FYS-STK 4155 Sept 8

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

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

$\lambda > 0$

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

$\beta \in \mathbb{R}^{p}$ $\quad I \in \mathbb{R}^{p \times p}$

SVD

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

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

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

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

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

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

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

$$U = [u_0 \; u_1 \; u_2 \; \dots \; u_{n-1}]$$

$$\langle u_i | u_j \rangle = u_i^T u_j$$
$$= \delta_{ij}$$

$$V = [v_0 \; v_1 \; v_2 \; \dots \; v_{p-1}]$$

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

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

$$n \geq p \qquad (n \gg p)$$

Example

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

3x2 matrix

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

$$\underset{2\times 3}{\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}} \underset{3\times 2}{\begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

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

OLS : $X^T X =$

$$V \Sigma^T \underbrace{U^T U}_{I} \Sigma V^T$$

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

$$\hat{y} = X\hat{\beta} = X\left(X^TX\right)^{-1}X^Ty$$

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

$$\times\ V\Sigma^T U^T y$$

$$\left(\begin{array}{l} A, B \text{ square, invertible} \\ (A\cdot B)^{-1} = B^{-1}A^{-1} \\ VV^T = V^TV = I \qquad V = (V^T)^{-1} \end{array}\right)$$

$$\hat{y} = U\Sigma \underbrace{V^T\cdot V}_{1}\left(\Sigma^T\Sigma\right)^{-1}\underbrace{V^TV}_{1}\Sigma^T U^T y$$

note sum

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

Ridge

$$\hat{\beta} = \left(X^TX + \lambda I\right)^{-1}X^T\cdot y$$

$$\tilde{y}_{Ridge} = U \Sigma V^T \left(V \Sigma^T \Sigma V^T + \lambda \underset{VV^T}{I} \right)^{-1}$$

$$\times (U \Sigma V^T)^T y$$

$$\tilde{y}_{Ridge} = \left[\sum_{j=0}^{p-1} u_j u_j^T \frac{\sigma_j^2}{\sigma_j^2 + \lambda} \right] y$$

— sums to $p-1$

$$\tilde{y}_{OLS} = \left[\sum_{j=0}^{p-1} u_j u_j^T \right] y$$

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

$\sigma_p \dots \sigma_{n-1} = 0$

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

what happens when λ is large

Example:

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

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

$$\tilde{y}_{Ridge} = ?$$

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

$$= \frac{1}{1 + \lambda 1} \underline{\hat{\beta}_{OLS}}$$

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

Covariance matrix

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

$$= \begin{bmatrix} x_0 & x_1 & \cdots & x_{p-1} \end{bmatrix}$$

$$\mathrm{cov}[x_i\, x_j] = \frac{1}{n} \sum_{k=0}^{n-1} (x_{ki} - \mu_i)(x_{kj} - \mu_j)$$

covariance matrix

$$C[X] = \frac{1}{n} X^T X$$

$$\left(\mathbb{E}[xx^T] \right)$$

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

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

where is the SVD?

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

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

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

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