Chapter 2 The Lasso for Linear Models

2.2 LS

LASSO coef bias to 0

LS subset coef debias away from 0 relaxed Lasso

2.3 CV

↓

fit model for $\{t_i\}_{i=1}^{T}$

Test = K ER_1^k ... ER_T^k

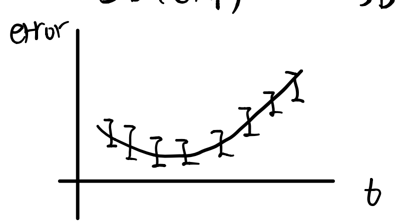
⋮

Test = 1 ER_1^1 ... ER_T^1

↓ ↓

$\overline{ER}_1$ ... $\overline{ER}_T$

$SD(ER_1)$ $SD(ER_1)$



2.4 Computation of LASSO

QP problem
$$\min_{\beta} \left\{ \frac{1}{2N} \| y - X\beta \|_2^2 \right\}$$
$$\text{s.t.} \quad \|\beta\|_1 \leq t$$

Lagrangian
$$\min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2N} \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$
$$\min_{\beta} \left\{ \frac{1}{2N} \| y - X\beta \|_2^2 + \lambda \|\beta\|_1 \right\}$$

where $\frac{1}{N} \Sigma_i y_i = 0$ $\quad \frac{1}{N} \Sigma_i x_{ij} = 0$ $\quad \frac{1}{N} \Sigma_i x_{ij}^2 = 1$

EX 2.2 Derivation for LASSO by inspection

Since X has been standardized, the $\hat{\beta}_{LS} = (X^T X)^{-1} X^T y = X^T y$

Expanding the first ~~term~~ of the Lagrangian form,

$$\frac{1}{2N} \left[(y - X\beta)^T (y - X\beta) \right]$$
$$= \frac{1}{2N} \left[y^T y - (X\beta)^T y - y^T X\beta + (X\beta)^T (X\beta) \right]$$
$$= \frac{1}{2N} \left[y^T y - 2 (X\beta)^T y + \beta^T X^T X \beta \right]$$
$$= \frac{1}{N} \left[\underbrace{\frac{1}{2} y^T y}_{\text{no } \beta \text{ here}} - y^T X\beta + \frac{1}{2} \beta^T \beta \right]$$

The problem changes to

$$\min_{\beta} \quad \frac{1}{N}\left[-y^T X\beta + \frac{1}{2}\beta^T\beta\right] + \lambda\|\beta\|_1 \qquad \|\beta\|_1 = \sum_{i=1}^{N}|\beta_i|$$

$$\min_{\beta} \quad \frac{1}{N}\left[-\hat{\beta}_{LS}^T\beta + \frac{1}{2}\beta^T\beta\right] + \lambda\|\beta\|_1$$

$$\min_{\beta} \quad \frac{1}{N}\sum_{i=1}^{N}\left(-\hat{\beta}_{LS_i}\beta_i + \frac{1}{2}\beta_i^2 + N\lambda|\beta_i|\right)$$

So, the problem can be solved as individual problems indexed by i.

For a certain i, $\min L_i = -\hat{\beta}_{LS_i}\beta_i + \frac{1}{2}\beta_i^2 + N\lambda|\beta_i|$

If $\hat{\beta}_{LS_i} > 0$, we must have $\beta_i \geq 0$

< 0 $\qquad \leq 0$

Case 1, If $\hat{\beta}_{LS_i} > 0$, since $\beta_i \geq 0$,

$$L_i = -\hat{\beta}_{LS_i}\beta_i + \frac{1}{2}\beta_i^2 + N\lambda\beta_i$$

$$\frac{\partial L_i}{\partial \beta_i} = -\hat{\beta}_{LS_i} + \beta_i + \lambda N = 0$$

$$\beta_i = \hat{\beta}_{LS_i} - \lambda N$$

with the assumption,

$$\beta_i = (\hat{\beta}_{LS_i} - \lambda N)_+ = \mathrm{sgn}(\hat{\beta}_{LS_i})(|\hat{\beta}_{LS_i}| - \lambda N)$$

Case 2. If $\hat{\beta}_{LS_i} < 0$, since $\beta_i \leq 0$

$$L_i = -\hat{\beta}_{LS_i}\beta_i + \frac{1}{2}\beta_i^2 - N\lambda\beta_i$$

$$\beta_i = (\hat{\beta}_{LS_i} + \lambda)_- = \operatorname{sgn}(\hat{\beta}_{LS_i})(|\hat{\beta}_{LS_i}| - \lambda)_+$$

To combine them together,

$$\hat{\beta} = \begin{cases} \frac{1}{N}\hat{\beta}_{LS} - \lambda & \text{if } \frac{1}{N}\hat{\beta}_{LS} > \lambda \\ 0 & \text{if } \frac{1}{N}|\hat{\beta}_{LS}| \leq \lambda \\ \frac{1}{N}\hat{\beta}_{LS} + \lambda & \text{if } \frac{1}{N}\hat{\beta}_{LS} < -\lambda \end{cases}$$

$$\hat{\beta} = S_\lambda\left(\frac{1}{N}\hat{\beta}_{LS}\right)$$

$$S_\lambda(x) = \operatorname{sign}(x)(|x| - \lambda)_+$$

By KKT conditions