times. The average l2 norm error is 0.7144043; the average number of false positives is 7; and the average number of false negatives is 0. From the result we can see that scdADMM is a better approach in terms of both l2 norm error and false positives. As scdADMM and glmnet both have 0 false negatives, their performances tie up in terms of false negatives.

References

[1] Yuwen Gu, Jun Fan, Lingchen Kong, Shiqian Ma & Hui Zou (2018) ADMM for High-Dimension Sparse Penalized Quantile Regression, *Technometrics*, 60:3, 319-331, DOI: 10.1080/00401706.2017.1345703