Lecture October 25

1st example

$$X = \begin{bmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & & \vdots \\ 1 & & \vdots \\ \vdots & & \\ 1 & & x_{n-1}^2 \end{bmatrix}$$

$$X \in \mathbb{R}^{n \times 3}$$

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

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

2nd example

$$X = \begin{bmatrix} x_0 & x_0^2 \\ \vdots & \\ \vdots & \\ \vdots & \\ x_{n-1} & x_{n-1}^2 \end{bmatrix}$$

Lasso & Ridge

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

Sklearn: $\underline{\lambda \sum_{j=1}^{p-1} \beta_j^2}$

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

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

Lasso

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

Sklearn $\lambda \sum_{j=1}^{p-1} |\beta_j|$

Math of the SVD

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

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

$X^T X$

$$\hat{x} \quad \hat{}$$

OLS: $C(\beta) = \frac{1}{n} \sum \left(y_i - \sum_j x_{ij}\beta_j\right)^2$

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

$$\frac{2}{n} X^T X = \frac{\partial^2 C}{\partial \beta^T \partial \beta} = \text{Hessian matrix}$$

$\mathrm{cov}(X)$

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

$$= \begin{bmatrix} | & | & | & & | \\ x_0 & x_1 & x_2 & \cdots & x_{p-1} \\ | & | & | & & | \end{bmatrix}$$

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

$$\sum_{i=0}^{n} \qquad \frac{1}{n-1} \approx \frac{1}{n} \qquad - \mu_{x_j})$$

$$\mu_x = \int_{x \in D} x \, p(x) \, dx$$

sample mean

$$\bar{\mu} = \frac{1}{n} \sum_{i=0}^{n-1} x_i \neq \mu_x$$

sample variance

$$\bar{\sigma}^2 = \frac{1}{n} \sum_{i=0}^{n-1} (x_i - \bar{\mu})^2$$

$$cov(x_i, x_i) = \frac{1}{n} \sum_{l=0}^{n-1} (x_{lj} - \bar{\mu}_{x_i})^2$$

$$= var(x_i)$$

covariance matrix

$$= \begin{bmatrix} var(x_0) & cov(x_0, x_1) & \cdots & cov(x_0, x_{p-1}) \\ \vdots & var(x_1) & & \\ \vdots & & \ddots & \\ cov(x_{n-1}, x_0) & \cdots & \cdots & var(x_{p-1}) \end{bmatrix}$$

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

correlation

$$\mathrm{corr}(x_i, x_j) = \frac{\mathrm{cov}(x_i, x_j)}{\sqrt{\mathrm{var}(x_i)\,\mathrm{var}(x_j)}}$$

$$= \begin{bmatrix} 1 & \mathrm{corr}(x_0, x_1) & \cdots \\ \mathrm{corr}(x_1, x_0) & 1 & \\ \vdots & & 1 & \\ \vdots & & & \ddots & \\ & & & & 1 \end{bmatrix}$$

$\Rightarrow$ unsupervised method

Principal component analysis (PCA)

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

$$\mathrm{var}(\hat{\beta}_j) \propto (X^T X)^{-1}_{j,j}$$

<u>SVD</u>

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

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

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

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

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

$$UU^T = U^T U = I_{n \times n} = I_n$$

$$VV^T = V^T V = I_{p \times p} = I_p$$

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

$$= \begin{bmatrix} \tilde{\Sigma} \\ 0 \end{bmatrix}$$

$$\tilde{\Sigma} \in \mathbb{R}^{p \times p} = \begin{bmatrix} \sigma_0 & & \\ & \ddots & \\ & & \sigma_{p-1} \end{bmatrix}$$

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

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

$$X^T X = V \Sigma^T \underbrace{U^T U}_{I_n} \Sigma V^T$$

$$\Sigma^T I_n \Sigma$$

$$\mathbb{R}^{p \times n} \times I_n \quad \mathbb{R}^{n \times p}$$

$$\tilde{\Sigma}^2 = \Sigma^T \Sigma$$

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

$$X^T X = V \tilde{\Sigma}^2 V^T \qquad | \; V$$

$$V^T V = I_p$$

$$(X^T X) V = V \tilde{\Sigma}^2 = \tilde{\Sigma}^2 V$$

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

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

covariance matrix has

eigenpairs given by the singular values of X and the orthogonal vectors of V^T.

$$XX^T = U\Sigma \underbrace{V^T \cdot V}\Sigma U^T$$

$$= U \Sigma^2 U^T \quad (\Sigma^2 \neq \tilde{\Sigma}^2)$$

$$| U$$

$$(XX^T)U = U\Sigma^2$$

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

$$\tilde{y}_{OLS} = X(X^TX)^{-1}X^Ty$$

$$= U\Sigma V^T (V\tilde{\Sigma}^2 V^T)^{-1} V \Sigma^T U^T$$

$$= UU^T y \in \mathbb{R}^{p \times p}$$

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

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

$$j=0$$

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

Simple example

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

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

OLS

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

$$\frac{\partial C}{\partial \beta_j} = 0 \Rightarrow y_j = \hat{\beta}_j$$

Ridge

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

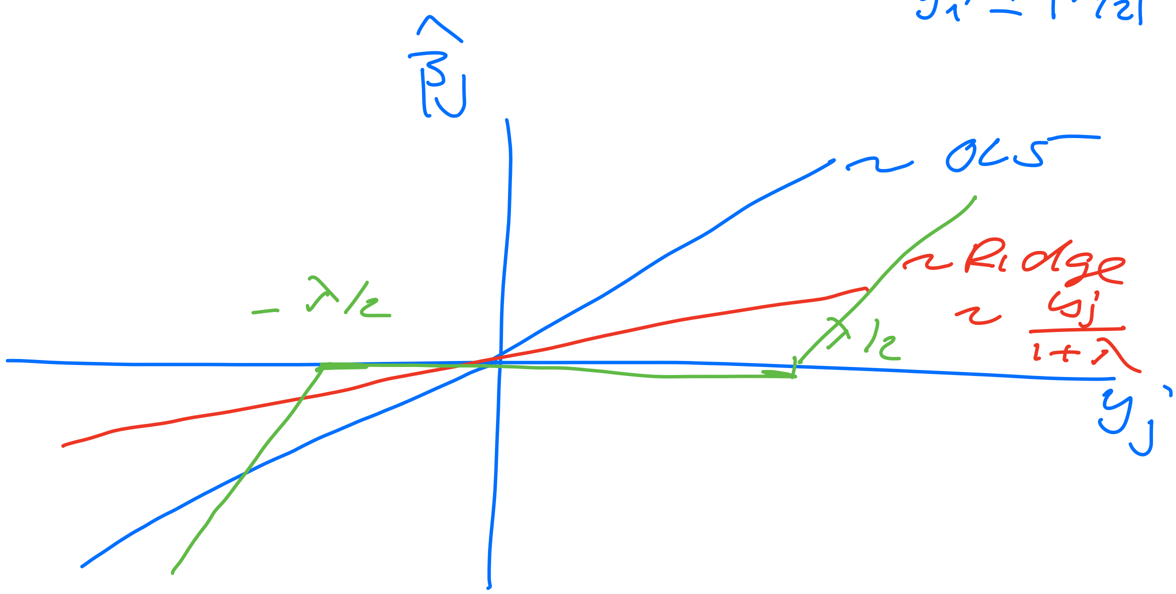
$$\frac{\partial C}{\partial \beta_j} = 0 \Rightarrow \hat{\beta}_j = \frac{y_j}{1+\lambda}$$

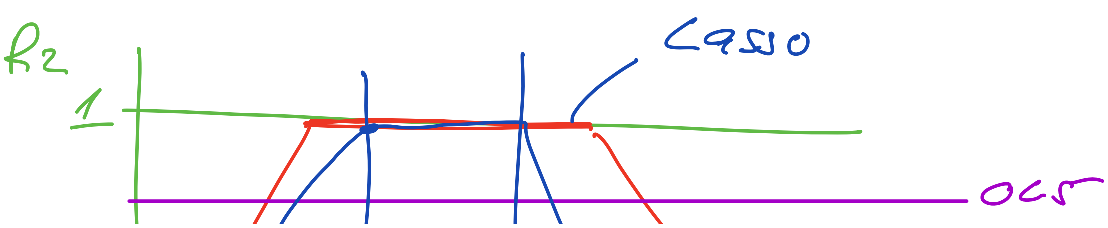
Lasso

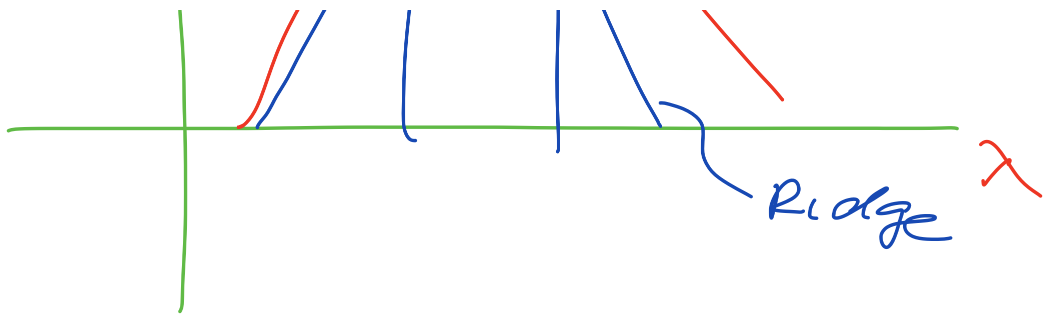
$$C(\beta) = \sum_{i=0}^{n-1} (y_i - \beta_i)^2 + \lambda \sum |\beta_j|$$

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

$$\hat{\beta}_j = \begin{cases} y_i - \lambda/2 & y_i > \lambda/2 \\ y_i + \lambda/2 & y_i < \lambda/2 \\ 0 & y_i \leq |\lambda/2| \end{cases}$$






Ridge
λ