Prop: (Optimality Condition)

For $x^* \in D(f)$,

$$f(x^*) = \min_{x \in D(f)} f(x) \iff 0 \in \partial f(x^*)$$

Proof: x^* is a minimizer

$$\iff f(x^*) = \min_{x \in D(f)} f(x)$$

$$\overset{\text{def of minimizer}}{\iff} f(y) \geq f(x^*) \quad \forall y \in D(f)$$

$$\iff f(y) \geq f(x^*) + \underbrace{0^T(y - x^*)}_{=0}$$

$$\overset{\text{def of subgradient}}{\iff} 0 \text{ is a subgradient of } f \text{ at } x^*$$

$$\overset{\text{def of subdifferential}}{\iff} 0 \in \partial f(x^*)$$

∎

2.3.2 Theory

There is no closed-form solution for β^{lasso} in general, but there is when X satisfies $XX^T = I_p$ where I_p is the $p \times p$ identity matrix. In this case,

$$\beta^{ls} = (\underbrace{XX^T}_{I_p})^{-1} X y = Xy$$

Recall: $\beta^{lasso} = \arg\min \underbrace{\| y - X^T\beta \|^2 + \lambda \|\beta\|_1}_{f(\beta)}$

where $f(\beta) = (y - X^T\beta)^T (y - X^T\beta) + \lambda \|\beta\|_1$

$$= \beta^T \underbrace{XX^T}_{I_p} \beta - 2 \underbrace{y^T X^T}_{=(Xy)^T = \beta^{lsT}} \beta + y^T y + \lambda \|\beta\|_1$$

$XX^T = I_p$, $\beta^{ls} = Xy$

$$= \beta^T\beta - 2\beta^{lsT}\beta + \lambda\|\beta\|_1 + y^Ty$$

$$= \sum_{j=1}^{p} \beta_j^2 - 2\sum_{j=1}^{p} \beta_j^{ls}\beta_j + \lambda\sum_{j=1}^{p} |\beta_j| + y^Ty$$

$$= \sum_{j=1}^{p} \left(\beta_j^2 - 2\beta_j^{ls}\beta_j + \lambda|\beta_j| \right) + y^T y$$

Therefore:

$$\beta^{lasso} = \arg\min_{\beta} f(\beta)$$

$$= \arg\min_{\beta} \left[\sum_{j=1}^{p} \left(\beta_j^2 - 2\beta_j^{ls}\beta_j + \lambda|\beta_j| \right) + y^T y \right]$$

$y^T y$ indep. of β

$$= \arg\min_{\beta_1, \dots, \beta_p} \left[\sum_{j=1}^{p} \left(\beta_j^2 - 2\beta_j^{ls}\beta_j + \lambda|\beta_j| \right) \right]$$

$$= \sum_{j=1}^{p} \arg\min_{\beta_j} \left(\beta_j^2 - 2\beta_j^{ls}\beta_j + \lambda|\beta_j| \right)$$

It suffices to minimize $\beta_j^2 - 2\beta_j^{ls}\beta_j + \lambda|\beta_j|$ for each $j = 1, \dots, p$. For a fixed j,

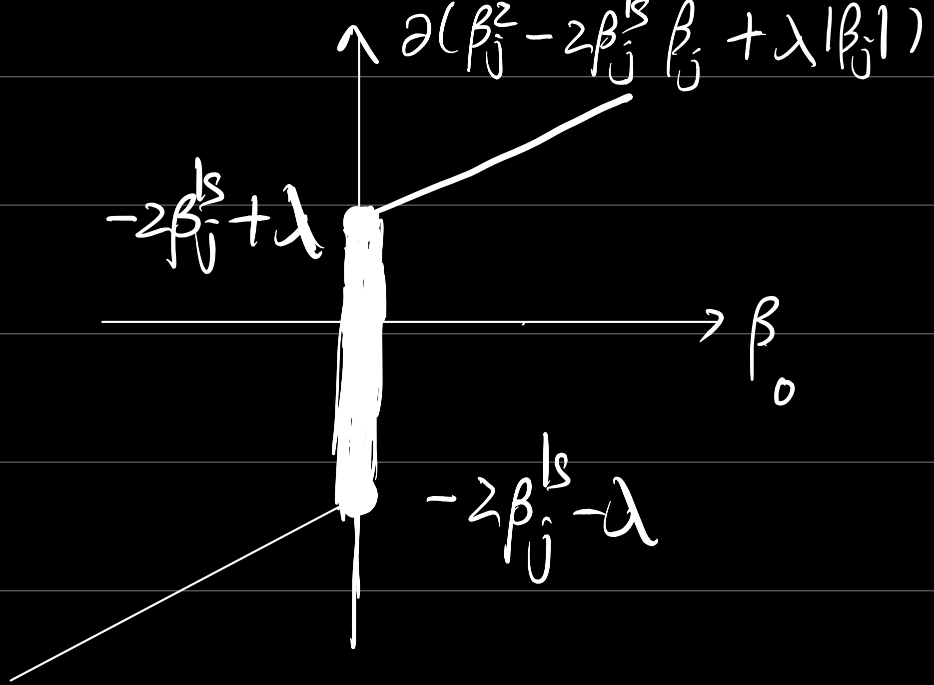
$$\partial\left(\beta_j^2 - 2\beta_j^{ls}\beta_j + \lambda|\beta_j| \right)(\beta_0)$$

$$= \partial(\beta_j^2)(\beta_0) + \partial(-2\beta_j^{ls}\beta_j)(\beta_0) + \partial(\lambda|\beta_j|)(\beta_0)$$

subdifferential of $|\cdot|$

$$= 2\beta_{j0} - 2\beta_j^{ls} + \begin{cases} \lambda & \beta_0 > 0 \\ \lambda[-1,1] & \beta_0 = 0 \\ -\lambda & \beta_0 < 0 \end{cases}$$

$$= \begin{cases} 2\beta_0 - 2\beta_j^{ls} + \lambda & \beta_0 > 0 \\ -2\beta_j^{ls} + \lambda[-1,1] & \beta_0 = 0 \\ 2\beta_0 - 2\beta_j^{ls} - \lambda & \beta_0 < 0 \end{cases}$$



To find the minimizer, we need $0 \in \partial f(\beta_0)$ by the optimality condition.

If $\beta_0 > 0$, $\quad 0 \in \partial f(\beta_0)$

$$\Leftrightarrow \quad 0 = 2\beta_0 - 2\beta_j^{ls} + \lambda$$

$$\Leftrightarrow \quad \beta_0 = \beta_j^{ls} - \frac{\lambda}{2}$$

If $\beta_0 < 0$, $\quad 0 \in \partial f(\beta_0)$

$$\Leftrightarrow \quad 0 = 2\beta_0 - 2\beta_j^{ls} - \lambda$$

$$\Leftrightarrow \quad \beta_0 = \beta_j^{ls} + \frac{\lambda}{2}$$

If $\beta_0 = 0$, $\quad 0 \in \partial f(\beta_0)$

$$\Leftrightarrow \quad 0 \in [-2\beta_j^{ls} - \lambda, \ -2\beta_j^{ls} + \lambda]$$

$$\Leftrightarrow \quad \begin{cases} -2\beta_j^{ls} - \lambda \leq 0 \\ -2\beta_j^{ls} + \lambda \geq 0 \end{cases} \quad \Leftrightarrow \quad \begin{cases} \lambda \geq -2\beta_j^{ls} \\ \lambda \geq 2\beta_j^{ls} \end{cases}$$

Putting together:

$$\beta_i^{lasso} = \begin{cases} \beta_j^{ls} - \frac{\lambda}{2} & \beta_j^{ls} - \frac{\lambda}{2} > 0 \\ 0 & \frac{\lambda}{2} \geq \beta_j^{ls} \geq -\frac{\lambda}{2} \end{cases}$$

$\left(\quad \beta^{ls}_{j} + \frac{\lambda}{2} \qquad \beta^{ls}_{j} + \frac{\lambda}{2} < 0 \right.$