Supranenko, HA 3, gr. 3

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 $\lambda \to 0$?

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 $\mathbb{V}\text{ar}(u \mid X) = \sigma^2 W$ 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 $\mathbb{V}\text{ar}(\hat{\beta} \mid X)$.

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

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

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!

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!

g_1

$$\hat{y} = \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 = X\hat{\beta}$$

$$L = \text{loss}(\hat{\beta}) = (y-\hat{y})^T (y-\hat{y})^T + \lambda \cdot \hat{\beta}^T \cdot \hat{\beta} \quad , \lambda \text{ is const}$$

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

$$dL = d(y - X\hat{\beta})^T \cdot (y - X\hat{\beta}) + (y - X\hat{\beta})^T \cdot d(y - X\hat{\beta}) +$$
$$+ \lambda \left(\underbrace{d\hat{\beta}^T}_{[1\times k]} \cdot \underbrace{\hat{\beta}}_{[k\times 1]} + \underbrace{\hat{\beta}^T}_{[1\times k]} \cdot \underbrace{d\hat{\beta}}_{[k\times 1]} \right)$$

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

$$dL = \underbrace{d\hat{\beta}^T}_{[1\times k]} \cdot \underbrace{(-X^T)}_{[k\times n]} \cdot \underbrace{(y - \hat{y})}_{[n\times 1]} + \underbrace{(y - \hat{y})^T}_{[1\times n]} \cdot \underbrace{(-X)}_{[n\times k]} \cdot \underbrace{d\hat{\beta}}_{[k\times 1]} +$$
$$+ 2\lambda \cdot \hat{\beta}^T \cdot d\hat{\beta}$$

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

$$\left(\lambda \cdot \hat{\beta}^T - (y - X\hat{\beta})^T \cdot X\right) \cdot d\hat{\beta} = 0 \qquad \forall d\hat{\beta}$$

$$\left(\lambda \cdot \hat{\beta}^T - (y - X\hat{\beta})^T \cdot X\right) = 0$$

$$\left(\lambda\cdot\hat{\beta}^T-\left(y-X\hat{\beta}\right)^T\cdot X\right)^T=0^T$$

$$\lambda\hat{\beta}-X^T\cdot\left(y-X\hat{\beta}\right)=0$$

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

$$\underbrace{\lambda\hat{\beta}}_{k\times 1}+\underbrace{\underset{k\times n}{X^T}\cdot\underset{n\times k}{X}\cdot\underset{k\times 1}{\hat{\beta}}}_{k\times 1}=\underbrace{\underset{k\times n}{X^T}\cdot\underset{n\times 1}{y}}_{k\times 1}$$

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

$$\hat{\beta}=\left(I\lambda+X^T\cdot X\right)^{-1}\cdot X^T\cdot y\ ,\quad \text{if } \left(I\lambda+X^T\cdot X\right)^{-1} \text{ exists}$$

$$\text{if } L=\sum_i\left(y_i-x_i\left(\hat{\beta}_1+\hat{\beta}_2\right)\right)^2+\lambda\left(\hat{\beta}_1^2+\hat{\beta}_2^2\right)\rightarrow\min_{\hat{\beta}_1;\hat{\beta}_2}$$

$$\frac{\partial L}{\partial\hat{\beta}_1}=-2\sum x_i\left(y_i-x_i\left(\hat{\beta}_1+\hat{\beta}_2\right)\right)+2\lambda\hat{\beta}_1$$

$$-\sum x_iy_i+\sum x_i^2\left(\hat{\beta}_1+\hat{\beta}_2\right)+\lambda\hat{\beta}_1=0$$

$$\frac{\partial L}{\partial\hat{\beta}_2}:-\sum x_iy_i+\sum x_i^2\left(\hat{\beta}_1+\hat{\beta}_2\right)+\lambda\hat{\beta}_2=0$$

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

$$-\sum x_iy_i+\sum x_i^2\cdot 2\hat{\beta}+\lambda\hat{\beta}$$

$$\Rightarrow \hat{\beta}\left(2\sum x_i^2+\lambda\right)=\sum x_i y_i$$

$$\boxed{\hat{\beta}=\frac{\sum x_i y_i}{2\sum x_i^2+\lambda}} \quad =\hat{\beta}_1=\hat{\beta}_2$$

b) $\lim\limits_{\lambda\to\infty}\hat{\beta}=0$

c) $\lim\limits_{\lambda\to 0}\left(\hat{\beta}_1+\hat{\beta}_2\right)=\frac{\sum x_i y_i}{2\sum x_i^2}\cdot 2=\frac{\sum x_i y_i}{\sum x_i^2}\rightarrow \hat{\beta}$ for regular OLS

Q2

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

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

$$E(\hat{\beta}|X)=E\left(\left(X^TX\right)^{-1}\cdot X^T(X\beta+u)|X\right)=$$

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

2. $E(\hat{\beta})=E\left(E(\hat{\beta}|X)\right)=\beta$

b) $Var(\hat{\beta}|X)=Var\left(\left(X^TX\right)^{-1}\cdot X^T\cdot(X\beta+u)|X\right)=$

$$=Var\left(\left(X^TX\right)^{-1}\cdot X^T\cdot u\,|X\right)=$$

$$=\left(X^TX\right)^{-1}\cdot X^T\cdot X\cdot\left(X^TX\right)^{-1}\cdot Var(u|X)=\left(X^T\cdot X\right)^{-1}\cdot\sigma^2\cdot W$$

c) standart confidence intervals for $\hat{\beta}$ would not be valid in this case, as they assume $W = I$

d) $Cov(y; \hat{\beta}|X) = Cov(X\beta + u; \hat{\beta}|X) =$

$= \underbrace{Cov(X\beta; \hat{\beta}|X)}_{=0} + Cov(u; \hat{\beta}|X) = Cov(u; \hat{\beta}|X) =$

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

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

Q3

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

a) $$X^T X = \underset{2\times 3}{\begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}} \times \underset{3\times 2}{\begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}} =$$

$$= \begin{pmatrix} 4+1+1 & 2-2+1 \\ 2-2+1 & 1+4+1 \end{pmatrix} = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$$

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

$$(6-\lambda)^2 - 1 = 0$$
$$(5-\lambda)(7-\lambda) = 0$$
$$\lambda_1 = 5 \quad \lambda_2 = 7$$

1. $\lambda_1 = 5$

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

$$x + y = 0 \rightarrow y = -1; \ x = 1$$

2. $\lambda_2 = 7$

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \rightarrow \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \rightarrow v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$-x + y = 0 \rightarrow y = 1 ; \ x = 1$$

$$\rightarrow D = \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix} \qquad V = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Chech: $\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1} \times \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix} \times \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix}$

b) $X_1 = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$ $X_2 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$

$\overline{X}_1 = \frac{2}{3} \approx 0.667$ $\overline{X}_2 = \frac{4}{3} = 1.333$

$\delta_1 = \frac{1}{2} \sum (X_i - \overline{X})^2$

$= 1.528$

$\delta_2 = 0.577$

$X_2^{std} = \begin{pmatrix} -0.577 \\ 1.154 \\ -0.577 \end{pmatrix}$

$X_1^{std} = \begin{pmatrix} 0.873 \\ -1.091 \\ 0.218 \end{pmatrix}$

$X^* = \begin{pmatrix} 0.873 & -0.577 \\ -1.091 & 1.154 \\ 0.218 & -0.577 \end{pmatrix}$

$X^{*T} \cdot X^* = \begin{pmatrix} 2 & -1.889 \\ -1.889 & 2 \end{pmatrix}$

$\lambda_1 = 3.889$ $V_1 = \begin{pmatrix} 0.707 \\ -0.707 \end{pmatrix}$ $D = \begin{pmatrix} 3.889 & 0 \\ 0 & 0.11 \end{pmatrix}$

$\lambda_2 = 0.11$ $V_2 = \begin{pmatrix} 0.707 \\ 0.707 \end{pmatrix}$

$V = \begin{pmatrix} 0.707 & 0.707 \\ -0.707 & 0.707 \end{pmatrix}$

$$X^* \cdot X^{*T} = \begin{pmatrix} 1.095 & -1.619 & 0.523 \\ -1.619 & 2.523 & -0.905 \\ 0.523 & -0.905 & 0.381 \end{pmatrix}$$

$$\lambda_1 = 3.889 \qquad U = \begin{pmatrix} -0.519 & 0.629 & 0.577 \\ 0.805 & 0.135 & 0.577 \\ -0.285 & -0.765 & 0.577 \end{pmatrix}$$

$$\lambda_2 = 0.11$$

$$\lambda_3 \approx 0$$

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

$$\underset{3\times2}{\begin{pmatrix} 0.873 & -0.577 \\ -1.091 & 1.154 \\ 0.218 & -0.577 \end{pmatrix}} = \underset{3\times3}{\begin{pmatrix} -0.519 & 0.629 & 0.577 \\ 0.805 & 0.135 & 0.577 \\ -0.285 & -0.765 & 0.577 \end{pmatrix}} \times \underset{3\times2}{\Sigma} \times \underset{2\times2}{\begin{pmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{pmatrix}}$$

$$\Sigma = \begin{pmatrix} -1.973 & 0 \\ 0 & 0.332 \\ 0 & 0 \end{pmatrix}$$

c) $$P_1 = X^* \times V_1 = \underset{3\times2}{\begin{pmatrix} 0.873 & -0.577 \\ -1.091 & 1.154 \\ 0.218 & -0.577 \end{pmatrix}} \times \underset{2\times1}{\begin{pmatrix} 0.707 \\ -0.707 \end{pmatrix}} =$$

$$= \begin{pmatrix} 1.025 \\ 1.587 \\ 0.562 \end{pmatrix} \qquad \text{rank}(P_1) = 1$$

g_4

$X = UDV^T$

a) $\|X \cdot w\|^2 \xrightarrow[w]{} \max$

s.t. $\|w\|^2 = 1$

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

$$L = \|UDV^T \cdot w\|^2 - \lambda(\|w\|^2 - 1) = 0$$

$$L = \|UDV^T \cdot w\|^2 - \lambda\|w\|^2 + \lambda = 0$$

b) $\underbrace{\|X \cdot w\|^2}_{1\times1} = \underbrace{(X \cdot w)^T}_{1\times n} \times \underbrace{(X \cdot w)}_{n\times1} = w^T \cdot X^T \cdot X \cdot w$

$\underbrace{\|w\|^2}_{1\times1} = \underbrace{w^T}_{1\times n} \cdot \underbrace{w}_{n\times1}$

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

$$d(w^T \cdot X^T \cdot X \cdot w) = d(w^T \cdot X^T) \cdot X \cdot w + w^T \cdot X^T \cdot d(X \cdot w) =$$

$$= \underbrace{d(w)^T}_{1\times K} \cdot \underbrace{X^T}_{K\times n} \cdot \underbrace{X}_{n\times K} \cdot \underbrace{w}_{K\times1} + \underbrace{w^T}_{1\times K} \cdot \underbrace{X^T}_{K\times n} \cdot \underbrace{X}_{n\times K} \cdot \underbrace{d(w)}_{K\times1} = 2w^T \cdot X^T \cdot X \cdot d(w)$$

$$d(\lambda(w^T \cdot w) - \lambda) = d(\lambda(w^T \cdot w)) - d\lambda =$$

$$= d(\lambda) \cdot w^T \cdot w + \lambda \cdot d(w^T \cdot w) - d\lambda$$

$$^{*}d(w^T \cdot w) = dw^T \cdot w + w^T \cdot dw = 2w^T \cdot d(w)$$

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

$$= d(\lambda) \cdot (w^T \cdot w - 1) + 2\lambda \cdot w^T \cdot dw$$

$$d(L) = 2w^T \cdot X^T \cdot X \cdot d(w) - \left(d(\lambda) \cdot (w^T \cdot w - 1) + 2\lambda \cdot w^T \cdot dw\right) =$$

$$= \left(2w^T \cdot X^T \cdot X - 2\lambda \cdot w^T\right) \cdot d(w) - d(\lambda) \cdot (w^T \cdot w - 1) = 0$$

$\forall d(w); d(\lambda)$:

$$\begin{cases} 2w^T \cdot X^T \cdot X - 2\lambda \cdot w^T = 0 \\ w^T \cdot w = 1 \end{cases}$$

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

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

$\Rightarrow$ w - eigenvector of $X^T X$, with corresponding eigenvalue λ

c) eigenvectors of $X^T X \rightarrow$ columns of V, and since the first column of V (V_1) corresponds to the largest $\lambda \Rightarrow w^* = V_1$