# HA #3

Ilya Demskiy,
1st group

## Problem 1

$$\hat{y} = X\hat{\beta}$$

a) $\hat{\mathcal{L}}(\beta) = (y-\hat{y})^T(y-\hat{y}) + \lambda\hat{\beta}^T\hat{\beta}$; $\hat{y} = \hat{\beta}_1 x + \hat{\beta}_2 x$

$$d\hat{\mathcal{L}} = d(y-\hat{y})^T(y-\hat{y}) + \lambda\, d\hat{\beta}^T\cdot\hat{\beta} =$$

$$= -2(y-\hat{y})^T d\hat{y} + 2\lambda\hat{\beta}^T d\hat{\beta} \quad ; \quad d\hat{y} = \underbrace{dX}_{0}\cdot\hat{\beta} + X\cdot d\hat{\beta}$$

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

$$= \left(-2(y-X\hat{\beta})^T\cdot X\cdot\cancel{I} + 2\lambda\cdot\hat{\beta}^T\right)\cdot d\hat{\beta} = 0 \text{ by FOC}$$

$$\Rightarrow 2\lambda\cdot\hat{\beta}^T = 2(y-X\hat{\beta})^T\cdot X \quad | :2$$

$$\lambda\cdot\hat{\beta}^T = \cancel{2}(y-X\hat{\beta})^T\cdot X \quad | (\cdot)^T$$

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

$$X^TX\hat{\beta} + \lambda\cdot\hat{\beta} = X^Ty$$

$$(X^TX+\lambda I)\cdot\hat{\beta} = X^Ty$$

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

Assuming according to the problem X are absolutely identical

$$\lambda I + X^TX = \begin{pmatrix} \sum x_i^2+\lambda & \sum x_i^2 \\ \sum x_i^2 & \sum x_i^2+\lambda \end{pmatrix}$$

$$(X^TX+\lambda I)^{-1} = \frac{1}{\lambda^2+2\lambda\sum x_i^2}\cdot\begin{pmatrix} \sum x_i^2+\lambda & -\sum x_i^2 \\ -\sum x_i^2 & \sum x_i^2+\lambda \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix}\hat{\beta}_1 \\ \hat{\beta}_2\end{pmatrix} = \frac{1}{\lambda^2+2\lambda\sum x_i^2}\begin{pmatrix} \lambda\sum x_iy_i + \sum x_i^2\sum x_iy_i - \sum x_i^2\sum x_iy_i \\ -\cancel{x}\sum x_iy_i\cdot\sum x_i^2 + \sum x_i^2\sum x_iy_i + \lambda\sum x_iy_i \end{pmatrix}$$

$$= \frac{1}{\lambda^2+2\lambda\sum x_i^2}\begin{pmatrix} \lambda\sum x_iy_i \\ \lambda\sum x_iy_i \end{pmatrix} = \frac{\lambda}{\lambda^2+2\lambda\sum x_i^2}\begin{pmatrix} \sum x_iy_i \\ \sum x_iy_i \end{pmatrix}$$

b) $\begin{pmatrix} \hat{\beta}_1 \\ \beta_2 \end{pmatrix} = \frac{1}{\lambda + 2\sum x_i^2} \begin{pmatrix} \sum x_i y_i \\ \sum x_i y_i \end{pmatrix}$

$\lim_{\lambda \to \infty} \Rightarrow \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

c) $\hat{\beta}_1 + \hat{\beta}_2 = \frac{2 \sum x_i y_i}{\lambda + 2 \sum x_i^2}$ if $\lambda = 0$

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

Problem 2

$$y = X\beta + u \qquad Var(u) = \sigma^2 W \; ; \; W \neq I$$

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

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

$$= (X^T X)^{-1} X^T \cdot X \cdot E(\beta) + (X^T X)^{-1} X^T \cdot \underbrace{E(u \mid X)}_{= 0}$$

$$= \beta$$

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

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

$$= Var\big(\underbrace{(X^T X)^{-1} \cdot X^T}_{A} \cdot u \mid X\big) = A \; Var(u \mid X) \cdot A^T =$$

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

c) Let $W = I + M$

$$\Rightarrow Var(\hat{\beta} \mid X) = \sigma^2 \cdot I (X^T X)^{-1} + (X^T X)^{-1} \cdot X^T \sigma^2 M X (X^T X)^{-1}$$

Since $\sigma^2 \cdot I (X^T X)^{-1}$ is standard for standard estimato and we have additional term, it is obvious that standard confidence level will be invalid for this case (i.e. it should be larger)

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

X is known

$$= Cov(y, X^T X^{-1} X^T y \mid X) = (X^T X)^{-1} X^T \, Var(y \mid X) =$$

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

Problem 3

a) $X = [3 \times 2] = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}$ $X = U \cdot D \cdot V^T$ $X = [n \times k]$

$U = [n \times n]$ $V^T = [k \times k]$

$D = [n \times k]$

$X^T X = V \cdot D^T \cdot D \cdot V^T$

$X^T X = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$ $\begin{pmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{pmatrix} = \lambda^2 - 12\lambda + 35$

$=$

$\lambda_1 = 7$ $v_1 = \cancel{\begin{pmatrix} -1 \\ 1 \end{pmatrix}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$\lambda_2 = 5$ $v_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

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

$X \cdot X^T = U \cdot D \cdot D^T \cdot U^T$

$X \cdot X^T = \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix}$ $\det(A - \lambda I) = -\lambda^3 + 12\lambda^2 - 35\lambda$

$\lambda_1 = 7$ $v_1 = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} \approx \frac{1}{\sqrt{14}} \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$

$\lambda_2 = 5$ $v_2 = \begin{pmatrix} -1 \\ 3 \\ 0 \end{pmatrix} \approx \frac{1}{\sqrt{10}} \begin{pmatrix} -1 \\ 3 \\ 0 \end{pmatrix}$

$\lambda_3 = 0$ $v_3 = \begin{pmatrix} -3 \\ -1 \\ 5 \end{pmatrix} \approx \frac{1}{\sqrt{35}} \begin{pmatrix} -3 \\ -1 \\ 5 \end{pmatrix}$

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

b $\Rightarrow X = U \cdot D \cdot V^T =$

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

c)

$$X = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} \quad x_{ik} = \frac{x_{ik} - \bar{x}_k}{\sqrt{\frac{1}{n-1} \Sigma (x_{ik} - \bar{x}_k)^2}}$$

$$\bar{x}_1 = \frac{2-1+1}{3} = \frac{2}{3} \quad \bar{x}_2 = \frac{1+2+1}{3} = \frac{4}{3}$$

$$s_1^2 = \frac{1}{2}\left(\left(2-\frac{2}{3}\right)^2 + \left(-1-\frac{2}{3}\right)^2 + \left(1-\frac{2}{3}\right)^2\right) = 2\frac{1}{3} = \frac{7}{3}$$

$$s_2^2 = \frac{1}{3}$$

$$\hat{X} = \begin{pmatrix} \frac{4\sqrt{21}}{21} & -\frac{1}{\sqrt{3}} \\ -\frac{5}{\sqrt{21}} & \frac{2}{\sqrt{3}} \\ \frac{\sqrt{21}}{21} & -\frac{1}{\sqrt{3}} \end{pmatrix} = \begin{pmatrix} \frac{4}{\sqrt{21}} & -\frac{1}{\sqrt{3}} \\ -\frac{5}{\sqrt{21}} & \frac{2}{\sqrt{3}} \\ \cancel{\sqrt{2}}\,\frac{1}{\sqrt{21}} & -\frac{1}{\sqrt{3}} \end{pmatrix}$$

$rk = 1$

But I will use SVD rank 1 approx

using $X = U D \cdot V^T$ choose biggest lambda

$\lambda = \sqrt{7}$

$$u_1 = \begin{pmatrix} \frac{3}{\sqrt{14}} \\ \frac{1}{\sqrt{14}} \\ \frac{2}{\sqrt{14}} \end{pmatrix} \quad V_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$$

$$\Rightarrow \hat{X} = \begin{pmatrix} \frac{3}{\sqrt{14}} \\ \frac{1}{\sqrt{14}} \\ \frac{2}{\sqrt{14}} \end{pmatrix} \times \left(\sqrt{7}\right) \times \begin{pmatrix} \cancel{\frac{1}{\sqrt{2}}} \\ \cancel{-\frac{1}{\sqrt{2}}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^{\mathbb{Z}}$$

$$= \begin{pmatrix} \frac{3 \cdot 7}{14} & \frac{3 \cdot 7}{14} \\ \frac{7}{14} & \frac{7}{14} \\ \frac{7}{7} & \frac{7}{7} \end{pmatrix} = \begin{pmatrix} 1{,}5 & 1.5 \\ 0.5 & 0.5 \\ 1 & 1 \end{pmatrix}$$

Problem 4

$X = U \cdot D \cdot V^T \quad X^T \cdot X = V D^T \cdot D \cdot V^T$

d) $\mathcal{L} = (Xw)^T (Xw) - \lambda (w^T \cdot w - 1)$

$\mathcal{L} = w^T V \cdot D^T \cdot D \cdot V^T \cdot w - \lambda w^T \cdot w + \lambda$

e) $d\mathcal{L} = dw^T \cdot V \cdot D^T \cdot D \cdot V^T \cdot w + w^T \cdot V \cdot D^T \cdot D \cdot V^T dw$ $\left(w^T \cdot V \cdot D^T \cdot D^T V^T dw = dw^T \cdot V \cdot D^T \cdot D V^T w \text{ due to matrix squaring}\right)$

$- \lambda \cdot dw^T \cdot w - \lambda w^T \cdot dw$

$\Rightarrow 2 w^T \cdot V \cdot D^T \cdot D \cdot V^T \cdot dw - 2\lambda \cdot w^T dw = 0$

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

FOC: $\begin{cases} w^T (V \cdot D^T \cdot D \cdot V^T) \\ w^T \cdot w = 1 \end{cases}$

f) $\begin{cases} V \cdot D^T \cdot D \cdot V^T \cdot w = \lambda \cdot w \\ w^T \cdot w = 1 \end{cases}$

$V \cdot D^T \cdot D \cdot V^T = V \cdot \underset{\text{diagonal}}{Z} \cdot V^T = A$

$V \cdot Z \cdot V^T w = \lambda \cdot w$

$A \cdot w = \lambda w$

$\Rightarrow w$ is the eigen vector of $A$

For $(Xw)^T (Xw)$ to be maximal we should choose eigen vector corresponding to biggest eigen value

because $(Xw)^T (Xw) = y \cdot D \cdot y^T = \lambda_1^2 y_1 + \lambda_2^2 y_2 + \dots + \lambda_n^2 y_n$

$\Rightarrow w$ is the first column of $V$

as this yields $\lambda_1^2$ which is the biggest