Summer Course Project III: Nonconvex Regulated Linear Regression

Yinyu Ye

June 2, 2018

1 Optimization over Convex Cones

We consider the following optimization problem in the non-nagative cone:

$$\begin{array}{ll} \text{Minimize} & f(\mathbf{x}) \\ \text{Subject To} & \mathbf{x} \geq 0 \text{ (or free).} \end{array} \tag{1}$$

where we have $f(\mathbf{x}) = \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|^2$ for some given data matrix $A \in R^{m \times n}$ and $\mathbf{b} \in R^m$. When $n > m$, the optimal solution may not unique so that we aim to find the sparsest optimal solution.

One approach is called LASSO [8]:

$$\begin{array}{ll} \text{Minimize} & f(\mathbf{x}) + \mu\|\mathbf{x}\|_1 \\ \text{Subject To} & \mathbf{x} \geq 0 \text{ (or free).} \end{array} \tag{2}$$

This remains a convex optimization for any given μ.

Recently, a class of "Folded Nonconvex Regularization/Penalty" functions have been introduced to replace $P(\mathbf{x}) = \|\mathbf{x}\|_1$. For example, the L_p quasi norm function with $p = 1/2$, that is, $P(\mathbf{x}) = \|\mathbf{x}\|_{1/2}^{1/2} = \sum_j |x_j|^{1/2}$, and many others in ([6, 1, 3, 5] and references therein).

2 KKT Solution Structures

Question 1: Write down the first-order KKT conditions. Note that the function is not differentiable when $x_j = 0$, but it must satisfy the classical first-order KKT conditions at $x_j \neq 0$.

Question 2: Write down the second-order KKT conditions. Note that the function is not differentiable when $x_j = 0$, but it must satisfy the classical second-order KKT conditions at $x_j \neq 0$.

Compare your results to those in [2].

3 Computational Experiments

Question 3: Implement any first-order and/or second-order algorithms on solving randomly generated test data, or other benchmark problems that you may find, and compare results among LASSO and different concave penalty functions.

4 Theoretical and Statistical Analyses

Question 4: There are some analyses on the performance and solution qualities of using the concave penalties, see [7, 5]. Read the two papers and understanding their findings. Furthermore, any improved results and new findings could be made?

5 Extensions over SDP Cone

We consider the regression problems over SDP cone:

$$\begin{array}{ll} \text{Minimize} & \frac{1}{2}\|\mathcal{A}X - \mathbf{b}\|^2 \\ \text{Subject To} & X \succeq 0, \end{array} \tag{3}$$

where

$$\mathcal{A}X = \begin{pmatrix} A_1 \bullet X \\ \dots \\ A_m \bullet X \end{pmatrix}$$

for given data matrices $A_i \in S^n$, $i = 1, ..., m$, and $\mathbf{b} \in R^m$. In many applications, we like to find a lowest rank solution matrix for the SDP regression, similar to find a sparsest solution to the linear regression. There has been a analog concave penalty $P(X)$, called matrix Schatten quasi-norm (see, e.g., [4]), was introduced and analyzed:

$$\begin{array}{ll} \text{Minimize} & \frac{1}{2}\|\mathcal{A}X - \mathbf{b}\|^2 + \mu P(X) \\ \text{Subject To} & X \succeq 0, \end{array}$$

Question 5: Do computational tests with few anchors and a few sensor points for solving Sensor Network Localization. Does the addition of the penalty help? Can you find other more effective concave penalty functions?

References

[1] R. Chartrand, Exact reconstructions of sparse signals via nonconvex minimization, IEEE Signal Process. Lett., 14 (2007), pp. 707–710.

[2] Chen, X., Xu, F., Y, Lower bound theory of non-zero entries in solutions of L2-Lp minimization. SIAM J. Sci. Comput. 32(5), 28322852 (2010).

[3] J. Fan and R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, J. Amer. Statist. Assoc., 96 (2001), pp. 1348-1360.

[4] Senshan Ji, Kam-Fung Sze, Zirui Zhou, Anthony Man-Cho So, Y, Beyond convex relaxation: A polynomial-time non-convex optimization approach to network localization. INFOCOM, 2013 Proceedings IEEE p. 2499-2507, 2013.

[5] Liu, Yao, Li, Y, .Folded concave penalized sparse linear regression: sparsity, statistical performance, and algorithmic theory for local solutions, *Math. Programming* (2017) 1-34.

`http://http://link.springer.com/article/10.1007%2Fs10107-017-1114-y`

[6] A. E. Hoerl and R. Kennard, Ridge regression: biased estimation for nonorthogonal problems, Technometrics, 12 (1970), pp. 55-67.

[7] Loh, P.-L., Wainwright, M.J. Regularized M-estimators with nonconvexity: statistical and algorithmic theory for local optima. J. Mach. Learn. Res. 16, 559616 (2015).

[8] Tibshirani, R. Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. B 58(1), 267288 (1996).