HW #3 theory Evdokushin Ilya gr 4

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.

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

$$loss(\hat{\beta}_1; \hat{\beta}_2) = y^T \cdot y - 2(\hat{\beta}_1 + \hat{\beta}_2)y^T \cdot x + (\hat{\beta}_1 + \hat{\beta}_2)^2 \cdot 1 + \lambda(\hat{\beta}_1^2 + \hat{\beta}_2^2)$$

FOC:

$$\left.\begin{aligned} \frac{\partial loss}{\partial \hat{\beta}_1} &= -2y^T \cdot x + 2(\hat{\beta}_1 + \hat{\beta}_2) + \lambda \cdot 2 \cdot \hat{\beta}_1 = 0 \\ \frac{\partial loss}{\partial \beta_2} &= -2y^T \cdot x + (\hat{\beta}_1 + \hat{\beta}_2) + \lambda \cdot 2\hat{\beta}_2 = 0 \end{aligned}\right\} \Rightarrow \hat{\beta}_1 = \hat{\beta}_2 = \frac{y_T \cdot x}{2 + \lambda}$$

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

$$\lim_{\lambda\to\infty} \hat{\beta}_1 = \lim_{\lambda\to\infty} \hat{\beta}_2 = \lim \frac{y^T \cdot x}{2+\lambda} = 0$$

(c) What happens to the sum $\hat{\beta}_1 + \hat{\beta}_2$ when $\lambda \to 0$?

$$\lim_{\lambda\to 0}(\hat{\beta}_1 + \hat{\beta}_2) = \lim_{\lambda\to 0} \frac{2 \cdot y^T \cdot x}{2+\lambda} = y^T \cdot x$$

estimation is unbiased

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

$\hat{\beta} = (x^T \cdot x)^{-1} \cdot x^T \cdot y$, where $y = x \cdot \beta + u$

So, $\hat{\beta} = (x^T \cdot x)^{-1} \cdot x^T \cdot (x + \beta + u) = (x^T \cdot x)^{-1} \cdot x^T \cdot x \cdot \beta + (x^T \cdot x)^{-1} \cdot x^T \cdot u$

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

Consequently, $E(\hat{\beta} \mid x) = \beta$ & $E(\hat{\beta}) = \hat{\beta}$

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

$Var(\beta + (x^T \cdot x)^{-1} \cdot x^T \cdot u \mid x) = Var(A \cdot u \mid x) = A\,Var(u \mid x) \cdot A^T = A \cdot \sigma^2 W \cdot A^T =$

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

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

CI can not be valid, as there is heteroscedasticity when $W \neq I$

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

$$\mathrm{Cov}(y, \hat{\beta} \mid X) = \mathrm{Cov}(X\beta + u, \beta + (X^TX)^{-1} \cdot X^T \cdot u \mid X) =$$

$$= \mathrm{Cov}(X\beta; \beta \mid X) + \mathrm{Cov}(X \cdot \beta, (X^TX)^{-1} \cdot X^T \cdot u \mid X) + \mathrm{Cov}(u, \beta \mid X) + \mathrm{Cov}(u, (X^TX)^{-1} \cdot X^T \cdot u \mid X) =$$

$$= \mathrm{Cov}(u, (X^TX)^{-1} \cdot X^T \cdot u \mid X) = (X^TX)^{-1} \cdot X^T \cdot \mathrm{Var}(u) = \sigma^2 \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot W$$

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.

$$A = X^T \cdot X = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$$

$$\det(A - \lambda I) = \det\begin{pmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{pmatrix} = (6-\lambda)^2 - 1 = 0 \Rightarrow \left[\begin{array}{l} \lambda_1 = 5 \\ \lambda_2 = 7 \end{array} \right.$$

$\lambda_1 = 5$:

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0 \rightarrow v_1 = v_2, \text{ so } v = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$\lambda_2 = 7$

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0 \rightarrow v_1 = v_2 \rightarrow v \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Consequently, $P \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$; $D \begin{pmatrix} 5 & 0 \\ 0 & 7 \end{pmatrix}$; $P^{-1} \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$

$$X^T X = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 7 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

(b) Find the SVD of X.

$$v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \Rightarrow \quad V = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \end{pmatrix}$$

$$v_1^N = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad v_2^N = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad X \cdot X^T = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix}$$

$$\det(X \cdot X^T - \lambda \cdot I) = (5-\lambda) \cdot (7-\lambda) \cdot \lambda \Rightarrow \lambda_1 = 5; \lambda_2 = 7; \lambda_3 = 0$$

$$\lambda_1 = 5 \quad v_1 = \begin{pmatrix} -1 \\ 3 \\ 0 \end{pmatrix} \rightarrow v_1^N = \frac{1}{\sqrt{10}} \begin{pmatrix} -1 \\ 3 \\ 0 \end{pmatrix}$$

$$\lambda_2 = 7 \quad v_2 \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} \rightarrow v_2^N = \frac{1}{\sqrt{14}} \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$$

$$\lambda_3 = 0 \quad v_3 \begin{pmatrix} -3 \\ -1 \\ 5 \end{pmatrix} \rightarrow v_3^N \frac{1}{\sqrt{35}} \begin{pmatrix} -3 \\ -1 \\ 5 \end{pmatrix}$$

$$X = \begin{pmatrix} \frac{3}{\sqrt{14}} & -\frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{35}} \\ \frac{1}{\sqrt{14}} & \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{35}} \\ \frac{2}{\sqrt{14}} & 0 & \frac{5}{\sqrt{35}} \end{pmatrix} \cdot \begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} =$$

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

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

$$\hat{X} = \begin{pmatrix} \frac{3}{\sqrt{14}} \\ \frac{1}{\sqrt{14}} \\ \frac{2}{\sqrt{14}} \end{pmatrix} \cdot (\sqrt{7}) \cdot \left(\frac{1}{\sqrt{2}} \;\; \frac{1}{\sqrt{2}} \right) = \begin{pmatrix} \frac{3}{2} & \frac{3}{2} \\ \frac{1}{2} & \frac{1}{2} \\ 1 & 1 \end{pmatrix}$$

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.

$$L = \|Xw\|^2 + \lambda(1 - \|w\|^2; \quad \text{as } \|a\|^2 = \langle a, a \rangle$$

$$L = \langle Xw, Xw \rangle + \lambda(1 - \langle w, w \rangle) = w^T X^T X w + \lambda(1 - w^T w)$$

e) FOC: $\begin{cases} \frac{\partial L}{\partial w} = 0 \\ \frac{\partial L}{\partial \lambda} = 0 \end{cases} \Rightarrow \begin{cases} dL = 0 \\ w^T w = 1 \end{cases} \Rightarrow \begin{cases} d(w^T) X^T X w + w^T X^T X d(w) - \lambda d(w^T) w = \lambda w^T d(w) \\ w^T w = 1 \end{cases}$

$d(w^T)w = w^T d(w)$; $d(w^T) X^T X w = w^T (X^T X) d(w) \Rightarrow 2 w^T X^T X d(w) - 2\lambda w^T d(w) =$

$= 2w^T (X^T X - 2I) d(w) = 0 \Rightarrow w^T (X^T X - \lambda I) = 0$; $w \neq 0$ as $w^T w = 1 \Rightarrow w^T (X^T X) = \lambda w^T \Rightarrow$

$\Rightarrow (X^T X) w = \lambda w \Rightarrow \begin{cases} X^T X w = \lambda w \\ w^T w = 1 \end{cases}$

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

w is an eigenvector of $X^T X$ as $X^T X = U D^T U^T U D U^T = V D^T D V^T$

columns of V are eigenvectors of $X^T X$ and $d_{11}, d_{22}, d_{33} \ldots$ are eigenvalues.

The best eigenvector should correspond to the highest eigenvalue $\Rightarrow$ 1st column of V is w^*.