CompSci, Nov 9, 2022

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

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

$X = U \Sigma V^T$

$UU^T = U^T U = \mathbb{1}$

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

$U \in \mathbb{R}^{n \times n}$

$V \in \mathbb{R}^{p \times p}$

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

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

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

$U = [u_0 \; u_1 \; \ldots \; u_{n-1}]$

$V = [v_0 \; v_1 \; \ldots \; v_{p-1}]$

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

$$= V \Sigma^T \underbrace{U^T U}\, \Sigma V^T$$

$$= V \Sigma^T \Sigma V^T$$

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

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

$$\Sigma^T \Sigma = \begin{bmatrix} \sigma_0^2 & & 0 \\ & \sigma_1^2 & \\ 0 & & \ddots \; \sigma_{p-1}^2 \end{bmatrix}$$

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

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

<u>intermezzo</u>

multiply from the right with V

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

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

Eigenvalues of $X^T X$ are $\sigma_i^2 > 0$

$X^T X$ is positive definite.

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

convex problem with a global minimum.

In Ridge regression

$X^T X + \lambda \mathbb{I}$, also a convex problem.

<u>intermezzo 2</u>

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

$$\mathrm{cov}(x_i, x_j) = \int_D dx_i dx_j \times (x_i - \mu_i)(x_j - \mu_j)$$

$$\times\ p(x_i, x_j)$$

if x_i, x_j are i.i.d. $\left(p(x_i)\, p(x_j) = p(x_i, x_j) \right)$

then $\operatorname{cov}(x_i, x_j) = 0$

$$X = \left[\mathbf{x}_0\ \mathbf{x}_1\ x_2\ \dots\ x_{p-1} \right]$$

$$\operatorname{cov}(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{n} \sum_{k=0}^{n-1} (x_{ki} - \mu_i)(x_{kj} - \mu_j)$$

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

$$\quad \mathbf{x}_0 \quad \mathbf{x}_1 \quad \dots \quad \mathbf{x}_{p-1}$$

$$\operatorname{cov}(x_i, x_j) \longrightarrow \operatorname{cov}[X]$$

$$= \frac{1}{n} \underline{X^T X}$$

The eigenvalues are proportional with σ_i^2

$$\mathrm{cov}[x] = \begin{bmatrix} \sigma_0^2 & \mathrm{cov}[x_0 x_1] & \cdots & \mathrm{cov}[x_0 x_{p-1}] \\ x & \sigma_1^2 & x & x \\ x & x & x & x \\ x & x & x & x \\ & x & x & x & \sigma_{p-1}^2 \end{bmatrix}$$

$$S\,\mathrm{cov}[x]\,S^T = D$$

$$= \begin{bmatrix} \sigma_0^2 & & & 0 \\ & \sigma_1^2 & & \\ 0 & & \ddots & \end{bmatrix}$$

$$\mathrm{cov}[Y] = S\,\mathrm{cov}[X]\,S^T$$

$$Ax = \lambda x$$

$$SAS^T = D = \begin{bmatrix} \lambda_0 & & \\ & \ddots & \\ & & \lambda_{n-1} \end{bmatrix}$$

$$SAx = S\lambda x = \lambda Sx$$

$$SS^T = \mathbb{1} = S^T S$$

$$\underbrace{SAS^T}_{B}\underbrace{Sx}_{y} = \lambda \underbrace{Sx}_{y}$$

$$B \cdot y = \lambda y$$

Back to OLS and Ridge with SVD

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

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

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

$$= X (X^T X)^{-1} X^T y$$

$$= U \Sigma V^T (V \Sigma^T \Sigma V^T)^{-1} V \Sigma^T U^T y$$

$$(AB)^{-1} = B^{-1} A^{-1}$$

A & B are square matrices and both are invertible

$$V V^T = \mathbb{1} \qquad V = (V^T)^{-1}$$

$$(V \Sigma^T \Sigma V^T)^{-1} = (V^T)^{-1} (\Sigma^T \Sigma)^{-1} V^{-1}$$

$$= V (\Sigma^T \Sigma)^{-1} V^T$$

$$\tilde{y} = X \hat{\beta}_{OLS} = u u^T y$$

$$= \left(\sum_{i=0}^{p-1} u_i u_i^T \right) y$$

Ridge

$$\hat{\beta}_{Ridge} = (X^T X + \lambda \mathbb{I})^{-1} X^T y$$

SVD :

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

$$= \sum_{j=0}^{p-1} u_j u_j^T \frac{\sigma_j^2}{\sigma_j^2 + \lambda}$$

$$\sigma_0^2 > \sigma_1^2 > \dots > \sigma_{p-1}^2 > 0$$

Simple example

$$X = \mathbb{1} \qquad n = p$$

$$\tilde{y}_{OLS} = \hat{\beta}_{OLS}$$

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

$$\tilde{y}_i = \sum_{j=0}^{p-1} x_{ij} \beta_j = \beta_i$$

$$x_{ij} = 1 \text{ if } i = j$$

minimize with respect to $\beta_j \Rightarrow$

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

$$\frac{\partial C_{OLS}}{\partial \beta_j} = -\frac{2}{n}(y_j - \beta_j) = 0$$

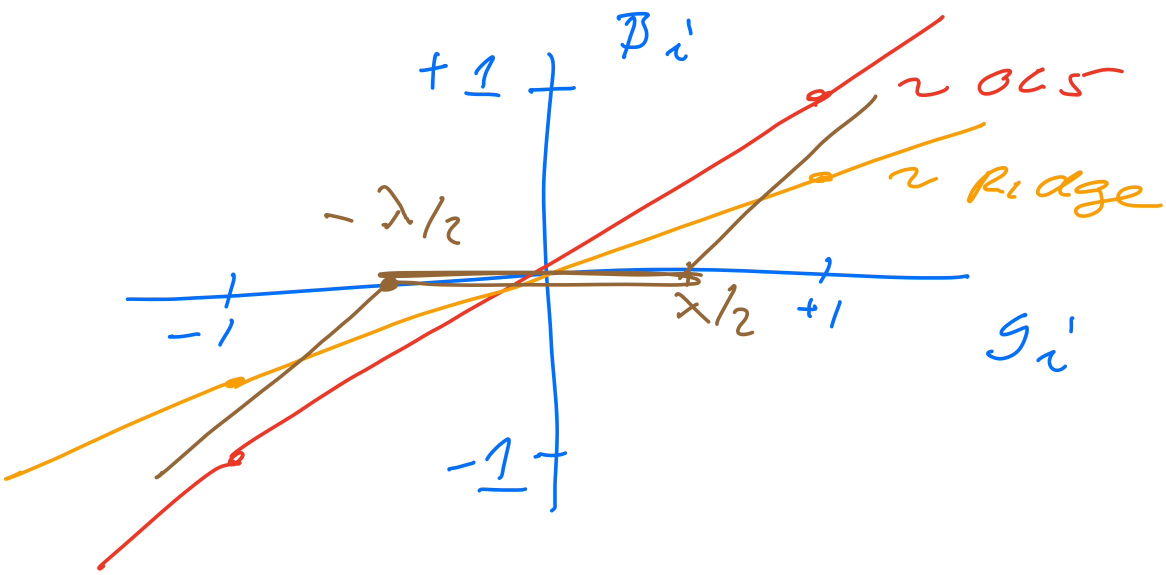
Ridge regression

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

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

$\Rightarrow$

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



Lasso Regression

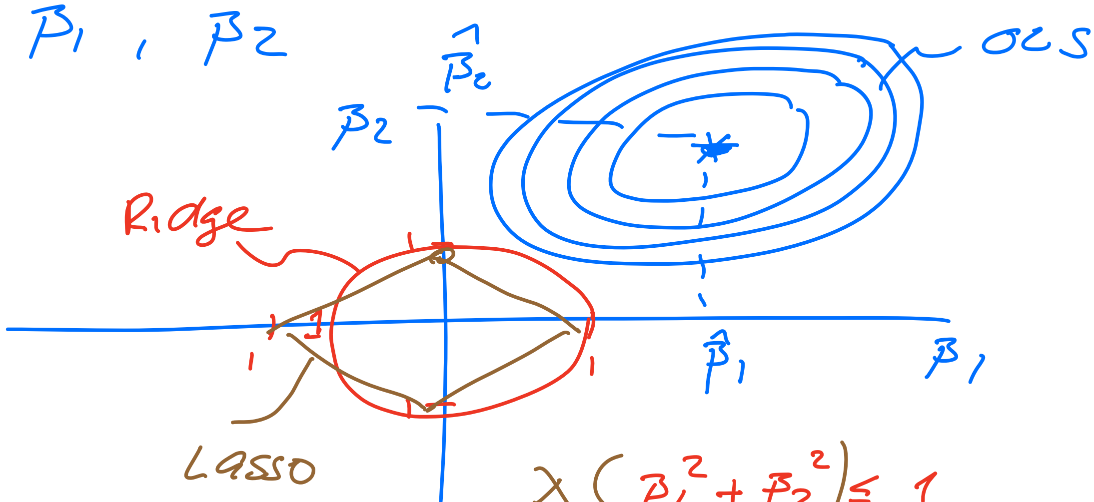
$$C(\beta)_{Lasso} = \frac{1}{n}\sum_{j=0}^{p-1}(y_i-\beta_i)^2 + \lambda\sum_{j=0}^{p-1}|\beta_j|$$

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

$$\frac{d|x|}{dx} = \begin{cases} +1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \end{cases}$$

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

β_1 , β_2



$\lambda(\beta_1^2 + \beta_2^2) \le 1$