Topic 11: Lasso

by Sai Zhang

Key points:

Disclaimer:

Lasso (Least absolute Shrinkage and Selection Operator), proposed by Tibshirani (1996), aims to minimize the **SSR** *(sum of residual squares)* subject to the $L1-$**norm** *(sum of the absolute value)* of the coefficients being less than a constant.

11.1 Set up

For data $\left(\mathbf{x}_i, y_i\right)_{i=1}^{n}$, where

- y_i is the outcome for individual i
- $\mathbf{x}_i = \left(x_{i1}, \cdots, x_{ip}\right)'$ is the $p \times 1$ vector of predictors
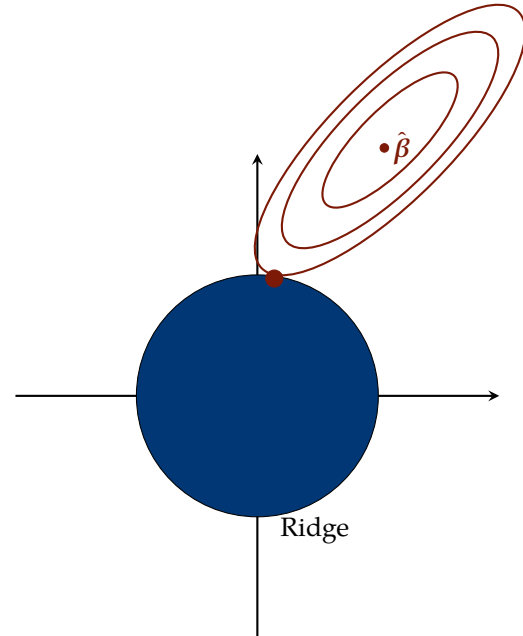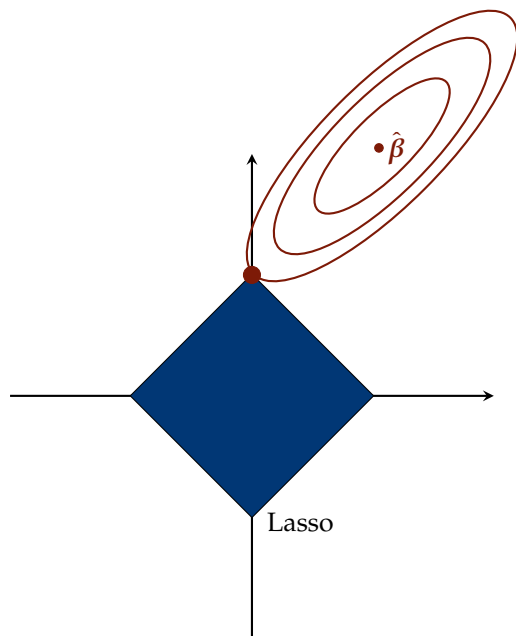
Then the Lasso estimator $\left(\hat{\alpha}, \hat{\boldsymbol{\beta}}\right)$ is defined as

$$\left(\hat{\alpha}, \hat{\boldsymbol{\beta}}\right) = \arg\min_{\alpha, \boldsymbol{\beta}} \left\{ \sum_{i=1}^{n} \left(y_i - \alpha - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \right\} \qquad \text{s.t.} \ \sum_{j=1}^{p} |\beta_j| \leq t$$

for the $n \times 1$ response vector $\mathbf{y} = \left(y_1, \cdots, y_n\right)'$, the $n \times p$ design matrix $\mathbf{X} = \left(\mathbf{x}_1, \cdots, \mathbf{x}_n\right)'$ where $\mathbf{x}_i = \left(x_{i1}, cdots, x_{ip}\right)'$ is a $p \times 1$ vector.

11.2 Penalized Least Square Estimation

Lasso is one special class of Penalized Least Square (PLS) Estimation. For the linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, if $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, we have PLS as

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \sum_{j=1}^{p} p_\lambda \left(|\beta_j| \right) \right\}$$

References

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.