Notes on Approximate Leave-One-Out for Elastic Net

July 24, 2018

This part is to be put after section 3

1 ALO for Elastic Net, without Penalty on Intercept through Generalized LASSO

Elastic net problem (without penalty on intercept):

$$\begin{aligned}\begin{bmatrix}\hat{\boldsymbol{\beta}}_0\\ \hat{\boldsymbol{\beta}}\end{bmatrix} &= arg\,min \frac{1}{2}\|y-\boldsymbol{\beta}_0 - X\boldsymbol{\beta}\|_2^2 + \lambda_1\|\boldsymbol{\beta}\|_1 + \lambda_2\|\boldsymbol{\beta}\|_2^2\\ &= arg\,min \frac{1}{2}\begin{bmatrix}\boldsymbol{\beta}_0\\ \boldsymbol{\beta}\end{bmatrix}^T \left(\begin{bmatrix}1 & X\end{bmatrix}^\top \begin{bmatrix}1 & X\end{bmatrix} + \lambda_2\, diag(0;\mathbf{1}_p)\right)\begin{bmatrix}\boldsymbol{\beta}_0\\ \boldsymbol{\beta}\end{bmatrix} - y^T\begin{bmatrix}1 & X\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_0\\ \boldsymbol{\beta}\end{bmatrix} + \lambda_1\|\boldsymbol{\beta}\|_1\end{aligned}$$

(Here with assumption that the size of $\boldsymbol{X}$ is $n\times p$)

While the LASSO problem (also without penalty on intercept) is:

$$\begin{aligned}\begin{bmatrix}\hat{\beta}_0\\ \hat{\boldsymbol{\beta}}\end{bmatrix} &= arg\,min \frac{1}{2}\|y-\beta_0 - X\boldsymbol{\beta}\|_2^2 + \lambda_1\|\boldsymbol{\beta}\|_1\\ &= arg\,min \frac{1}{2}\begin{bmatrix}\beta_0\\ \boldsymbol{\beta}\end{bmatrix}^T \begin{bmatrix}1 & X\end{bmatrix}^\top \begin{bmatrix}1 & X\end{bmatrix}\begin{bmatrix}\beta_0\\ \boldsymbol{\beta}\end{bmatrix} - y^T\begin{bmatrix}1 & X\end{bmatrix}\begin{bmatrix}\beta_0\\ \boldsymbol{\beta}\end{bmatrix} + \lambda_1\|\boldsymbol{\beta}\|_1\end{aligned}$$

Thus we can add some "observations" to the data and let

$$y_2 = \begin{bmatrix}y\\ \mathbf{0}_p\end{bmatrix}, \qquad X_2 = \begin{bmatrix}X\\ \sqrt{\lambda_2}\boldsymbol{I}_p\end{bmatrix}$$

Then the elastic net becomes

$$\begin{aligned}\begin{bmatrix}\hat{\beta}_0\\ \hat{\boldsymbol{\beta}}\end{bmatrix} &= arg\,min \frac{1}{2}\left\|y_2 - \boldsymbol{\beta}_0\begin{bmatrix}\mathbf{1}_n\\ \mathbf{0}_p\end{bmatrix} - X_2\boldsymbol{\beta}\right\|_2^2 + \lambda_1\|\boldsymbol{\beta}\|_1\\ &= arg\,min \frac{1}{2}\left\|\begin{bmatrix}y\\ \mathbf{0}_p\end{bmatrix} - \begin{bmatrix}\mathbf{1}_n & X\\ \mathbf{0}_p & \sqrt{\lambda_2}\boldsymbol{I}_p\end{bmatrix}\begin{bmatrix}\beta_0\\ \boldsymbol{\beta}\end{bmatrix}\right\|_2^2 + \lambda_1\|\boldsymbol{\beta}\|_1\end{aligned}$$

as a special case of general LASSO.