HA-03

GROUP 2

MIRZOEVA SARA

(1) standardised regressors x and x y $x \sim N(0,1)$

y - centered

ridge reg.:

$$loss(\hat{\beta}) = (y - \hat{y})^T (y - \hat{y}) +$$

$$+ \lambda \hat{\beta}^T \hat{\beta}$$

regularization

$$\hat{y} = \hat{\beta}_1 x + \hat{\beta}_2 x = x(\hat{\beta}_1 + \hat{\beta}_2)$$

(a) optimal $\hat{\beta}_1, \hat{\beta}_2$

λ - fixed

$$\frac{\partial loss}{\partial \beta_1} = 0 \qquad \frac{\partial loss}{\partial \beta_2} = 0$$

$$loss(\hat{\beta}) = (y - x(\hat{\beta}_1 + \hat{\beta}_2))^T (y - x(\hat{\beta}_1 + \hat{\beta}_2)) + \lambda \hat{\beta}^T \hat{\beta} =$$

$$\hat{\beta} = \hat{\beta}_1 + \hat{\beta}_2$$

$$\hat{\beta}^T \hat{\beta} = (\hat{\beta}_1 + \hat{\beta}_2)^T (\hat{\beta}_1 + \hat{\beta}_2)$$

$$= (y - X(\hat{\beta}_1 + \hat{\beta}_2))$$
$$(y - X(\hat{\beta}_1 + \hat{\beta}_2)) +$$
$$+ \lambda (\hat{\beta}_1 + \hat{\beta}_2)^T (\hat{\beta}_1 + \hat{\beta}_2)$$

$$\beta^T \beta = ?$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_h \end{pmatrix} \qquad \beta^T = (\beta_1, \beta_2 \ldots \beta_h)$$

$[h \times 1]$ $\qquad$ $[1 \times h]$

$$\hat{\beta}^T \cdot \hat{\beta} = (\hat{\beta}_1 \; \hat{\beta}_2 \; \ldots \; \hat{\beta}_h) \cdot \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_h \end{pmatrix} = \hat{\beta}_1^2 + \hat{\beta}_2^2 + \ldots + \hat{\beta}_h^2$$

So, $(\hat\beta_1+\hat\beta_2)^T(\hat\beta_1+\hat\beta_2) =$

$$= \hat\beta_1^2 + \hat\beta_2^2 \;\therefore$$

$$= (y - X(\hat\beta_1+\hat\beta_2))^T (y - X(\hat\beta_1+\hat\beta_2)) +$$

$$+ \lambda(\hat\beta_1^2 + \beta_2^2)$$

$$(y - X(\hat\beta_1+\hat\beta_2))^T \cdot (y - X(\hat\beta_1+\hat\beta_2)) =$$

$$= y^T y - y^T (X(\hat\beta_1+\hat\beta_2)) -$$

$$- (X(\hat\beta_1+\hat\beta_2))^T \cdot y +$$

$$+ (X(\beta_1+\beta_2))^T X(\hat\beta_1+\hat\beta_2)$$

$$= y^T y - y^T (\hat\beta_1+\hat\beta_2)^T X^T -$$

$$- (\hat\beta_1 + \hat\beta_2) X^T y +$$

$$+ (\hat\beta_1 + \hat\beta_2)^T X^T X (\hat\beta_1 + \hat\beta_2)$$

$$= (\beta_1^2 + \beta_2^2)$$

$$= y^T y - y^T X (\hat\beta_1 + \hat\beta_2) -$$

$$- X^T y (\hat\beta_1 + \hat\beta_2) +$$

$$+ (\beta_1^2 + \beta_2^2) X^T X =$$

$$y^T X (\hat\beta_1 + \hat\beta_2) =$$

$$= X^T y (\hat\beta_1 + \hat\beta_2)$$

$$y^T X = (y_1 y_2 \cdot y_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$\shortparallel$$

$$X^T y = (x_1 x_2 \cdots x_n) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$= y^T y - 2(\hat{\beta}_1 + \hat{\beta}_2) X^T y + (\hat{\beta}_1 + \hat{\beta}_2)^2 X^T X$$

$$\text{loss}(\hat{\beta}_1, \hat{\beta}_2) = y^T y - 2(\hat{\beta}_1 + \hat{\beta}_2) X^T y + (\hat{\beta}_1 + \hat{\beta}_2)^2 X^T X + \lambda(\hat{\beta}_1^2 + \hat{\beta}_2^2)$$

$$\frac{\partial \text{loss}(\hat{\beta}_1, \hat{\beta}_2)}{\partial \hat{\beta}_1} = -2X^T y + 2(\hat{\beta}_1 + \hat{\beta}_2) X^T X + 2\lambda \hat{\beta}_1 = 0$$

$$-2X^T y + 2\hat{\beta}_1 X^T X +$$
$$+ 2\hat{\beta}_2 X^T X + 2\lambda\hat{\beta}_1 = 0$$

$$\frac{\partial \operatorname{loss}(\hat{\beta}_1, \hat{\beta}_2)}{\partial \hat{\beta}_2} =$$

$$= -2X^T y + 2(\hat{\beta}_1 + \hat{\beta}_2)X^T \cdot X + \lambda 2\hat{\beta}_2 =$$
$$= -2X^T y + 2\hat{\beta}_1 X^T X +$$
$$+ 2\hat{\beta}_2 X^T X + \lambda 2\hat{\beta}_2 =$$
$$= 0$$

$$\begin{cases} \dfrac{\partial \operatorname{loss}(\hat{\beta}_1, \hat{\beta}_2)}{\partial \hat{\beta}_1} = 0 \quad (1) \\ \dfrac{\partial \operatorname{loss}(\hat{\beta}_1, \hat{\beta}_2)}{\partial \hat{\beta}_2} = 0 \quad (2) \end{cases}$$

$(1) - (2) \rightarrow$

$$2\lambda\hat{\beta}_1 - 2\lambda\hat{\beta}_2 = 0$$

$$\hat{\beta}_1 = \hat{\beta}_2$$

$$-2X^T y + 2\hat{\beta}_1 X^T X +$$

$$+ 2\hat{\beta}_1 X^T X + \lambda 2\hat{\beta}_1 = 0$$

$$-2X^T y + 4\hat{\beta}_1 X^T X + \lambda 2\hat{\beta}_1 = 0$$

$$-X^T y + 2\hat{\beta}_1 X^T X + \lambda\hat{\beta}_1 = 0$$

$$\hat{\beta}_1 (2X^T X + \lambda) = X^T y$$

$$\hat{\beta}_1 = \frac{X^T y}{2X^T X + \lambda}$$

$$\boxed{\hat{\beta}_1 = \hat{\beta}_2 = \frac{X^T y}{2X^T X + \lambda}}$$

(b) $\lambda \to \infty$

$$\lim_{\lambda \to \infty} \hat{\beta}_1 = \lim_{\lambda \to \infty} \hat{\beta}_2 =$$

$$= \lim_{\lambda \to \infty} \frac{X^T y}{\underbrace{2X^T X + \lambda}_{\infty}} =$$

$$= 0$$

(c) $\lambda \to 0$

$\hat{\beta}_1 + \hat{\beta}_2 \to$

$\hat{\beta}_1 = \hat{\beta}_2$

$$\hat{\beta}_1 + \hat{\beta}_2 = \frac{X^T y}{2X^T X + \lambda} + \frac{X^T y}{2X^T X + \lambda} = \frac{2X^T y}{2X^T X + \lambda}$$

$$\lim_{\lambda \to 0} \hat{\beta}_1 + \hat{\beta}_2 =$$

$$= \lim_{\lambda \to 0} \frac{2x^T y}{2x^T x + \lambda} =$$

$$= \frac{2x^T y}{2x^T x} = \frac{x^T y}{x^T x}$$

② $y = X\beta + u$

β - non-random

$E(u|X) = 0$

X $\quad$ rank $X = k$

$[n \times k]$

$Var(u|X) = \sigma^2 W$

$W \neq I$

$\hat{\beta}$ standard OLS estimator

(a) $E(\hat{\beta}|X) = $

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

const

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

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

$$= (X^T X)^{-1} X^T E(X\beta + u|X) =$$

$$= (X^T X)^{-1} X^T \big(X\beta + \overbrace{E(u|X)}^{0}\big) =$$

$$= \underbrace{(X^T X)^{-1} X^T X}_{I}\beta =$$

$$= \beta$$

$$E(\hat{\beta}) = E(E(\hat{\beta}|X)) =$$

$$= E(\beta) = \beta$$

(b) $Var(\hat{\beta}|X) =$

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

$$= (X^T X)^{-1} X^T \operatorname{Var}(y|X) \cdot \left((X^T X)^{-1} X^T\right)^T =$$

$$= (X^T X)^{-1} X^T \operatorname{Var}(X\beta + u|X) \cdot X (X^T X)^{-1} =$$

$$= (X^T X)^{-1} X^T \operatorname{Var}(u|X) \cdot X (X^T X)^{-1} =$$

$$= (X^T X)^{-1} X^T \sigma^2 W \cdot X (X^T X)^{-1} =$$

$$= \sigma^2 (X^T X)^{-1} X^T W \cdot X (X^T X)^{-1}$$

(c) CI for β will not be valid

se not correct (because uses Var)

because $Var(u|X) = \sigma^2 W + \sigma^2 I$

not min var of unbiased est

$\hat\beta$ is unbiased, not efficient

(d) $Cov(y, \hat\beta | X) =$

$= Cov(X\beta + u, (X^T X)^{-1} X^T y | X) =$

$X\beta = 0$

$y = X\beta + u$

$= Cov(u, (X^T X)^{-1} X^T (X\beta + u) | X) =$

$= Cov(u, (X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T u | X) =$

$(X^T X)^{-1} X^T X \beta = 0$

$= Cov(u, (X^T X)^{-1} X^T u | X)$

$= (X^T X)^{-1} X^T Cov(u, u | X)$

$= X^{T T} (X^T X)^{-1}$ $\quad Var(u|X) =$

$$= (X^T X)^{-1} X^T \delta^2 w =$$

$$= \delta^2 w X (X^T X)^{-1}$$

3

$$X = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}$$

(a) $X^T X$ and diagonalize is

$$X^T = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 2 \cdot 2 + 1 + 1 & 2 \cdot 1 - 2 + 1 \cdot 1 \\ 1 \cdot 2 + 2(-1) + 1 \cdot 1 & 1 \cdot 1 + 2 \cdot 2 + 1 \cdot 1 \end{pmatrix} =$$

$$= \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$$

diagonalize $X^T X$

$$\det(X^T X - \lambda I) = 0$$

$$\det \begin{pmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{pmatrix} =$$

$$= (6-\lambda)^2 - 1 = 0$$

$$36 - 12\lambda + \lambda^2 - 1 = 0$$

$$\lambda^2 - 12\lambda + 35 = 0$$

$$D = 144 - 4 \cdot 35 \cdot 1 =$$

$$= 4 \qquad \lambda_{1,2} = \frac{12 \pm 2}{2}$$

$$\lambda_1 = 5 \qquad \lambda_2 = 7.$$

$\lambda_1 = 5$

$$\begin{pmatrix} 6-5 & 1 \\ 1 & 6-5 \end{pmatrix} v_1 =$$

$$= \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} v_1 = 0$$

$$v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$\lambda_2 = 7$ $\qquad D = \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix}$

$$v_2 \begin{pmatrix} 6-7 & 1 \\ 1 & 6-7 \end{pmatrix} = v_2 \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

(b) SVD of X

$$SVD = \underset{[n\times n]}{U} \cdot \underset{[n\times h]}{D} \cdot \underset{[h\times h]}{V^T}$$

$$X X^T = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}.$$

X is standardized
for this let's use
python code
Code is provided here

$X_{standardized} =$

D is diagonal.

$$D = \begin{pmatrix} \sqrt{\lambda_2} & 0 \\ 0 & \sqrt{\lambda_1} \end{pmatrix} =$$

$$= \begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \end{pmatrix}$$

$$V^T = \begin{pmatrix} \frac{v_{11}}{\sqrt{2}} & \frac{v_{12}}{\sqrt{2}} \\ \frac{v_{21}}{\sqrt{2}} & \frac{v_{22}}{\sqrt{2}} \end{pmatrix} =$$

$\|v_1\| = \sqrt{1^2+1^2} = \sqrt{2}$

$\|v_2\| = \sqrt{1^2+(-1)^2} = \sqrt{2}$

$$= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

$$U = \begin{pmatrix} \frac{1}{\sqrt{\lambda_2}} & 0 \\ 0 & \frac{1}{\sqrt{\lambda_1}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{7}} & 0 \\ 0 & \frac{1}{\sqrt{5}} \end{pmatrix}$$

$SUD = UDV^T$

$$X^T X = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix} =$$

$$= \underbrace{\begin{pmatrix} \frac{1}{\sqrt{7}} & 0 \\ 0 & \frac{1}{\sqrt{5}} \end{pmatrix}}_{U} \cdot \underbrace{\begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \end{pmatrix}}_{D} \cdot \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}}_{V^T}$$

(c) best approxim to X with rank = 1

keep largest → others 0

$$D = \begin{pmatrix} \sqrt{7} & 0 \\ 0 & 0 \end{pmatrix}$$

$$X_{standard} = \frac{X - \bar{X}}{X_{std}}$$

$X_{approx} \approx X_i \cdot X_{std} + \bar{X}$

SVD $= UDV^T$

4

SVD of $X = UDV^T$

D positive from highest to lowest

$d_{11} > d_{22} > \ldots > 0$

$\|Xw\|^2 \to \max_w$ s.t. $\|w\|^2 = 1$

(d) Lagrangian

$$\begin{cases} \|Xw\|^2 \to \max\limits_{w} \\ \text{s.t} \\ \|w\|^2 = 1 \end{cases}$$

$$L = \|Xw\|^2 - \lambda(\|w\|^2 - 1) =$$

$$= w^T X^T \cdot X \cdot w - \lambda(w^T \cdot w - 1)$$

(e) F.O.C

$$\frac{\partial L}{\partial w} = (w^T X^T \cdot Xw)'_w - \lambda(w^T w - 1)'_w =$$

$$= 2X^T X w - 2\lambda w = 0$$

$$2X^T X w = 2\lambda w$$

$$\lambda = X^T X$$

(f) optimal w of columns of V

V - ?

SVD

$$X^T \cdot X = (UDV^T)^T \cdot UDV^T = V^{TT} \cdot D^T \cdot \underbrace{U^T \cdot U}_{I} \cdot DV^T =$$

$$= V D^T D V^T$$

$$2 X^T X w = \lambda w$$

$$V D^T D V^T w = \lambda w$$

$$(V D^T D \cdot V^T) w V^T = \lambda V^T w$$

$$D^T D U^T w = \lambda U^T w$$

$$\begin{pmatrix} \delta^2 & 0 & 0 & 0 \\ & \delta^2 & 0 & 0 \\ 0 & & \ddots & \\ 0 & & & \ddots \\ 0 & 0 & 0 & \delta^2 \end{pmatrix} U^T w = \lambda U^T w$$

$$\delta_i \cdot U^T w = \lambda U^T w$$

$U^T w = 0$ or $\delta_i = \lambda$

optimal w is the first col of U