Lecture FYS-STK3155/4155, September 7

$$C(\beta) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \tilde{y}_i)^2 + \lambda \sum_{j=0}^{p-1} \beta_j^2$$

$$\tilde{y}_i = \sum_j x_{ij} \beta_j$$

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

$$\tilde{y} = X\beta$$

$$\frac{\partial C}{\partial \beta} = -\frac{2}{n} X^T (y - X\beta) + 2\lambda\beta = 0$$

$$X^T y = X^T X \beta + \underbrace{n\lambda}_{\lambda} \beta$$

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

Simple case

$$X = \mathbb{1}$$

$$\tilde{y} = \beta \qquad p = n$$

$$\tilde{y}_i = \beta_i$$

OLS:

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

$\boxed{y_i = \hat{\beta}_i}$ (minimize wrt β)

Ridge

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

$$\frac{\partial C}{\partial \beta_i} = -\frac{2}{n}(y_i - \beta_i) + 2\lambda \beta_i = 0$$

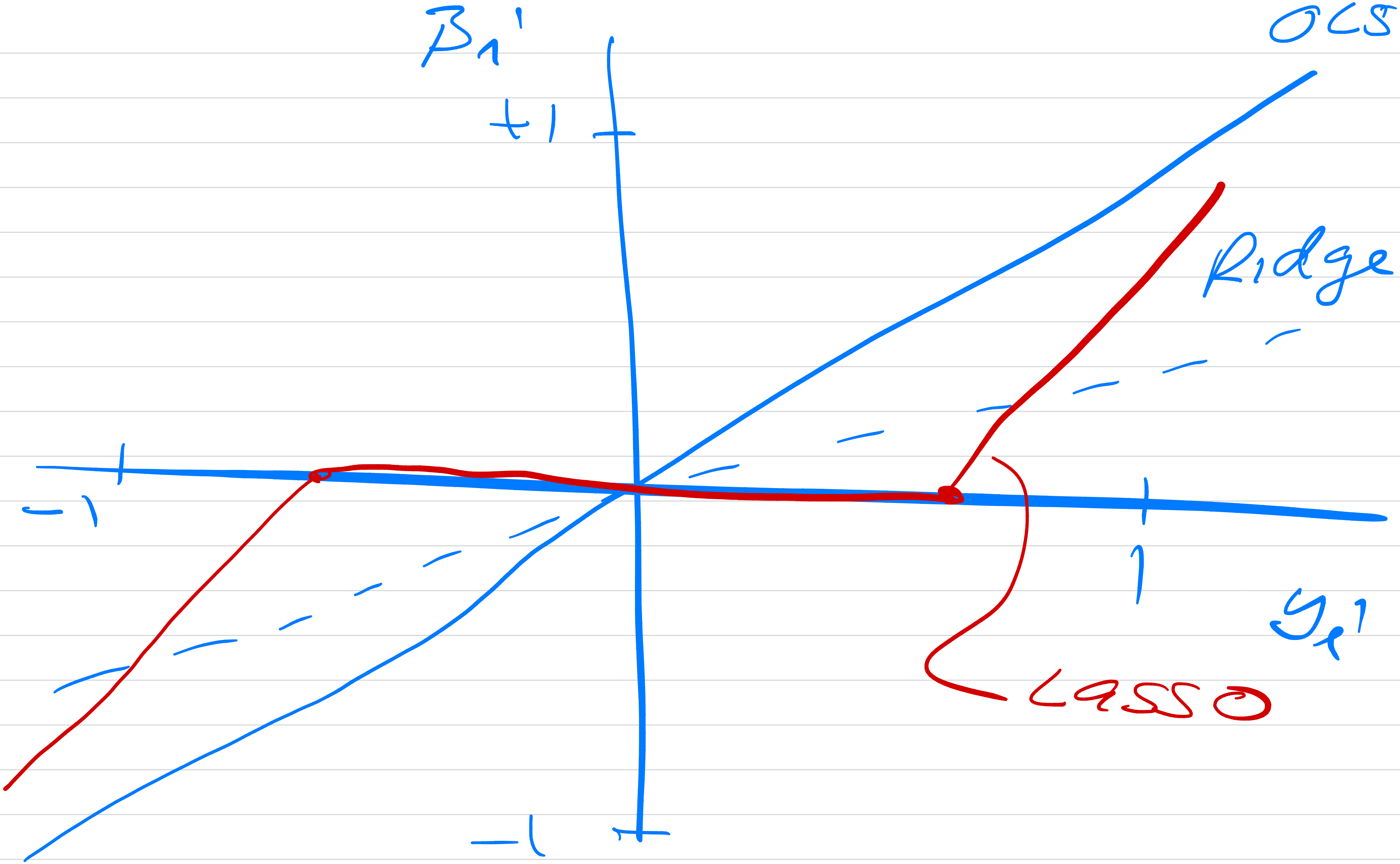
$$\Rightarrow \boxed{\beta_i^{\text{Ridge}} = \frac{y_i}{1+\lambda}}$$

Lasso

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

$$\frac{\partial C}{\partial \beta_i} = -\frac{2}{n}(y_i - \beta_i) + \lambda \frac{\beta_i}{|\beta_i|} = 0$$

$$\beta_i^{lasso} = \begin{cases} y_i - \lambda/2 & \text{if } y_i > \lambda/2 \\ y_i + \lambda/2 & \text{if } y_i < -1/2 \\ 0 & \text{if } |y_i| \leq \lambda/2 \end{cases}$$



β_1'
+1
OLS
Ridge
-1
1
y_1'
Lasso
-1

σ_j^2 are the eigenvalues of $X^T X$

$$\mathrm{var}(\beta) \propto (X^T X)^{-1}$$

model $\hat{y} = X\beta$

$$\beta_i \pm \mathrm{STD}(\beta_i)$$

$$\mathrm{STD}(\beta_i) = \sqrt{\mathrm{var}(\beta_i)} \propto \frac{1}{\sigma_i^2}$$

Links with statistics

$$y = f(x) + \varepsilon$$

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

$$\mathbb{E}[\varepsilon] = \int_{x \in D} dx \; \varepsilon(x) \, p(x)$$

$$= 0$$

$$\mathrm{var}[\varepsilon] = \sigma^2$$

$$\tilde{y} = X\beta \simeq f(x)$$

$$y \simeq X\beta + \varepsilon$$

X is non-stochastic

$$y_i = \sum_j x_{ij}\beta_j + \varepsilon_i$$

$$= x_{i*}\beta + \varepsilon_i$$

$$\mathbb{E}[y_i] = \mathbb{E}[x_{i*}\beta] + \underbrace{\mathbb{E}[\varepsilon_i]}_{=0}$$

$$= x_{i*}\beta$$

$$\operatorname{var}[y_i] = \mathbb{E}\left[(y_i - \mathbb{E}[y_i])^2\right]$$

$$= \mathbb{E}[y_i^2] - (\mathbb{E}[y_i])^2$$

$$= \mathbb{E}\left[(x_i * \beta + \varepsilon_i)^2\right] - (x_i * \beta)^2$$

$$= \mathbb{E}\left[\cancel{(x_i * \beta)^2} + 2 x_i * \beta \underbrace{\varepsilon_i}_{0} + \varepsilon_i^2\right] - \cancel{(x_i * \beta)^2}$$

$$= \mathbb{E}[\varepsilon_i^2] = \sigma^2 \Rightarrow$$

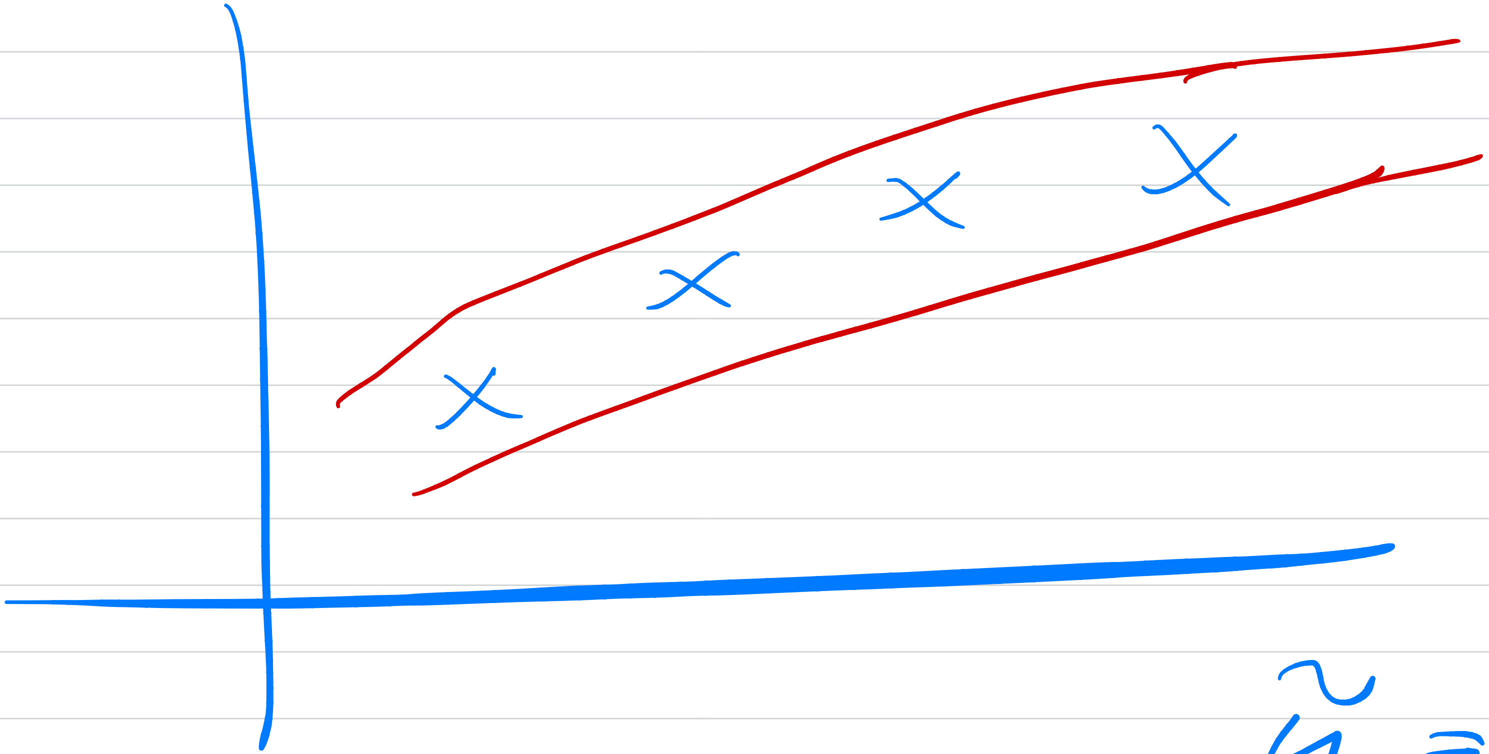
$$y_i \sim N(x_i * \beta, \sigma^2)$$

$$\mathbb{E}[\hat{\beta}] = \mathbb{E}\left[(X^T X)^{-1} X^T y\right]$$

$$= \mathbb{E}\left[(X^T X)^{-1} X^T X \beta\right]$$

$$= \mathbb{E}[\beta] = \beta$$

$$\operatorname{var}[\hat{\beta}] \propto (X^T X)^{-1} \qquad \text{OLS}$$



$$\tilde{y} = X\beta$$