

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

- y_i is the outcome for individual i
- $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ is the $p \times 1$ vector of predictors

Then the Lasso estimator $(\hat{\alpha}, \hat{\beta})$ is defined as

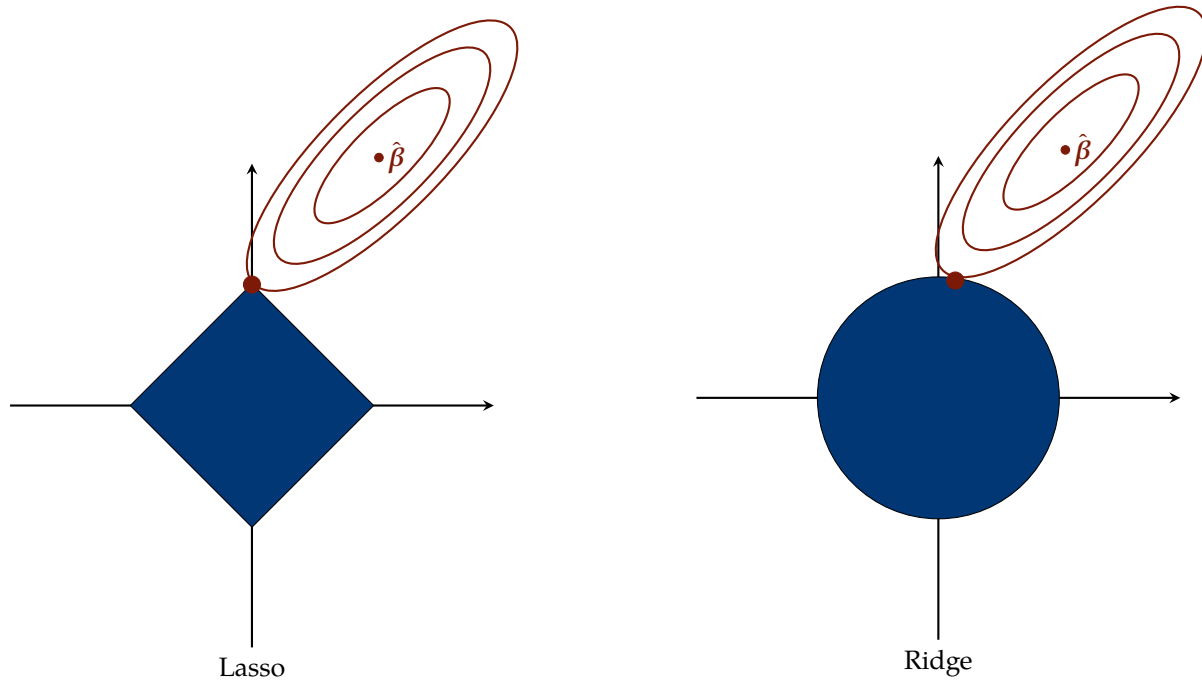
$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{\alpha, \beta} \left\{ \sum_{i=1}^n \left(y_i - \alpha - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \right\} \quad \text{s.t.} \quad \sum_{j=1}^p |\beta_j| \leq t$$

for the $n \times 1$ response vector $\mathbf{y} = (y_1, \dots, y_n)'$, the $n \times p$ design matrix $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$ where $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ 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 $\mathbf{y} = \mathbf{X}\beta + \epsilon$, if $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, we have PLS as

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



References

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