Lasso 问题的求解

Yimeng Ren

1 问题重述

To find β:

$$\min_{\beta}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{P}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j|\right\}$$

when $\beta_0=0$, it is equal to:

$$\min_{\beta}\left\{\frac{1}{n}\sum_{j=1}^{n}\left(y_i-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j|\right\} \tag{1}$$

After standardization:

$$\frac{1}{n}\sum_{i}y_i=0,\quad \frac{1}{n}\sum_{i}x_{ij}=0,\quad \frac{1}{n}\sum_{i}x_{ij}^2=1$$

2 单变量

$$\begin{aligned}(1)\Rightarrow &\min_{\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}(y_i-z_i\beta)^2+\lambda|\beta|\right\}\\ \Rightarrow &\min_{\beta}\left\{\frac{1}{2}\beta^2-\frac{1}{n}\langle z,y\rangle\beta+\lambda|\beta|+const\right\}\end{aligned}$$

$$\hat{\beta}=\begin{cases}\frac{1}{n}\langle z,y\rangle-\lambda, & \text{if } \frac{1}{n}\langle z,y\rangle>\lambda\\ 0 & \text{if } \frac{1}{n}|\langle z,y\rangle|\le\lambda\\ \frac{1}{n}\langle z,y\rangle+\lambda, & \text{if } \frac{1}{n}\langle z,y\rangle<-\lambda\end{cases}$$

define **soft threshold** as:

$$\hat{\beta} = S_\lambda \left(\frac{1}{n} \langle z, y \rangle \right)$$

where

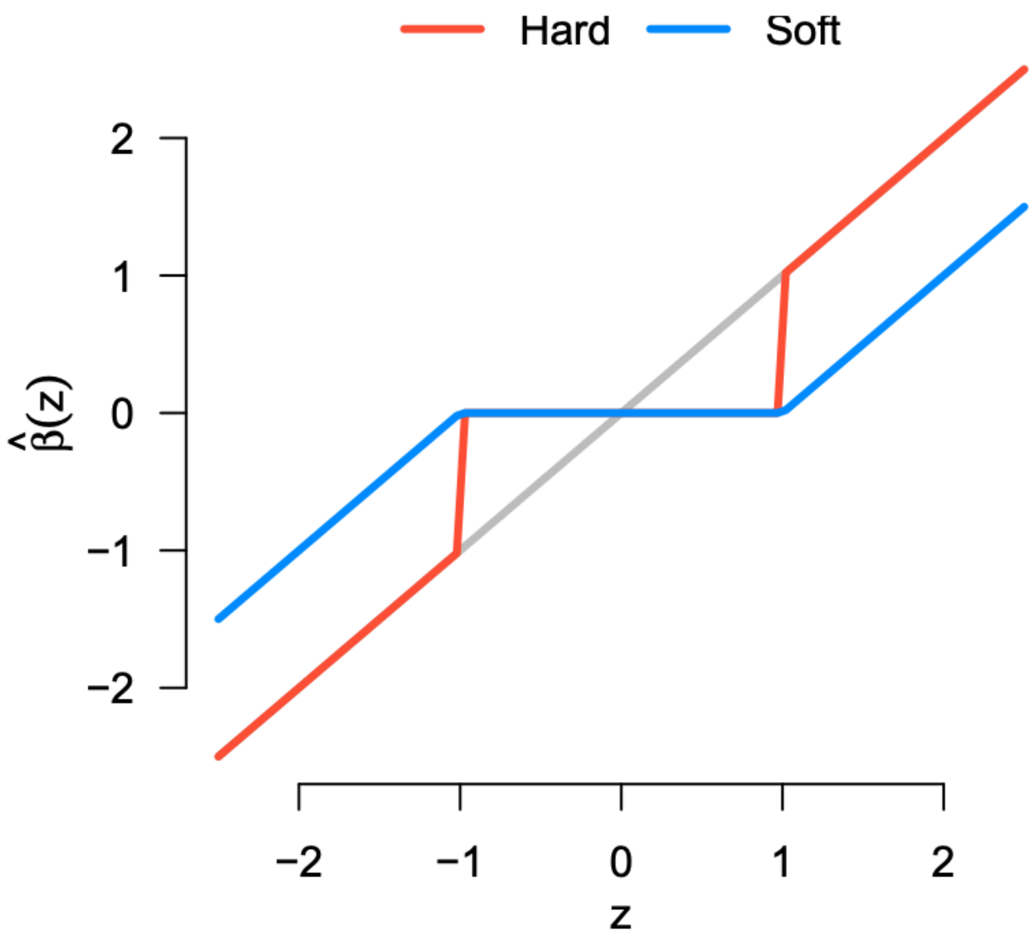
$$S_\lambda(x) = \text{sign}(x)(|x| - \lambda)_+$$

note that

$$(|x| - \lambda)_+ = \begin{cases} |x| - \lambda, & |x| - \lambda > 0 \\ 0 & |x| - \lambda < 0 \end{cases}$$

$$S_\lambda(x) = \begin{cases} \text{sign}(x)(|x| - \lambda), |x| > \lambda \\ 0, |x| \leq \lambda \end{cases}$$

$$= \begin{cases} x - \lambda, & x > \lambda > 0 \\ x + \lambda, & x < -\lambda < 0 \\ 0, & |x| \leq \lambda \end{cases}$$



3 多变量循环坐标下降

object function for multiple variables. at j^{th} step, we want to minimize:

$$\frac{1}{2n}\sum_{i=1}^{n}\left(y_i-\sum_{k\neq j}x_{ik}\beta_k-x_{ij}\beta_j\right)^2+\lambda\sum_{k\neq j}|\beta_k|+\lambda\left|\beta_j\right|$$

$$fix\left\{\hat{\beta}_k,k\neq j\right\},\text{ update }\beta_j$$

denote partial residual $\gamma_i^{(j)}=y_i-\sum_{k\neq j}x_{ik}\hat{\beta}_k$, which could be regarded as the y in single variable senario.

$$\Rightarrow \hat{\beta}_j = S_\lambda \left(\frac{1}{n} \left\langle x_j, r^{(j)} \right\rangle \right)$$
$$\Rightarrow \hat{\beta}_j \leftarrow S_\lambda \left(\hat{\beta}_j + \frac{1}{n} \left\langle x_j, r \right\rangle \right)$$

where $r_i = y_i - \sum_{j=1}^{p} x_{ij\beta j}$