Lecture September 9

$$\hat{\beta}_{OLS} = \underline{(X^T X)}^{-1} X^T y$$

$X \in \mathbb{R}^{n \times p}$ $\quad y \in \mathbb{R}^n$

$\beta \in \mathbb{R}^p$

$X^T X$ can be singular

$$X = U \Sigma V^T$$

$$U^T U = U U^T = \mathbb{1}_{n \times n}$$

$$V^T V = V V^T = \mathbb{1}_{p \times p}$$

$$\Sigma = \begin{bmatrix} \sigma_0 & & & \\ & \sigma_1 & & \\ & & \ddots & \sigma_{p-1} \\ & & & 0 \end{bmatrix}$$

$$\Sigma \in \mathbb{R}^{n \times p}$$

$$\frac{1}{} X^T X = \mathrm{cov}[\underline{x}]$$

n

$$X = [\, x_0 \; x_1 \; x_2 \; \dots \; x_{p-1} \,]$$

$$\mathrm{cov}(x_i, x_j) = \underline{\frac{1}{n} \sum_{l=0}^{n-1} (x_{li} - \bar{\mu}_x)(x_{lj} - \bar{\mu}_x)}$$

$$\mathrm{cov}[X] = \begin{array}{c|c} & \begin{bmatrix} \mathrm{var}[x_0] & \mathrm{cov}[x_0, x_1] & \dots \\ \vdots & & \\ \mathrm{cov}[x_{p-1}, x_0] & \dots & \mathrm{cov}[x_{p-1}, x_{p-1}] \end{bmatrix} \end{array}$$

$0, 1, 2, \dots, n-1$

$p=0 \quad p=1 \quad \dots \quad p-1$

$X^T \in \mathbb{R}^{p \times n}$

$X \in \mathbb{R}^{n \times p}$

$$\frac{1}{n} [X^T X] V = \Sigma^2 V \frac{1}{n}$$

$$\mathrm{var}(\hat{\beta}_{OLS}) \sim (X^T X)^{-1}$$

$$\frac{\partial^2 C}{\partial \beta^T \partial \beta} = X^T X$$

Ridge Regression

$$\hat{\beta}_{Ridge} = (X^T X + \lambda I_{p \times p})^{-1} X^T y$$

$$\hat{\beta}_{Ridge} = \underset{\beta \in \mathbb{R}^p}{\arg\min} \; \| (y - X\beta) \|_2^2 + \lambda \|\beta\|_{\underline{2}}^2$$

$$\sum_{i=0}^{p-1} \beta_i^2 \leq t < \infty$$

Lasso regression

$$\hat{\beta}_{Lasso} = \underset{\beta \in \mathbb{R}^p}{\arg\min} \; \| (y - X\beta) \|_2^2 + \lambda \|\beta\|_1$$

$$\sum_{i=0}^{p-1} |\beta_i|$$

$$\frac{d|\beta_i|}{d\beta_i} := \begin{cases} 1 & \text{if } \beta_i > 0 \\ 0 & \text{if } \beta_i = 0 \\ -1 & \text{if } \beta_i < 0 \end{cases}$$

$= \mathrm{sgn}(\beta_i)$

Example

$$X = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & 1 & \\ 0 & & \ddots & \\ & & & 1 \end{bmatrix}$$

OLS - cost function

$$C(\beta) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \beta_i)^2$$

$$\boxed{p = n} \qquad \frac{\partial C}{\partial \beta_i} = 0 = -\frac{1}{n} 2(y_i - \beta_i)$$

$$\hat{\beta}_i^{OLS} = y_i$$

Ridge

$$C(\beta) = \sum_{i=0}^{n-1} (y_i - \beta_i)^2 + \lambda \sum_{i=0}^{n-1} \beta_i^2$$

$$\hat{\beta}_i^{Ridge} = \frac{y_i}{(1+\lambda)}$$

$1+\lambda$

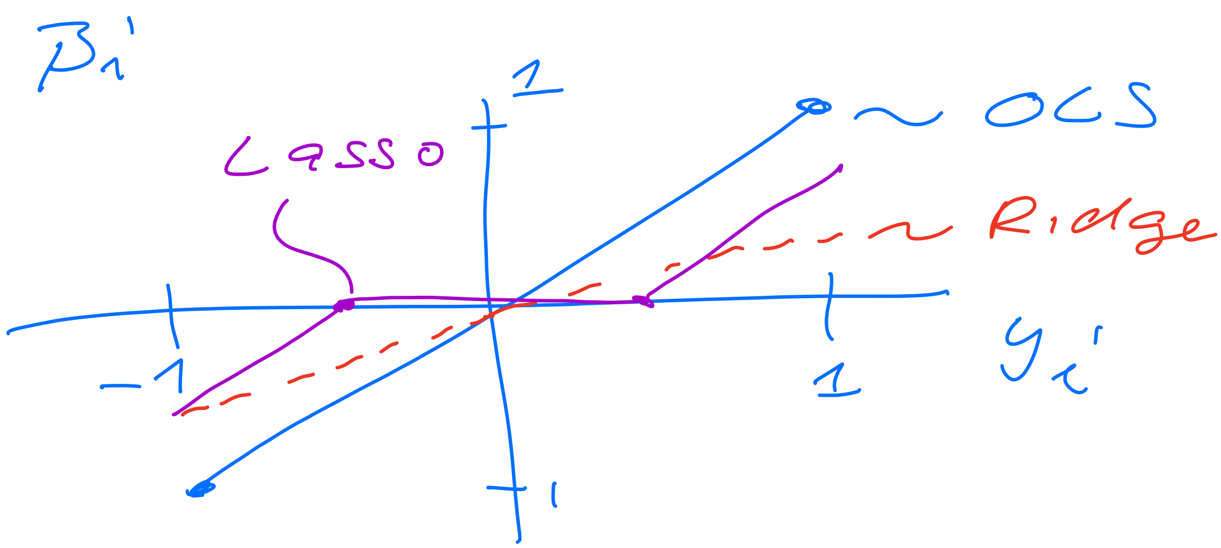
Lasso

$$\frac{\partial}{\partial \beta_i}\left[\sum_i (y_i-\beta_i)^2+\lambda\sum_i|\beta_i|\right]$$

$$= 0 \qquad \lambda > 0$$

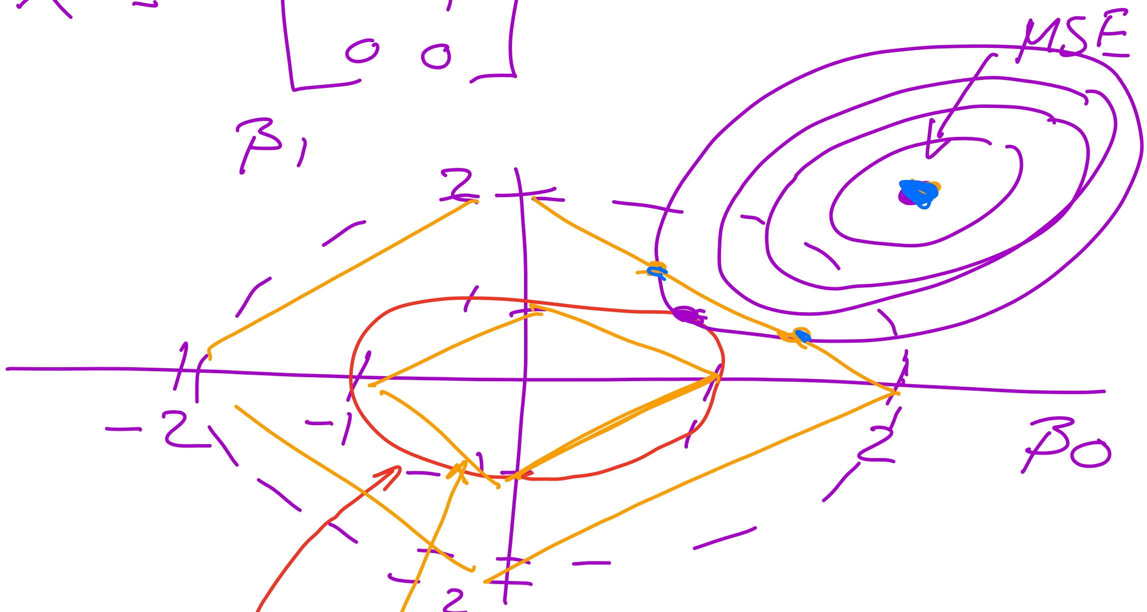
$$-2\sum_i (y_i-\beta_i)+\lambda\sum_i \mathrm{sgn}(\beta_i)$$

$$\hat{\beta}_i^{\,lasso} = \begin{cases} \underline{y_i-\lambda/2} & \text{if } y_i>\lambda/2 \\ y_i+\lambda/2 & \text{if } y_i<-\lambda/2 \\ 0 & \text{if } |y_i|\le\lambda/2 \end{cases}$$



$$y = \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix} \qquad \hat{\beta}_{OLS} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$X = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$



Ridge $\underline{(\beta_0^2 + \beta_1^2)} = 1$

$$\lambda \sum_{i=0}^{p-1} \beta_i^2$$

Lasso

$$\underline{|\beta_0| + |\beta_1|} = 1$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

$$y_i = f(x_i) + \varepsilon_i$$

$$\mathbb{E}[y_i] = \underbrace{\mathbb{E}[f(x_i)]}_{\mathbb{E}\left[X_{i*}\hat{\beta}\right]} + \underbrace{\mathbb{E}[\varepsilon_i]}_{= 0}$$

$$\mathbb{E}[X] = \sum_{i=0}^{n-1} x_i \, p(x_i)$$

$$x_{1961} \, p(1961) = 1$$

$$\mathbb{E}[y_i] = X_{i*}\hat{\beta} + 0$$

$$= X_{i*}\hat{\beta}$$

$$\mathrm{var}[y_i] = \mathrm{var}[\varepsilon_i]$$

$$= \sigma^2$$

$$y_i \sim N(x_i \beta, \sigma^2)$$