Initialization of Constrained LASSO path

안승환

2019년 9월

[1] Department of Statistics
University of Seoul

Goal

Constrained lasso problem with only equality constraints:

$$\begin{aligned} &\text{minimize} \quad && L(\boldsymbol{\beta}) + \rho||\boldsymbol{\beta}||_1 \\ &\text{subject to} \quad && \mathbf{A}\boldsymbol{\beta} = 0 \end{aligned} \tag{1}$$

(We could think of $L(\boldsymbol{\beta})$ as $\frac{1}{2}||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||_2^2$.)
Since we perform path following in the decreasing direction, an initializing value for the parameter ρ is needed.
$(\mathbf{y} \in \mathbb{R}^n, \mathbf{X} \in \mathbb{R}^{n \times p}, \boldsymbol{\beta} \in \mathbb{R}^p, \mathbf{A} \in \mathbb{R}^{m \times p})$

Goal

As $\rho \to \infty$, the solution $\boldsymbol{\beta}$ to the original problem is given by

$$\begin{aligned} &\text{minimize} && ||\boldsymbol{\beta}||_1 \\ &\text{subject to} && \mathbf{A}\boldsymbol{\beta} = 0 \end{aligned} \tag{2}$$

And obviously, the solution $\hat{\boldsymbol{\beta}}$ for the above problem is 0_p.

KKT condition

The stationarity condition of KKT conditions is as follows:

$$\nabla L(\boldsymbol{\beta}) + \rho sign(\boldsymbol{\beta}) + \mathbf{A}^T \lambda = 0_p$$

, where $\lambda \in \mathbb{R}^m$ is lagrangian multiplier and $sign(\boldsymbol{\beta})$ means the subgradient of $||\boldsymbol{\beta}||_1$. And $|sign(\boldsymbol{\beta})| \leq 1_p$, we can transform above condition as follows:

$$|\nabla L(\boldsymbol{\beta}) + \mathbf{A}^T \lambda| \leq \rho 1_p$$

Lemma

For fixed ρ, $\boldsymbol{\beta}$, let $\mathcal{E}_\rho(\boldsymbol{\beta}) = \{\lambda \in \mathbb{R}^m : |\nabla L(\boldsymbol{\beta}) + \mathbf{A}^T\lambda| \leq \rho 1_p\}$, and let $\rho_{max} = inf\{\rho \in \mathbb{R} : \mathcal{E}_\rho(\boldsymbol{\beta}) \neq \varnothing\}$. Then for $\rho < \rho_{max}$, $\boldsymbol{\beta} = 0_p$ is not solution of (1).

Corollary

The minimizer of (1) for a ρ is $\boldsymbol{\beta} = 0_p$ if and only if $\rho \geq \rho_{max}$, where ρ_{max} is the solution of (3). And also, we can get the solution λ_{max} corresponding to ρ_{max} by (3).

$$\begin{aligned} \text{minimize} \quad & \rho \\ \text{subject to} \quad & z = \mathbf{A}^T \lambda \\ & z \leq -X^T y + \rho 1 \\ & z \geq -X^T y - \rho 1 \\ & \rho \geq 0 \end{aligned} \tag{3}$$

Active Set

So, we can initialize active set $\mathcal{A}$ as follows:

$$\mathcal{A} = \{j : |\nabla L(\beta)_j + a_j^T \lambda_{max}| = \rho_{max}\}$$

where $\mathbf{A} = [a_1, \cdots a_p]$.

Because If ρ is decreased very little as $\rho < \rho_{max}$, $\hat{\beta}_j = 0$ cannot be the coefficient of solution (1) for predictor x_j. So, predictor x_j must be activated.

Uniqueness of λ

λ_{max} is the unique solution to (3) if the solution for $\mathbf{A}_{\mathcal{A}}^T \tilde{\lambda} = 0$ is only $\tilde{\lambda} = 0$. ($\Leftrightarrow \mathbf{A}_{\mathcal{A}}^T$ is full column).

Because we can formulate an equation for predictors which are set on boundaries of the stationarity condition(which are also included in active set $\mathcal{A}$) as follows:

$$|\mathbf{X}_{\mathcal{A}}^T \mathbf{y} + \mathbf{A}_{\mathcal{A}}^T(\lambda_{max} + \tilde{\lambda})| = \rho_{max} 1_{\mathcal{A}}$$