ML ERASMUS, OCT 17, 2022

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

$$X = U \Sigma V^T \quad \text{(SVD)}$$

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

$$U \in \mathbb{R}^{n \times n} \qquad U U^T = U^T U = \mathbb{1}$$

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

$$\sigma_0 > \sigma_1 > \sigma_2 > \dots > \sigma_{p-1} > 0$$

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

$$V \in \mathbb{R}^{p \times p} \qquad V V^T = V^T V = \mathbb{1}$$

$$U = \begin{bmatrix} | & | & & | \\ u_0 & u_1 & \dots & u_{n-1} \\ | & | & & | \end{bmatrix}$$

$$u_i^T u_j = \delta_{ij}$$

$$V = \begin{bmatrix} | & | & & | \\ v_0 & v_1 & \cdots & v_{p-1} \\ | & | & & | \end{bmatrix}$$

$$v_i^T v_j = \delta_{ij}$$

$$\tilde{y}_{OLS} = X \hat{\beta}_{OLS} = \left(\sum_{i=0}^{p-1} u_i u_i^T \right) y$$

$$\tilde{y}_{Ridge} = X \hat{\beta}_{Ridge} =$$

$$\left(\sum_{i=0}^{p-1} u_i u_i^T \frac{\sigma_i^2}{\sigma_i^2 + \lambda} \right) y$$

Example

$$\tilde{y} = \hat{\beta} \qquad X = \mathbb{1}$$

$$n = p$$

OLS :

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

$$\tilde{y} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$$

$$\frac{\partial C}{\partial \beta_i} = 0 \quad \Rightarrow \quad y_i = \beta_i$$

Ridge

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

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

$$\hat{\beta}_i^{ridge} = \frac{y_i}{1+\lambda}$$

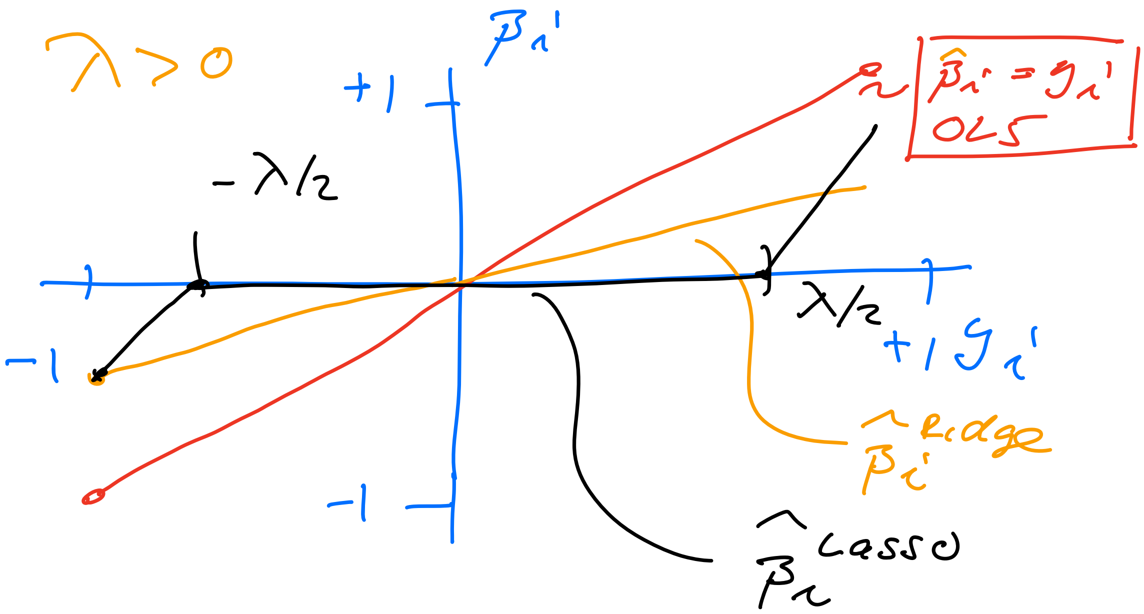
Lasso

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

$$+ \lambda\sum_{i=0}^{p-1}\frac{\sqrt{\beta_i^2}}{|\beta_i|}$$

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

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



Statistical analysis

SVD : $X^T X = V \Sigma^T \Sigma V^T$

$$V^T V = V V^T = \mathbb{1}$$

$$(X^T X) V = V \Sigma^T \Sigma$$

$$\Sigma^T \Sigma = \begin{bmatrix} \sigma_0^2 & & \\ & \ddots & \\ & & \sigma_{p-1}^2 \end{bmatrix}$$

$$V = \begin{bmatrix} | & | & & | \\ v_0 & v_1 & \cdots & v_{p-1} \\ | & | & & | \end{bmatrix}$$

$$(X^T X)\, v_i = \sigma_i^2\, v_i$$

The eigenvalues of $X^T X$ are the singular values of $X = U \Sigma V^T$

$(X^T X)$ is a positive definite matrix.

$$\frac{\partial^2 C(\beta)}{\partial \beta\, \partial \beta^T} = \frac{2}{n} X^T X$$

= Hessian matrix for OLS

Expectation values

$$\mathbb{E}[x^n] = \langle x^n \rangle = \int_D p(x)\, x^n\, dx$$

$$\mu = \langle x \rangle = \int_D p(x)\, x\, dx$$

$$\left(\sum_{i=1}^{D} p(x_i)\, x_i \right)$$

sample mean

$$\bar{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i \neq \mu$$

$p(x)$ is unknown

sample variance

$$\bar{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{\mu})^2$$

$$\text{cov}(x,y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{\mu}_x) \times (y_i - \bar{\mu}_y)$$

(Biased formula $\frac{1}{n} \to \frac{1}{n-1}$)

i.i.d. = independent and identicaly distributed variables.

i.i.d. : $P(x,y)\,dx\,dy$

$$= p(x)\,p(y)\,dx\,dy$$

$$\underline{\text{cov}(x,y)} = \int_D p(x)\,p(y)\,(x-\mu)(y-\mu) \times dx\,dy$$

i.i.d : $\mu = \langle x \rangle = \langle y \rangle =$

$$\int dx\,dy\, p(x)\,p(y)\,x$$

$$= \underbrace{\int dx\, p(x)\, x}_{\mu} \underbrace{\int dy\, p(y)}_{=1}$$

$\mathrm{cov}(x,y) = 0$ if iid

Design matrix

$$X = \begin{bmatrix} x_{00} & x_{01} & \cdots & x_{0\,p-1} \\ x_{10} & & & \\ x_{20} & & & \\ \vdots & & & \\ x_{n-1\,0} & - & - & x_{n-1\,p-1} \end{bmatrix}$$

$$= \begin{bmatrix} | & | & & | \\ \mathbf{x}_0 & \mathbf{x}_1 & - - & \mathbf{x}_{p-1} \\ | & | & & | \end{bmatrix}$$

$$\mathrm{cov}(\mathbf{x}_i, \mathbf{x}_j) = \mathrm{cov}(X)$$

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

$$\mathrm{cov}(X_i, X_j) = \frac{1}{n} \sum_{k=0}^{n-1} (x_{ki} - \mu_i) \times (x_{kj} - \mu_j)$$

$$\boxed{\mathrm{cov}[X] = \frac{1}{n} X^T X \in \mathbb{R}^{p \times p} \quad (X \in \mathbb{R}^{n \times p})}$$

$$\boxed{\mathrm{cov}[X] = \frac{1}{n} X X^T \quad (X \in \mathbb{R}^{p \times n})}$$

Hessian in OLS

$$H = \frac{2}{n} X^T X \propto \mathrm{cov}[X]$$

From SVD we found

$$(X^T X) v_i = \sigma_i^2 v_i$$

eigenvalues of $\mathrm{cov}[X]$

are proportional with σ_i^2

Derivation of OLS, Ridge and Lasso from statistics:

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

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

$f(x)$ is a deterministic function.

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

$$\tilde{y}_i = \sum_{j=0}^{p-1} x_{ij} \beta_j \quad \text{(scalar)}$$

deterministic

$$\langle \tilde{y}_i \rangle = \langle \underbrace{\sum_{j=0} x_{ij} \beta_j}_{} \rangle + \underbrace{\langle \varepsilon_i \rangle}_{= 0}$$

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

$\varepsilon_i \sim N(0, \sigma^2)$ we can deduce that

$$y_i \sim N(x_{i*}\beta, \ ?)$$

<u>Exercise</u>: show

$$\text{var}[y_i] = \sigma^2 \Rightarrow$$

$$y_i \sim N(x_{i*}\beta, \sigma^2)$$

$$p(y_i \mid X, \beta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\left(\frac{(y_i - x_{i*}\beta)^2}{2\sigma^2} \right) \right]$$

we assume that all

$y_i \sim$ i.i.d. $\Rightarrow$

$$p(y \mid X\beta) = \prod_{i=0}^{n-1} p(y_i \mid X\beta)$$

$$= \prod_{i=0}^{n-1} p(y_i) = \prod_{i=0}^{n-1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - x_{i*}\beta)^2}{2\sigma^2}}$$

We want to find β which maximize $P(y|X\beta)$

$$\hat{\beta} = \arg\max_{\beta \in \mathbb{R}^p} P(y|X\beta)$$

$$\frac{\partial P}{\partial \beta} = 0$$

$$\ln P = \sum_{i=0}^{n-1} \ln p(y_i)$$

minimize:

$$\frac{\partial C(\beta)}{\partial \beta} = - \frac{\partial \ln P}{\partial \beta}$$