HA #3 Gatsalova Eva

1. We have two absolutely identical preliminary standardized regressors x and x. The dependent variable y is centered.

   In the ridge regression one minimizes the loss function ← regularization element, $\lambda = 0 \to$ OLS

$$\text{loss}(\hat\beta) = (y-\hat y)^T(y-\hat y) + \lambda\underbrace{\hat\beta^T\hat\beta}_{\beta^2}, \quad \hat y = \hat\beta_1 x + \hat\beta_2 x. = (\hat\beta_1 + \hat\beta_2)\cdot x$$

   (a) Find the optimal $\hat\beta_1$ and $\hat\beta_2$ for fixed λ.

   (b) What happens to the estimates when $\lambda \to \infty$?

   (c) What happens to the sum $\hat\beta_1 + \hat\beta_2$ when $\lambda \to 0$?

a)

$$\text{loss}(\hat\beta) = (y - (\hat\beta_1 + \hat\beta_2)x)^T(y - (\hat\beta_1 + \hat\beta_2)x) + \lambda(\hat\beta_1^2 + \hat\beta_2^2)$$

$$(1)\ \frac{\partial L}{\partial \hat\beta_1} = \underbrace{2\cdot(-x)^T\cdot(y - (\hat\beta_1 + \hat\beta_2)x)}_{\Updownarrow \text{ same}} + 2\lambda\hat\beta_1 = 0$$

$$(2)\ \frac{\partial L}{\partial \hat\beta_2} = 2(-x)^T\cdot(y - (\hat\beta_1 + \hat\beta_2)x) + 2\lambda\hat\beta_2 = 0$$

(the function is symmetrical, so the result for $\hat\beta_2$ would be symmetrical too) $\Rightarrow \lambda\hat\beta_2 = \lambda\hat\beta_1 \Rightarrow$

$\hat\beta_2 = \hat\beta_1 \Rightarrow$ substituting into (1)

$$2(-x)^T(y - 2\hat\beta_1 x) + 2\lambda\hat\beta_1 = 0$$

$$-2x^Ty + 4x^T\cdot x\hat\beta_1 + 2\lambda\hat\beta_1$$

$$(2\lambda + 4x^T\cdot x)\hat\beta_1 = 2x^Ty$$

$$\hat\beta_1 = 2(\lambda + 2x^T\cdot x)^{-1}2x^Ty = \hat\beta_2$$

b) $\lim\limits_{\lambda\to\infty}(\lambda + 2x^T\cdot x)^{-1}x^Ty = 0 \longrightarrow \hat\beta_1 \xrightarrow[\lambda\to\infty]{} 0$, $\hat\beta_2 \xrightarrow[\lambda\to\infty]{} 0$

λ is a regularization element, which penalizes large coefficients, so it is logical that the $\lambda\to\infty$ would lead to decreasing coefficients.

c) $\lim\limits_{\lambda\to 0}(\lambda + 2x^T\cdot x)^{-1}x^Ty = \hat\beta_{OLS} \to$ so the sum would be $\hat\beta_1^{OLS} + \hat\beta_2^{OLS}$

As $\lambda\to 0$, the function would be as if without it, resulting in OLS estimators.

2. Consider the model $y = X\beta + u$ where β is non-random, $\mathbb{E}(u \mid X) = 0$, the matrix X of size $n \times k$ has rank $X = k$, but $\mathbb{V}\mathrm{ar}(u \mid X) = \sigma^2 W$ with $W \neq I$. Let $\hat{\beta}$ be the standard OLS estimator of β.

(a) Find $\mathbb{E}(\hat{\beta} \mid X)$, $\mathbb{E}(\hat{\beta})$.

(b) Find $\mathbb{V}\mathrm{ar}(\hat{\beta} \mid X)$.

(c) How do you think, will the standard confidence interval for β be valid in this case?

(d) Find $\mathbb{C}\mathrm{ov}(y, \hat{\beta} \mid X)$.

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

a) $E(\hat{\beta} | X) = E((X^T X)^{-1} X^T y \,|\, X) = (X^T X)^{-1} X^T E(X\beta + u \,|\, X) = \overbrace{(X^T X)^{-1} X^T \cdot X}^{I}\beta + 0 = \underline{\beta}$

$E(\hat{\beta}) = E((X^T X)^{-1} X^T y) = (X^T X)^{-1} X^T E(X\beta + u) = \underline{(X^T X)^{-1} X^T \cdot X \cdot \beta + (X^T X)^{-1} X^T \cdot E(u)}$

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

b) $Var(\hat{\beta} | X) = Var((X^T X)^{-1} X^T \cdot y | X) = (X^T X)^{-1} X^T \, Var(y|X) ((X^T X)^{-1} X^T)^T = (X^T X)^{-1} X^T \cdot \sigma^2 W \cdot X(X^T X)^{-1} = \underline{\sigma^2 (X^T X)^{-1} X^T W X (X^T X)^{-1}}$

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

(T)

c) No, the standard CI will be invalid, since error terms are not constant but heteroscedastic as they vary across the sample $\Rightarrow$
$\Leftrightarrow var(u|X) = \sigma^2 W$

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

d) $Cov(y, \hat{\beta} | X) = Cov(y, (X^T X)^{-1} X^T y | X) = Cov(y; y | X) ((X^T X)^{-1} X^T)^T \,(=)$

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

(T)

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

3. Consider the matrix

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

(a) Find the matrix $X^T X$ and diagonalize it.

a) $X^T = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}$, $A = X^T \cdot X = \underset{2\times 3}{\begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}} \underset{3\times 2}{\begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}} = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$, $\det(A) = 35$

① ↳ need to find eigenvalues of A:

$$\begin{pmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{pmatrix} = (6-\lambda)^2 - 1 = 36 - 12\lambda + \lambda^2 - 1 = \lambda^2 - 12\lambda + 35 = 0 \rightarrow \begin{cases} \lambda_1 = 5 \\ \lambda_2 = 7 \end{cases}$$

② ↳ find eigenvector: $\lambda = 5$:

$$(A - \lambda E)x = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} x = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} x = 0 \Rightarrow x_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\lambda = 7: (A - \lambda E)x = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} x = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} x = 0 \Rightarrow x_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

↳ Diagonalizing: $X^{-1} A X = D$

$$\begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} -5 & 5 \\ 7 & 7 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 10 & 0 \\ 0 & 14 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 7 \end{pmatrix}$$

(b) Find the SVD of X.

$X = \underset{3\times 2}{\begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}} \Rightarrow X = UDV^T$, where $U^T U = I$ [3×3], $U \rightarrow$ columns are eigen vectors of XX^T

$V^T V = I$ [2×2], $V \rightarrow$ columns are eigen.v-s of $X^T X$

D-diagonal $= \begin{pmatrix} 5 & 0 \\ 0 & 7 \end{pmatrix}$ [2×2]

① $XX^T = \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix}$

② $\lambda_1 = 7 \rightarrow x_1 = \begin{pmatrix} 3/2 \\ 1/2 \\ 1 \end{pmatrix}$

$\lambda_2 = 5 \rightarrow x_2 = \begin{pmatrix} -1/3 \\ 1 \\ 0 \end{pmatrix}$

$\lambda_3 = 0 \rightarrow x_3 = \begin{pmatrix} -3/5 \\ -1/5 \\ 1 \end{pmatrix}$

③ Normalizing:

$\Rightarrow x_1 = \sqrt{\left(\frac{3}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + 1} = \frac{\sqrt{14}}{2}$

$x_2 = \sqrt{\frac{1}{9} + \frac{9}{9}} = \frac{\sqrt{10}}{3}$

$x_3 = \sqrt{\frac{9+1+25}{25}} = \frac{\sqrt{35}}{5}$

④ Vectors (devided by normal-zed)

$\rightarrow n_1 = (3/\sqrt{14}, 1/\sqrt{14}, 2/\sqrt{14})^T$

$\rightarrow n_2 = (-1/\sqrt{10}, 3/\sqrt{10}, 0)^T$

$\rightarrow n_3 = (-3/\sqrt{35}, -1/\sqrt{35}, 5/\sqrt{35})^T$

⑤ Normalizing for V:

$$\begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \rightarrow : \sqrt{2} \Rightarrow \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

⑥ $D = \begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{pmatrix}$

⑦ Finally, SVD:

$$\begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} = \underset{\boxed{3\times 3}}{\begin{pmatrix} 3/\sqrt{14} & -1/\sqrt{10} & -3/\sqrt{35} \\ 1/\sqrt{14} & 3/\sqrt{10} & -1/\sqrt{35} \\ 2/\sqrt{14} & 0 & 5/\sqrt{35} \end{pmatrix}} \underset{\boxed{3\times 2}}{\begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{pmatrix}} \underset{\boxed{2\times 2}}{\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}}$$

c) Best approxim-n to X with $rk = 1$

Again, SVD: $\begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} = \underbrace{\overset{U}{\begin{pmatrix} 3/\sqrt{14} & -1/\sqrt{10} & -3/\sqrt{35} \\ 1/\sqrt{14} & 3/\sqrt{10} & -1/\sqrt{35} \\ 2/\sqrt{14} & 0 & 5/\sqrt{35} \end{pmatrix}}}_{\boxed{3\times 3}} \underbrace{\overset{D}{\begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{pmatrix}}}_{\boxed{3\times 2}} \underbrace{\overset{V^T}{\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}}}_{\boxed{2\times 2}}$

$k = 1$

$V_1^T = (1/\sqrt{2}\,;\ 1/\sqrt{2})$ (keep first k rows)

$U_1 = \begin{pmatrix} 3/\sqrt{14} \\ 1/\sqrt{14} \\ 2/\sqrt{14} \end{pmatrix}$ (keep first k columns)

$D_1 = (\sqrt{7})$ (keep first k rows and columns)

$\Rightarrow$ Approximation is then:

$X_1 = \underbrace{\begin{pmatrix} 3/\sqrt{14} \\ 1/\sqrt{14} \\ 2/\sqrt{14} \end{pmatrix}}_{3\times 1} \underbrace{(\sqrt{7})}_{1\times 1} (1/\sqrt{2}\,;\ 1/\sqrt{2})$

$\Rightarrow \begin{pmatrix} 3/2 & 3/2 \\ 1/2 & 1/2 \\ 1 & 1 \end{pmatrix} \Rightarrow (1\,;\ 1)$ $\underline{rk = 1}$

4. The columns of X are standardized. You know the SVD of the matrix $X = UDV^T$. The diagonal elements of D are positive and ordered from highest to lowest, $d_{11} > d_{22} > \cdots > 0$. → unique

Let's maximize $\|Xw\|^2$ by choosing an optimal vector w subject to $\|w\|^2 = 1$.

(d) Write the Lagrangian function for this problem.

(e) Find the first order conditions. Differential is your friend!

(f) Find the optimal w in terms of columns of V.

Hint: one may interpret the FOC in terms of eigenvalues and eigenvectors!

d) $X = UDV^T$

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

$\|w\|\cdot\|w\| = w^T w$

$\underline{\mathcal{L}(w,\lambda) = \|Xw\|^2 - \lambda(\|w\|^2 - 1)}$

$\|Xw\|^2 = \|Xw\|\,\|Xw\| = \sqrt{Xw}\,\sqrt{Xw} = (Xw)^T Xw$

$\mathcal{L} = w^T X^T X w - \lambda(w^T w - 1)$

$\underline{\underline{w^T X^T X w}}$

e) FOC:

$d(\underbrace{w^T X^T X w}_{①} - \lambda(w^T w)) \ ⊜$

① ↳ $d(w^T X^T X w) = dw^T (X^T X w) + (w^T) d\, X^T X w = dw^T \cdot (X^T X w) + w^T(dX^T X \cdot (w) + X^T X\, dw)) =$ constants

$= dw^T (X^T X w) + w^T \cdot X^T X\, dw = \underline{2w^T (X^T X)\, dw} \Rightarrow \underbrace{w^T}_{\text{eigenvector}} X^T X \to \underbrace{\lambda}_{\text{eigenvalue}} w$

$(X^T X w)^T = w^T X^T X$

Using: $d(v^T v) = d(v^T)\cdot v + v^T dv = 2d(v^T)\cdot v = 2d(v)\cdot v^T$

⊜ $2w^T(X^T X)dw - \lambda d(w^T w) = 2w^T(X^T X)dw - 2\lambda\, dw \cdot w^T = \underline{2w^T(X^T X - \lambda I)dw}$

↳ $\lambda(dw^T \cdot w + w^T dw) = 2\,d(w)\cdot w^T$

f) $X = UDV^T$

$\|XW\|^2 = W^T X^T X W = W^T \Lambda W = \lambda w^T w \to \max\limits_{w}$

W_{max} should be the biggest eigenvalue, so the $W_{max} = V$, where

V is normalized matrix of eigenvectors of $x^T x$