Solving Overlapping Group Lasso via Alternating Direction Method of Multipliers

Tianhong Sheng

Dec 10, 2019

Group Lasso

$$\min_{\beta} \frac{1}{2} \| Y - X\beta \|_2^2 + \lambda \sum_{i=1}^{N} \|\beta_i\|_2, \quad (1)$$

- Assume each variable joins only ONE group
- allows predefined groups of variables to be selected simultaneously
- Challenge: In micro-array gene expression data analysis, some genes may exist in more than one groups as each gene can participate in multiple pathways
- This motivates us to consider **overlapping Group Lasso**

Computing problem

In overlapping Group Lasso, optimization problem becomes challenging

$$\min_{\beta} \frac{1}{2} \| Y - X\beta \|_2^2 + \lambda \sum_{i=1}^{N} \|\beta_{G_i}\|_2, \tag{2}$$

where β_{G_i} is the sub-vector of β including the coefficients corresponding to the G_i-th group.

- Here we allow $G_i \cap G_j \neq 0$.
- For example, $G_1 = \{1, 2, 3, 4\}$ and $G_2 = \{3, 4, 5\}$
- Classic algorithms for Group Lasso does NOT work!

Solution: Alternating Direction Method of Multipliers

- **Basic Idea:** Create N new variables $z_i \in \mathbb{R}^{|G_i|}$,

$$\begin{aligned} \min_{\beta, z_1, \ldots, z_N} \quad & \frac{1}{2}\|Y - X\beta\|_2^2 + \lambda \sum_{i=1}^{N} \|z_i\|_2, \\ \text{s.t.} \quad & z_i - \tilde{\boldsymbol{\beta}}_i = 0, i = 1, \ldots, N, \end{aligned} \tag{3}$$

  where $\tilde{\boldsymbol{\beta}}_i = \beta_{G_i}$ contains the entries in β whose index is in group G_i. Let $\beta_{\mathscr{G}(i,j)}$ denote $(\tilde{\boldsymbol{\beta}}_i)_j$.
- β – global variable
- z_i, $i = 1, \ldots, N$ – local variables

ADMM algorithm for overlapping Group Lasso

Algorithm 1 ADMM algorithm for solving overlapping Group Lasso

1: Starting value for $z_i \in \mathbb{R}^{|G_i|}$ is generated from $|G_i|$ random normal variables $N(0,1)$, β and u_i are set as 0 vector in corresponding dimensions.

2: For each iteration $k = 1, 2, \ldots$ and $i = 1, \ldots, N$:

3: $z_i^{(k)} = \underset{z_i}{\operatorname{argmin}} \left(\lambda \|z_i\|_2^2 + \rho/2 \|z_i - \tilde{\boldsymbol{\beta}}_i^{(k-1)} + u_i^{(k-1)}\|_2^2\right)$,

4: $\beta^{(k)} = \underset{\beta}{\operatorname{argmin}} \left(\frac{1}{2}\|Y - X\beta\|_2^2 + \rho/2 \|z_i^{(k)} - \tilde{\boldsymbol{\beta}}_i + u_i^{(k-1)}\|_2^2\right)$,

5: $u_i^{(k)} = u_i^{(k-1)} + z_i^{(k)} - \tilde{\boldsymbol{\beta}}_i^{(k)}$,

6: Calculate the dual residual $s^{(k)} = \rho(\beta^{(k)} - \beta^{(k-1)})$ and the primal residual $r^{(k)} = (s_1^{(k)}, \ldots, s_N^{(k)})^{\mathrm{T}}$, $s_i^{(k)} = z_i^{(k)} - \tilde{\boldsymbol{\beta}}_i$;

7: Stop until convergence, stopping rule is $\left\|r^{(k)}\right\|_2 \leqslant \epsilon^{\mathrm{pri}}$ and $\left\|s^{(k)}\right\|_2 \leqslant \epsilon^{\mathrm{dual}}$ where $\epsilon^{\mathrm{pri}} = \sqrt{n}\epsilon^{\mathrm{abs}} + \epsilon^{\mathrm{rel}} \max\left\{\left\|z^{(k)}\right\|_2, \left\|\beta^{(k)}\right\|_2\right\}$ and $\epsilon^{\mathrm{dual}} = \sqrt{n}\epsilon^{\mathrm{abs}} + \epsilon^{\mathrm{rel}}\left\|u^{(k)}\right\|_2$; we set $\epsilon^{\mathrm{rel}} = 1e-3$ and $\epsilon^{\mathrm{abs}} = 1e-4$.

ADMM algorithm for overlapping Group Lasso

- z-step: block soft-thresholding

$$z_i^{(k)} = \arg\min_{z_i} (\lambda \|z_i\|_2^2 + \rho/2 \|z_i - \tilde{\beta}_i^{(k-1)} + u_i^{(k-1)}\|_2^2), \quad (4)$$

$$z_i^{(k)} = S_{\lambda/\rho}(\tilde{\beta}_i^{(k-1)} + u^{(k)}), \text{ where } \mathcal{S}_\kappa(a) = (1 - \kappa/\|a\|_2)_+ \, a.$$

- β-step: proximal operator evaluation

$$\beta^{(k)} = \arg\min_{\beta} (\frac{1}{2}\|Y - X\beta\|_2^2 + \rho/2\|z_i^{(k)} - \tilde{\beta}_i + u_i^{(k-1)}\|_2^2), \quad (5)$$

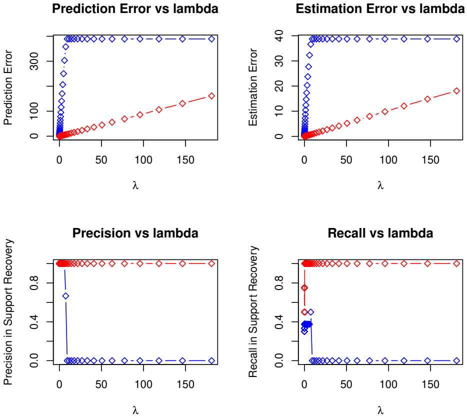
$\beta^{(k)} = (X^T X + \rho \xi I)^{-1}(X^T Y + \rho \eta)$, where
$\xi = (\xi_1, \ldots, \xi_N) \in^N$ and $\eta = (\eta_1, \ldots, \eta_N) \in^N$

$$\xi_g = \sum_{\mathscr{G}(i,j)=g} \mathbf{1} \quad \text{and} \quad \eta_g = \sum_{\mathscr{G}(i,j)=g} ((u_i^{(k)})_j + (z_i^{(k)})_j).$$

Simulation Study

- Synthetic data: $Y = X\beta_0 + \epsilon$, $\epsilon \sim N(0, \sigma^2 I)$, where $\sigma^2 = 0.04$ and each entry of design matrix $X \in \mathbb{R}^{n \times p} \sim N(0, 1)$.
- $n = 100$ and $p = 50$, $\beta_{0,i} = 10$ for $i \in \{1, 2, 3, 4, 5, 21, 22, 23, 24, 25, 36, 37, 38, 39, 40\}$
- $G_i = \{5(i-1) + j, j = 1, \ldots, 10\}$, $i = 1, \ldots, 9$
- Prediction error: $\|X\beta_0 - X\hat{\beta}\|_2$
- Estimation Error for β: $\|\beta_0 - \hat{\beta}\|_2$
- Precision of the recovery of the support: $\frac{\left|\left\{i:\hat{\beta}_i \neq 0\right\} \cap \left\{j:\beta_{0,j} \neq 0\right\}\right|}{\left\|\hat{\beta}\right\|_0}$
- Recall of the recovery of the support: $\frac{\left|\left\{i:\hat{\beta}_i \neq 0\right\} \cap \left\{j:\beta_{0,j} \neq 0\right\}\right|}{\|\beta_0\|_0}$

Simulation Results



The red line is our ADMM algorithm, the blue line is method by Zeng & Breheny (2016).

Summary

- **My contribution:** Derive the ADMM algorithm for overlapping group Lasso and design simulation study to compare its performance with state-of-art algorithm
- **Limitations of the algorithm:**
  1. For high dimensional problem, computational cost is high and storage of matrix may not be practical
  2. Algorithm suffers from singularity problem when the true support of predictors are not the complement of a union of groups (Jacob et al. 2009).

Take Home Message

- **Advantages of ADMM:** The iterative scheme of ADMM decomposes the problem into two separate minimization subproblems with only one variable to minimize for each.
- **Sparsity:** For general ℓ_1-norm problems which often appear in statistical learning, the solution of ADMM leads to a soft-thresholding operator, which introduces sparsity.
- **Limitations:** ADMM for overlapping group Lasso works for small and medium scale problems well, but it has limitations for high dimensional problems and singularity issue.