Svetlana Telnova 2 group

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

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

   (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 $\lambda \to 0$?

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

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

$$\frac{\partial(\lambda \hat{\beta}^T \hat{\beta})}{\partial \hat{\beta}} = \frac{\partial \lambda \|\hat{\beta}\|^2}{\partial \hat{\beta}} = 2\lambda \hat{\beta} \qquad \beta_{2\times 1} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

$$X\hat{\beta} = \begin{bmatrix} x_1 & x_1 \\ x_2 & x_2 \\ \vdots & \vdots \\ x_n & x_n \end{bmatrix}_{[n\times 2]} \times \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}_{[2\times 1]} = \begin{bmatrix} x_1\beta_1 + x_1\beta_2 \\ x_2\beta_1 + x_2\beta_2 \\ \vdots \\ x_n\beta_1 + x_n\beta_2 \end{bmatrix}_{[n\times 1]}$$ which is equivalent to

$$\hat{y} = \hat{\beta}_1 X + \hat{\beta}_2 X \longrightarrow \text{can represent } \hat{y} = X\hat{\beta}.$$

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

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

$[1\times n][n\times 2][2\times n]$ $\quad [1\times 2][2\times n][n\times 1]$

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

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

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

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

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

$$\hat{\beta} = (X^T X + \lambda)^{-1} X^T y \qquad \hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X^T X + \lambda)^{-1} X^T y.$$

$$loss(\hat{\beta}) = (y-\hat{y})^T(y-\hat{y}) + \lambda \hat{\beta}^T\hat{\beta} = (y-(\hat{\beta}_1+\hat{\beta}_2)x)^T(y-(\beta_1+\beta_2)x) +$$
$$+\lambda(\hat{\beta}_1^2+\hat{\beta}_2^2) = \left|(y-(\hat{\beta}_1+\hat{\beta}_2)x)\right|^2 + \lambda(\hat{\beta}_1^2+\hat{\beta}_2^2)$$

$$\begin{cases} \frac{\partial loss}{\partial \beta_1} = 2(-x^T)(y-(\hat{\beta}_1+\hat{\beta}_2)x) + 2\lambda\hat{\beta}_1 = 0 \\ \frac{\partial loss}{\partial \beta_2} = 2(-x^T)(y-(\hat{\beta}_1+\hat{\beta}_2)x) + 2\lambda\hat{\beta}_2 = 0 \end{cases}$$

The equations are symmetric $\Rightarrow \hat{\beta}^* = \hat{\beta}_1 = \hat{\beta}_2$.

$$-2x^T(y-2\hat{\beta}^*x) + 2\lambda\hat{\beta}^* = 0.$$
$$-2x^Ty + 4x^Tx\hat{\beta}^* + 2\lambda\hat{\beta}^* = 0.$$
$$(2x^Tx+\lambda)\hat{\beta}^* = x^Ty$$
$$\hat{\beta}^* = (2x^Tx+\lambda)^{-1}x^Ty = \hat{\beta}_1 = \hat{\beta}_2 \text{ - optimal } \hat{\beta}_1 \text{ and } \hat{\beta}_2 \text{ for fixed } \lambda.$$

b) $\lim\limits_{\lambda\to\infty} \underset{[1\times2]\,[2\times1]}{(2x^Tx+\lambda)^{-1}} \cdot x^Ty = 0. \Rightarrow (\hat{\beta}_1;\hat{\beta}_2) = (0;0)$

c) $\lim\limits_{\lambda\to0} 2\underset{[1\times2]\,[2\times1]}{(2x^Tx+\lambda)^{-1}} \cdot x^Ty = \hat{\beta}^* \Rightarrow$ The sum of estimators converges to the optimal estimator $\hat{\beta}^*$

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

$y = X\beta + u \quad E(u \mid X) = 0. \quad X_{n \times k} \quad E(X\beta \mid X)$

$\mathrm{Var}(u \mid X) = \sigma^2 W \quad W \neq 1$

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

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

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

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

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

B) $\mathrm{Var}(\hat{\beta} \mid X) = \mathrm{Var}((X^T X)^{-1} X^T y \mid X) = (X^T X)^{-1} \cdot X^T \mathrm{Var}(y \mid X) \cdot [(X^T X)^{-1} \cdot X^T]^T =$ (1)

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

(1) $\mathrm{Var}(y \mid X) = \mathrm{Var}(\underbrace{X\beta}_{\text{const}} + u \mid X) = 0 + \mathrm{Var}(u \mid X) \cdot X^T = \sigma^2 W$

c) No, the CI will be invalid since it is based on assumption that $\mathrm{Var}(u \mid X) = \sigma^2 I$. In our case $\mathrm{Var}(u \mid X) = \sigma^2 W$ $(W \neq I)$ indicates heteroscedasciby or correlation among errors. Thus, Gauss Markov conditions are violated, the standard estimators and CI are invalid

d) $\mathrm{Cov}(y, \hat{\beta} \mid X) = \mathrm{Cov}(y, (X^T X)^{-1} X^T y \mid X) = \mathrm{Cov}(y, y \mid X)[(X^T X)^{-1} X^T]^T =$

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

$\mathrm{Var}(y \mid X) = \mathrm{Var}(\underbrace{X\beta}_{\text{const}} + u \mid X) = \mathrm{Var}(u \mid X) = \sigma^2 W$

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.

(b) Find the SVD of X.

(c) Find the best approximation to X with rank equal to 1.

Remark: in the principal component analysis the variables in the matrix X should be standardized. If you can't do this by bare hands, feel free to use python, but provide code!

a) $X^T \times X = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \times \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$

$\begin{vmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{vmatrix} = 0$

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

$(6-\lambda-1)(6-\lambda+1) = 0$

$\lambda = 5 \quad \lambda = 7.$

① $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$ $\quad x_1 + x_2 = 0 \quad \begin{pmatrix} x_1 \\ -x_1 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

$x_1 = -x_2$

② $\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$ $\quad -x_1 + x_2 = 0 \quad \begin{pmatrix} x_1 \\ x_1 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$x_1 = x$

$A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$

$A^{-1} = \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ -1 & 1 & 0 & 1 \end{array}\right) \to \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 2 & 1 & 1 \end{array}\right) \to \left(\begin{array}{cc|cc} 2 & 0 & 1 & -1 \\ 0 & 2 & 1 & 1 \end{array}\right) \to \left(\begin{array}{cc|cc} 1 & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 & \frac{1}{2} & \frac{1}{2} \end{array}\right)$

$A \cdot A^{-1} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \times \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$D = \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix}$ $\quad \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \times \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix} \times \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$

Rectangular diagonal matrix:

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

в) $X \cdot X^T = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} \times \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix}$

$$\begin{vmatrix} 5-\lambda & 0 & 3 \\ 0 & 5-\lambda & 1 \\ 3 & 1 & 2-\lambda \end{vmatrix} = 0$$

$(5-\lambda)^2(2-\lambda) + 0 + 0 - (9(5-\lambda) + (5-\lambda)) = 0$

$(5-\lambda)^2(2-\lambda) - 10(5-\lambda) = 0$

$(5-\lambda)\left((5-\lambda)(2-\lambda) - 10\right) = 0$

$10 - 2\lambda - 5\lambda + \lambda^2 - 10 = 0$

$\lambda^2 - 7\lambda = 0$

$\lambda(\lambda - 7) = 0 \quad \lambda = 0 \quad \lambda = 7,$

① $\begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ $\lambda_3 = 5.$

$5x_1 + 3x_3 = 0.$

$x_1 = -\frac{3x_3}{5}$

$5x_2 + x_3 = 0. \Rightarrow x_2 = -\frac{x_3}{5}$

$v_3 = \begin{pmatrix} -\frac{3}{5} \\ -\frac{1}{5} \\ 1 \end{pmatrix}$

② $\begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 1 \\ 3 & 1 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \bar{0}$

$3x_3 = 0 \Rightarrow x_3 = 0.$

$3x_1 + x_2 - 3x_3 = 0.$

$3x_1 + x_2 = 0. \quad x_1 = -\frac{x_2}{3}$

$v_2 = \begin{pmatrix} -\frac{1}{3} \\ 1 \\ 0 \end{pmatrix}$

$\begin{pmatrix} -2 & 0 & 3 \\ 0 & -2 & 1 \\ 3 & 1 & -5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \bar{0}.$

$-2x_1 + 3x_3 = 0 \quad x_1 = \frac{3x_3}{2}$

$-2x_2 + x_3 = 0. \quad x_2 = \frac{x_3}{2}$

$v_1 = \begin{pmatrix} \frac{3}{2} \\ \frac{1}{2} \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$

Normalizing vectors:

$|v_1| = \sqrt{\left(\frac{3}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + 1} = \frac{\sqrt{14}}{2}$ $v_1 = \begin{pmatrix} \frac{3}{\sqrt{14}} \\ \frac{1}{\sqrt{14}} \\ \frac{2}{\sqrt{14}} \end{pmatrix}$

$|v_2| = \sqrt{\frac{1}{9} + 1} = \frac{\sqrt{10}}{3}$

$|v_3| = \sqrt{\frac{9}{25} + \frac{1}{25} + 1} = \frac{\sqrt{35}}{5}.$

$$v_2 = \begin{pmatrix} -\frac{1}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} \\ 0 \end{pmatrix} \qquad v_1 = \begin{pmatrix} -\frac{3}{\sqrt{35}} \\ -\frac{1}{\sqrt{35}} \\ \frac{5}{\sqrt{35}} \end{pmatrix}$$

Normalizing: $\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \to \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$

length $\sqrt{2}$, length $\sqrt{2}$

$$X = \underbrace{\begin{pmatrix} \frac{3}{\sqrt{14}} & -\frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{35}} \\ \frac{1}{\sqrt{14}} & \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{35}} \\ \frac{2}{\sqrt{14}} & 0 & \frac{5}{\sqrt{35}} \end{pmatrix}}_{U} \times \underbrace{\begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{pmatrix}}_{D} \times \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{V^T} = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} \Rightarrow$$

U: $[3\times 3]$, $n \times n$ — D: $[3\times 2]$, $n \times m$ — V^T: 2×2

SVD decomposition

$$\begin{pmatrix} \frac{3}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{3}{\sqrt{2}} \\ \frac{2}{\sqrt{2}} & 0 \end{pmatrix} \times \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}$$

c) $A_k = U_k S_k V_k^T$

U_k - keep the first k columns

S_k - keep only the top k singular val

V_k^T - keep the first k rows

$A = U \times S \times V^T$

$[k\times 1]$ $[n\times k]$ $k\times k$ $k\times m$

$$\underbrace{\begin{pmatrix} \frac{3}{\sqrt{14}} \\ \frac{1}{\sqrt{14}} \\ \frac{2}{\sqrt{14}} \end{pmatrix}}_{[3\times]} \times \underbrace{(\sqrt{7})}_{[1\times 1]} \times \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{[1\times 2]} = \underbrace{\begin{pmatrix} \frac{3}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{2}} \end{pmatrix}}_{[3\times 1]} \times \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{[1\times 2]} = \begin{pmatrix} \frac{3}{2} & \frac{3}{2} \\ \frac{1}{2} & \frac{1}{2} \\ 1 & 1 \end{pmatrix}$$

$\Rightarrow$ the matrix approximation of X with rank 1 is $(1;1)$

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, $d_{11} > d_{22} > \cdots > 0$.

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

$\|Xw\|^2 = (Xw)^T(Xw) = w^T X^T X w$

$\|w\|^2 = w^T w$

$\Rightarrow \begin{cases} \max\limits_{w} \; w^T X^T X w \\ \text{s.t. } w^T w = 1. \end{cases}$

$L = \|Xw\|^2 - \lambda\left(\|w\|^2 - 1\right) = w^T X^T X w - \lambda\left(w^T w - 1\right)$

e) Denote $X^T X$ as A $\quad (w_1\ w_2 \dots w_n) \times 1$

F.O.C $\begin{cases} \frac{\partial L}{\partial w} = 2X^T X w - 2\lambda w = 0. \Rightarrow \underbrace{X^T X}_{\text{matrix}} w = \underbrace{\lambda}_{\text{eigenvalue}} w \quad (1) \\ \frac{\partial L}{\partial \lambda} = -w^T w - 1 = 0. \Rightarrow w^T w = 1 \quad (2) \end{cases}$

f) From (1) w is an eigenvector of $X^T X \Rightarrow$

$\|Xw\|^2 = w^T X^T X w = w^T \lambda w = \lambda w^T w = \lambda \to \max\limits_{w} \Rightarrow$

the eigenvector of w should correspond to the maximum eigenvalue V is the normalized matrix of eigenvectors of $X^T X$ matrix. Thus, the optimal value of w corresponds to the maximum eigenvalue of matrix V