Econometrics **February 15, 2023**

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
- $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})'$ is the $p \times 1$ vector of predictors

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

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

for the $n \times 1$ response vector $\mathbf{y} = (y_1, \cdots, y_n)'$, the $n \times p$ design matrix $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)'$ where $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})'$ is a $p \times 1$ vector. Here $\hat{\alpha} = \overline{y}$, w.l.o.g., let $\overline{y} = 0$ and omit α for simplicity.

In matrix form, we have

- constrained form:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 \right\} \qquad \text{s.t. } \|\boldsymbol{\beta}\|_1 \leq \lambda$$

- unconstrained form:

$$\hat{\boldsymbol{\beta}}(\lambda) = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \right\}$$

where the regularization parameter $\lambda \geq 0$:

- $\lambda \to \infty$: $\hat{\boldsymbol{\beta}}_{lasso} \to \hat{\boldsymbol{\beta}}_{OLS}$
- $\lambda = 0$: $\hat{\boldsymbol{\beta}}_{lasso} = \mathbf{0}$

11.2 Solving Lasso

Lasso is essentially a quadratic optimization problem. Hence, the solution is given by taking the derivative (of the unconstrainted question) and set it equal to 0

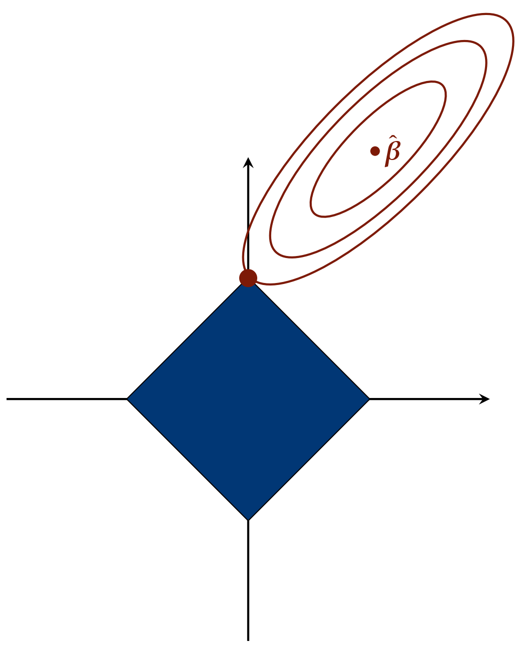
$$\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}}\left(\frac{1}{2n}\right)$$

11.3 Penalized Least Square Estimation

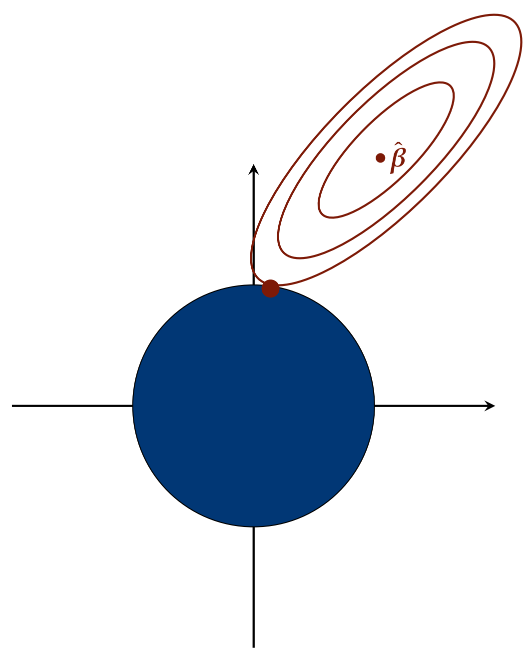
Lasso is one special class of Penalized Least Square (PLS) Estimation. For the linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, if $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_n)$, we have PLS as

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

where $p_\lambda(\cdot)$ is a penalty function indexed by the regularization parameter $\lambda \geq 0$



Lasso



Ridge

References

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