

LASSO

$$\hat{\beta}^{\text{lasso}} = \arg \min_{\beta} \left\{ \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\}$$

$\lambda = 2\lambda$

$$= \arg \min_{\beta} \left\{ \frac{1}{2} (y^T y - 2y^T X\beta + \beta^T X^T X \beta) + \lambda \|\beta\|_1 \right\}$$

$X^T X = I$

$$= \arg \min_{\beta} \left\{ \frac{1}{2} (\hat{\beta}^{\text{OLS}}{}^T \beta + \beta^T \hat{\beta}^{\text{OLS}}) + \lambda \|\beta\|_1 \right\}$$

Does not depend on  $\beta$

$$= \arg \min_{\beta} \left\{ \frac{1}{2} \sum_{j=1}^p (-2\hat{\beta}_j^{\text{OLS}} \beta_j + \beta_j^2 + 2\lambda |\beta_j|) \right\}$$

For each  $j$ ,  $\text{loss}(\beta_j) = -2\hat{\beta}_j^{\text{OLS}} \beta_j + \beta_j^2 + 2\lambda |\beta_j|$

$$\hat{\beta}_j^{\text{OLS}} > 0$$

$$\hat{\beta}_j^{\text{OLS}} < 0$$

$\hat{\beta}_j^{\text{OLS}} > 0$  ) if  $\hat{\beta}_j^{\text{OLS}} > 0$ , then  $\beta_j > 0$

(if not,  $-2\hat{\beta}_j^{\text{OLS}} \beta_j$  would be positive,  $\text{loss}(\beta_j) > \text{loss}(\beta_j)$ )

$\hat{\beta}_j^{\text{OLS}} < 0$  ) if  $\hat{\beta}_j^{\text{OLS}} < 0$ , then  $\beta_j < 0$  (for the same reason)

•  $\hat{\beta}_j^{\text{OLS}} > 0$       $\frac{d \text{loss}(\beta_j)}{d \beta_j} = -\hat{\beta}_j^{\text{OLS}} + \beta_j + \lambda = 0$

$$\hat{\beta}_j^{\text{lasso}} = \hat{\beta}_j^{\text{OLS}} - \lambda$$

$$= (|\hat{\beta}_j^{\text{OLS}}| - \lambda)_+$$

•  $\hat{\beta}_j^{\text{OLS}} < 0$       $\frac{d \text{loss}(\beta_j)}{d \beta_j} = -2\hat{\beta}_j^{\text{OLS}} + 2\beta_j - 2\lambda = 0$

$$\hat{\beta}_j^{\text{lasso}} = \hat{\beta}_j^{\text{OLS}} + \lambda$$

$$= (|\hat{\beta}_j^{\text{OLS}}| - \lambda)_+$$

Combining  $\hat{\beta}_j^{\text{lasso}} = (|\hat{\beta}_j^{\text{OLS}}| - \lambda)_+ \quad \text{if } \hat{\beta}_j^{\text{OLS}} > 0$

$-\hat{\beta}_j^{\text{lasso}} = (|\hat{\beta}_j^{\text{OLS}}| - \lambda)_+ \quad \text{if } \hat{\beta}_j^{\text{OLS}} < 0$

$\text{sgn}(\hat{\beta}_j^{\text{OLS}}) \hat{\beta}_j^{\text{lasso}} = (|\hat{\beta}_j^{\text{OLS}}| - \lambda)_+$

$\hat{\beta}_j^{\text{lasso}} = \text{sgn}(\hat{\beta}_j^{\text{OLS}}) (|\hat{\beta}_j^{\text{OLS}}| - \lambda)_+$