Dual Derivations of ALO for Elastic Net

Yuze Zhou

July 24, 2018

1 Dual Problem of Elastic Net

The original problem for elastic net is to solve for $\hat{\beta}$ such that:

$$\hat{\beta} = \arg\min_{\beta} \tfrac{1}{2}||y - X\beta||_2^2 + \lambda_1||\beta||_1 + \lambda_2||\beta||_2^2$$

By adding the Lagrangian, we get the formulation of L:

$$L = \tfrac{1}{2}||y - z||_2^2 + \lambda_1||\beta||_1 + \lambda_2||\beta||_2^2 + u^\tau(z - X\beta)$$

The original problem is solving the primal of the Lagrangian such that $p^* = \min_{\beta,z}\max_{u} L$ and the dual formulation $d^* = \max_{u}\min_{\beta,z} L$, to minimize over z:

$$\frac{\partial L}{\partial z} = z - y + u = 0$$
$$y = u + z$$

Since β is penalized element-wisely, we can minimize over β by minimizing over each β_i, that is, we have to minimize $\lambda_1|\beta_i| + \lambda_2\beta_i^2 - u^\tau X_i\beta$ for each dimension of β, where X_i denotes the ith column of X, therefore:

$$\min_{\beta} \lambda_1|\beta_i| + \lambda_2\beta_i^2 - u^\tau X_i\beta =$$
$$= \begin{cases} 0 & if \quad |u^\tau X_i| \le \lambda_1 \\ -\dfrac{(\lambda_1 - |u^\tau X_i|)^2}{4\lambda_2} & if \quad |u^\tau X_i| > \lambda_1 \end{cases}$$

By taking all the above to the Lagrangian, we could obtain the dual problem d^* as:

$$d^* = \min_{u} \tfrac{1}{2}||y - u||_2^2 + \sum_{j:|X_j^\tau u|>\lambda_1} \tfrac{(\lambda_1 - |u^\tau X_i|)^2}{4\lambda_2}$$

The minimizer $\hat{u}$ could also be obtained from the dual problem through a proximal approach:

$$\hat{u} = \mathbf{prox}_R(y) \quad where \quad R(u) = \sum_{j:|X_j^\tau u|>\lambda_1} \tfrac{(\lambda_1 - |u^\tau X_i|)^2}{4\lambda_2}$$

2 ALO Estimation for Elastic Net

By replacing the full data problem y with $y_\alpha = y+(y_i^{/i}-y_i)e_i$, where $y_i^{/i}$ is the true loo estimator and e_i is the ith standard vector, and let $u^{/i} = \mathbf{prox}_R(y_\alpha)$, therefore:

$$\begin{aligned}
0 &= e_i^\tau u^{/i} \\
&= e_i^\tau \mathbf{prox}_R(y_\alpha) \\
&\approx e_i^\tau(\mathbf{prox}_R(y) + J_R(y)(y_\alpha - y)) \\
&\approx \hat{u}_i + J_{ii}(y_i^{/i} - y_i)
\end{aligned}$$

Here $J_R(y)$ denotes the Jacobian matric of the proximal operator at y, thus the alo estimator $\tilde{y}_i$ is obtained as

$$\tilde{y}_i = y_i - \tfrac{\hat{u}_i}{J_{ii}}$$

The Jacobian could locally be obtained as:

$$\begin{aligned}
J_R(y) &= (I + \nabla^2 R(\mathbf{prox}_R(y)))^{-1} \\
&= (I + \nabla^2 R(\hat{u}))^{-1} \\
&= (I + \frac{1}{2\lambda_2} X_E X_E^\tau)^{-1}
\end{aligned}$$

Here $E = \{j : |X_j^\tau u| > \lambda_1\}$.