# ICEF. Data Science

Zaykovskaya Alexandra

## Home assignment 3

В данной работе я хочу использовать 1 горящий с нулями для переносов дедлайна (если это возможно)

(3)

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

$$X^TX = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$$

$a_{11} = 2\cdot 2 + (-1)(-1) + 1\cdot 1 = 6$ ; $a_{12} = 2\cdot 1 + (-1)\cdot 2 + 1\cdot 1 = 1$

$a_{21} = 1\cdot 2 + 2(-1) + 1\cdot 1 = 1$ ; $a_{22} = 1\cdot 1 + 2\cdot 2 + 1\cdot 1 = 6$

To diagonalize this matrix, we must first its eigenvalues:

$$\det(X^TX - \lambda I) = \det\begin{pmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{pmatrix} = 0 \Leftrightarrow$$

$$\Leftrightarrow (6-\lambda)^2 - 1 = 0 \Leftrightarrow 36 - 12\lambda + \lambda^2 - 1 = 0 \Leftrightarrow$$

$$\Leftrightarrow \lambda^2 - 12\lambda + 35 = 0$$

$$D = 144 - 140 = 4 \rightarrow \lambda = \frac{12 \pm \sqrt{4}}{2} \Leftrightarrow \begin{cases} \lambda_1 = 7 \\ \lambda_2 = 5 \end{cases}$$

Thus: $D = \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix}$

Finding eigenvectors of $X^TX$:

$\underline{\lambda_1 = 7}$:

$$(X^TX - 7\cdot I)v = 0 \Leftrightarrow \begin{pmatrix} 6-7 & 1 \\ 1 & 6-7 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0 \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} -1\cdot v_1 + 1\cdot v_2 = 0 \\ 1\cdot v_1 - 1\cdot v_2 = 0 \end{cases} \Leftrightarrow v_1 = v_2 \Rightarrow v = \begin{pmatrix} 1 \\ 1 \end{pmatrix} ; \|v\| = \sqrt{1^2 + 1^2} = \sqrt{2} \rightarrow$$

$$\rightarrow \text{norm.: } v = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

$\lambda_2 = 5$:

$$(X^TX - 5\cdot I)v = 0 \iff \begin{pmatrix}6-5 & 1\\ 1 & 6-5\end{pmatrix}\begin{pmatrix}v_1\\ v_2\end{pmatrix} = 0 \iff v_1 + v_2 = 0 \iff$$

$$\iff v_1 = -v_2 \Rightarrow v = \begin{pmatrix}-1\\ 1\end{pmatrix};\ \|v\| = \sqrt{(-1)^2+1^2} = \sqrt{2} \to \text{norm.}: v = \begin{pmatrix}-\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix}$$

$$\cancel{V = \begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix} \to V^T = \begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix}}$$

$$\cancel{X^TX = VDV^T = \begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}\begin{pmatrix}6 & 1\\ 1 & 6\end{pmatrix}\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix}}$$

$$\cancel{\text{Thus: } X^TX = \begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}\begin{pmatrix}7 & 0\\ 0 & 5\end{pmatrix}\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix}}$$

$$V = \begin{pmatrix}\text{norm. } v_{\lambda_1} & \text{norm. } v_{\lambda_2}\end{pmatrix} = \begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix} \to V^T = \begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}$$

Thus:

$$X^TX = VDV^T = \begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}\begin{pmatrix}7 & 0\\ 0 & 5\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix} \leftarrow \text{diagonalization}$$

3 By $X = U\Sigma V^T$ by SVD, $U^TU = I$; $V^TV = I$; $\Sigma$ - diagonal

$$X^TX = (U\Sigma V^T)^T U\Sigma V^T = V\Sigma^T \underbrace{U^TU}_{= I \text{ (by SVD)}} \Sigma V^T =$$

$$= V\Sigma^T\Sigma V^T$$

$$XX^T = U\Sigma V^T (U\Sigma V^T)^T = U\Sigma V^T V \Sigma^T U^T =$$

$$= U\Sigma\Sigma^T U^T.$$

Found in a): $V = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$; $V^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$

Let us find $U$ from $XX^T$:

$$XX^T = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$a_{11} = 2\cdot 2 + 1\cdot 1 = 5$; $a_{12} = 2(-1) + 1\cdot 2 = 0$; $a_{13} = 2\cdot 1 + 1\cdot 1 = 3$

$a_{21} = -2+2 = 0$; $a_{22} = 1+4 = 5$; $a_{23} = -1+2 = 1$

$a_{31} = 2+1 = 3$; $a_{32} = -1+2 = 1$; $a_{33} = 1+1 = 2$

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

Finding eigenvalues:

$\det(XX^T - \lambda I) = 0 \Leftrightarrow \det \begin{pmatrix} 5-\lambda & 0 & 3 \\ 0 & 5-\lambda & 1 \\ 3 & 1 & 2-\lambda \end{pmatrix} = 0 \Leftrightarrow$

$\Leftrightarrow (5-\lambda)(5-\lambda)(2-\lambda) + 0 + 0 - (3(5-\lambda)\cdot 3 + 0 + (5-\lambda)\cdot 1^2) =$

$= 0 \Leftrightarrow 50 - 45\lambda + 12\lambda^2 - \lambda^3 - 45 + 9\lambda - 5 + \lambda = 0 \Leftrightarrow$

$\Leftrightarrow -\lambda(\lambda^2 - 12\lambda + 35) = 0$

$D = (-12)^2 - 4\cdot 35 = 144 - 140 = 4 \rightarrow \lambda_{1,2} = \frac{12 \pm \sqrt{4}}{2} = \left[\begin{array}{l} 7 \\ 5 \end{array}\right.$

Thus: $D = \begin{pmatrix} 7 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \Sigma \Sigma^T$

Finding eigenvectors:

$\underline{\lambda_1 = 7:}$ $(XX^T - 7\cdot I)v = 0 \Leftrightarrow$

$\cancel{\Leftrightarrow \begin{pmatrix} 7-7 & 0 & 0 \\ 0 & 5-7 & 0 \\ 0 & 0 & 0-7 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0 \Leftrightarrow \begin{cases} -2\cdot v_2 = 0 \\ -7\cdot v_3 = 0 \\ 0\cdot v_1 = 0 \end{cases}}$

$$\Leftrightarrow \begin{pmatrix} 5-7 & 0 & 3 \\ 0 & 5-7 & 1 \\ 3 & 1 & 2-7 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0 \Leftrightarrow \begin{pmatrix} -2 & 0 & 3 \\ 0 & -2 & 1 \\ 3 & 1 & -5 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0 \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} -2v_1 + 3v_3 = 0 \\ -2v_2 + 1 \cdot v_3 = 0 \\ 3v_1 + 1 \cdot v_2 - 5 \cdot v_3 = 0 \end{cases} \Leftrightarrow \begin{cases} v_1 = \frac{3}{2} v_3 \\ v_2 = \frac{1}{2} v_3 \\ \frac{9}{2} \cdot v_3 + \frac{1}{2} v_3 - \frac{10}{2} v_3 = 0 \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} v_1 = \frac{3}{2} v_3 \\ v_2 = \frac{1}{2} v_3 \\ \frac{10-10}{2} v_3 = 0 \end{cases} \Leftrightarrow \begin{cases} v_1 = \frac{3}{2} t \\ v_2 = \frac{1}{2} t \\ v_3 = t \end{cases} \rightarrow v_{\lambda_1} = \begin{pmatrix} \frac{3}{2} \\ \frac{1}{2} \\ 1 \end{pmatrix}$$

$\underline{\lambda_2 = 5:}$

$$(XX^T - 5 \cdot I) \cdot v = 0 = \begin{pmatrix} 5-5 & 0 & 3 \\ 0 & 5-5 & 1 \\ 3 & 1 & 2-5 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0 \Leftrightarrow$$

$$\Leftrightarrow \begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 1 \\ 3 & 1 & -3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0 \Leftrightarrow \begin{cases} v_3 = 0 \\ 3v_1 + 1 \cdot v_2 - 3 \cdot v_3 = 0 \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} v_3 = 0 \\ v_1 = -\frac{1}{3} v_2 \end{cases} \rightarrow \begin{cases} v_1 = -\frac{1}{3} t \\ v_2 = t \\ v_3 = 0 \end{cases} \rightarrow v_{\lambda_2} = \begin{pmatrix} -\frac{1}{3} \\ 1 \\ 0 \end{pmatrix}$$

$\underline{\lambda_3 = 0:}$

$$(XX^T - 0 \cdot I) \cdot v = 0 \Leftrightarrow \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0 \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} 5v_1 + 3v_3 = 0 \\ 5v_2 + 1 \cdot v_3 = 0 \\ 3v_1 + 1 \cdot v_2 + 2 \cdot v_3 = 0 \end{cases} \Leftrightarrow \begin{cases} v_1 = -\frac{3}{5} v_3 \\ v_2 = -\frac{1}{5} v_3 \\ -\frac{9}{5} v_3 - \frac{1}{5} v_3 + \frac{10}{5} v_3 = 0 \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} v_1 = -\frac{3}{5}t \\ v_2 = -\frac{1}{5}t \\ v_3 = t \end{cases} \Leftrightarrow v_{\lambda_3} = \begin{pmatrix} -\frac{3}{5} \\ -\frac{1}{5} \\ 1 \end{pmatrix}$$

Normalized eigenvectors:

$$\|v_{\lambda_1}\| = \sqrt{\frac{9}{4} + \frac{1}{4} + \frac{4}{4}} = \sqrt{\frac{14}{4}} \Rightarrow$$

$$\Rightarrow \text{Normalized: } v_{\lambda_1} = \frac{1}{\frac{\sqrt{14}}{2}} \begin{pmatrix} 3/2 \\ 1/2 \\ 1 \end{pmatrix} = \begin{pmatrix} 3/\sqrt{14} \\ 1/\sqrt{14} \\ 2/\sqrt{14} \end{pmatrix}$$

$$\|v_{\lambda_2}\| = \sqrt{\frac{1}{9} + \frac{9}{9}} = \frac{\sqrt{10}}{3} \Rightarrow \text{Norm.: } v_{\lambda_2} = \frac{1}{\frac{\sqrt{10}}{3}} \cdot \begin{pmatrix} -\frac{1}{3} \\ 1 \\ 0 \end{pmatrix} =$$

$$= \begin{pmatrix} -1/\sqrt{10} \\ 3/\sqrt{10} \\ 0 \end{pmatrix}$$

$$\|v_{\lambda_3}\| = \sqrt{\frac{9}{25} + \frac{1}{25} + \frac{25}{25}} = \frac{\sqrt{35}}{5} \rightarrow \text{Norm.: } v_{\lambda_3} = \begin{pmatrix} -3/\sqrt{35} \\ -1/\sqrt{35} \\ 5/\sqrt{35} \end{pmatrix}$$

Thus:

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

$$\cdot \overbrace{\begin{pmatrix} \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{14}} & \frac{2}{\sqrt{14}} \\ -\frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} & 0 \\ -\frac{3}{\sqrt{35}} & -\frac{1}{\sqrt{35}} & \frac{5}{\sqrt{35}} \end{pmatrix}}^{U^T}$$

Continued

Thus:

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

And SVD of X is:

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

~~3 c~~ 3 c) $X = \sum_{j=1}^{rank} \sigma_j u_j v_j^T$

Here: need approximation of rank = 1, which is:

$$X = \sum_{j=1}^{1} \sigma_j u_j v_j^T = \sigma_j u_j v_j^T = \sqrt{7} \begin{pmatrix} \frac{3}{\sqrt{14}} \\ \frac{1}{\sqrt{14}} \\ \frac{2}{\sqrt{14}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} =$$

$$= \begin{pmatrix} 3/\sqrt{2} \\ \frac{1}{\sqrt{2}} \\ 2/\sqrt{2} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 3/2 & 3/2 \\ 1/2 & 1/2 \\ 1 & 1 \end{pmatrix}$$

(2) $y = X\beta + u$ ; $E(u|X) = 0$ ; $Var(u|X) = \sigma^2 W$, $W \neq I$.

$\hat{\beta}$ - standard OLS estimator $\Rightarrow$

$\Rightarrow$ in the matrix form:

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

a) $E(\hat{\beta}|X) = E(\underbrace{(X^T X)^{-1} X^T}_{\text{measurable} \to \text{take out}} y|X) = (X^T X)^{-1} X^T E(y|X) =$

$= \underbrace{(X^T X)^{-1} X^T}_{\text{Name "a"}} E(X\beta + u|X) = a E(\overset{\text{meas.}}{X}\underset{\text{const. (non-stoch.)}}{\beta}|X) +$

$+ a \underbrace{E(u|X)}_{=0} = aX\beta = \underbrace{(X^T X)^{-1} X^T X}_{= I}\beta = \beta$

Thus, tower property

$E(\hat{\beta}) = E(E(\hat{\beta}|X)) = E(\beta) = \beta \to \hat{\beta}$ is unbiased estimator of $\beta$.

non-stoch $\to Var(..|X) = 0$

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

$Var(\hat{y}|X) = \cancel{Var(X\hat{\beta}|X) = X\, Var(\hat{\beta}|X) X^T}$

$= Var(X(X^T X)^{-1} X^T \beta|X) =$

comm. transposed

Cont.

$$\Leftrightarrow X(X^TX)^{-1}X^T \overbrace{var(\beta|X)}^{=\sigma^2 W}\underbrace{(X(X^TX)^{-1}X^T)^T}_{=X(X^TX)^{-1}X^T} =$$

$$= X(X^TX)^{-1}X^T \underbrace{\sigma^2}_{\leftarrow \text{const.} \rightarrow \text{can move}} \overbrace{W}^{\neq I \rightarrow \text{cannot eliminate}} X(X^TX)^{-1}X^T \cancel{= \sigma^2 X(X^TX)^{-1}X^T W}$$

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

2c) The fact that $W \neq I$ means that the residuals $u$ are correlated, which signals of an autocorrelation problem. Under autocorrelation, although regression estimates remain unbiased, (which was seen in 2a)), regression becomes inefficient, and standard errors ~~become~~ are wrongly estimated, which leads to the invalidity of t-tests and F-tests. Thus, standard confidence intervals are invalid as well. What one can do to improve the model and remove the issue is perform an autoregressive transformation.

8

(2) d) $Cov(y;\beta|X) = E(y|X)\cdot E(\beta^T|X) - E(y\cdot\beta^T|X) =$

$$\boxed{\begin{array}{l} y = X\beta + u \\ E(u|X) = 0 \\ E(\beta|X) = \beta \end{array}}$$

$= E(X\beta + u|X)E(\beta^T|X) - E((X\beta + u)\beta^T|X) =$

measurable

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

$- E(X\beta\beta^T|X) - E(u\beta^T|X) =$

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

$- E(u\beta^T|X) = \cancel{X\beta\beta^T} - \cancel{X\beta\beta^T} - E(u\beta^T|X)$

$E(u\beta^T|X) = -Cov(u;\beta|X) + \underbrace{E(u|X)}_{=0}E(\beta|X) = -Cov(u;\beta|X)$

Thus:

$Cov(y;\beta|X) = Cov(u;\beta|X)$

(4) a) $X = UDV^T = U \begin{pmatrix} d_{11} & 0 & 0 & \cdots \\ 0 & d_{22} & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} V^T$

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

Let $\underset{[k\times 1]}{\overset{\text{vector}}{w}} = \begin{pmatrix} w_1 \\ \vdots \\ w_k \end{pmatrix}$ and $\underset{[n\times k]}{\overset{\text{matrix}}{X}} = \begin{pmatrix} x_{11} & x_{12} & \cdots \\ x_{21} & x_{22} & \cdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & x_{nk} \end{pmatrix}$. Then:

$\|w\| = \sqrt{w_1^2 + w_2^2 + \cdots + w_k^2} \longrightarrow \|w\|^2 = w_1^2 + w_2^2 + \cdots + w_k^2$

$w^T = (w_1 \ \ w_2 \ \cdots \ w_k) \to \underset{[k\times 1][1\times k]}{w w^T} = \underset{[k\times k]}{W}$ ; $\underset{[1\times k][k\times 1]}{w^T w} = \underset{[1\times 1]}{W} \leftarrow$ scalar

Thus: $\|w\|^2 = w^T w = 1$

$$Xw = \underset{[n\times k]}{\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & & \ddots & \vdots \\ x_{n1} & \cdots & \cdots & x_{nk} \end{pmatrix}} \underset{[k\times 1]}{\begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_k \end{pmatrix}} =$$

$$= \underset{[n\times 1]}{\begin{pmatrix} x_{11}w_1 + x_{12}w_2 + \cdots + x_{1k}w_k \\ x_{21}w_1 + x_{22}w_2 + \cdots + x_{2k}w_k \\ x_{31}w_1 + x_{32}w_2 + \cdots + x_{3k}w_k \\ \cdots \\ x_{n1}w_1 + x_{n2}w_2 + \cdots + x_{nk}w_k \end{pmatrix}}$$

$\Downarrow$

$$\|Xw\| = \sqrt{\left(\sum_{i=1}^{k} x_{1i}w_i\right)^2 + \left(\sum_{i=1}^{k} x_{2i}w_i\right)^2 + \cdots + \left(\sum_{i=1}^{k} x_{ni}w_i\right)^2} \Rightarrow$$

$$\Rightarrow \|Xw\|^2 = \left(\sum x_{1i}w_i\right)^2 + \cdots + \left(\sum_{i=1}^{k} x_{ni}w_i\right)^2$$

$\underset{[n\times 1]}{Xw}\underset{[1\times n]}{(Xw)^T} = \underset{[n\times n]}{A} \leftarrow$ matrix ; $\underset{[1\times n]}{(Xw)^T}\underset{[n\times 1]}{Xw} = \underset{[1\times 1]}{B} \leftarrow$ scalar. Thus:

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

Thus, the problem can be expressed as:

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

$$\Leftrightarrow \begin{cases} \left(\sum_{i=1}^{k} x_{1i} w_i\right)^2 + \dots + \left(\sum_{i=1}^{k} x_{ni} w_i\right)^2 \to \max\limits_{w_i} \\ s.t. \ \sum_{i=1}^{k} (w_i^2) = 1 \end{cases}$$

And the Lagrangean is:

$$L = \|Xw\|^2 + \lambda(1 - \|w\|^2) = (Xw)^T Xw + \lambda(1 - w^T w) =$$

$$= \left(\sum_{i=1}^{k} x_{1i} w_i\right)^2 + \dots + \left(\sum_{i=1}^{k} x_{ni} w_i\right)^2 + \lambda\left(1 - \sum_{i=1}^{k} (w_i)^2\right).$$

(1) by the scalar terms:

FOC:

$$\frac{\partial L}{\partial w_j} = 0 \ \forall w_i, \ i = 1, \dots, k.$$

$$\frac{\partial L}{\partial w_j} = \frac{\partial\left(\left(\sum_i x_{1i} w_i\right)^2 + \dots + \left(\sum_i x_{ni} w_i\right)^2 + \lambda\left(1 - \sum_i (w_i^2)\right)\right)}{\partial w_j} =$$

$$= 2x_{1j}\left(\sum_{i=1}^{k} x_{1i} w_i\right) + \dots + 2x_{nj}\left(\sum_{i=1}^{k} x_{ni} w_i\right) - 2\lambda w_j = 0 \to$$

$$\to w_j = \frac{x_{1j}\left(\sum_{i=1}^{k} x_{1i} w_i\right) + \dots + x_{nj}\left(\sum_{i=1}^{k} x_{ni} w_i\right)}{\lambda} =$$

$$= w_j = \frac{\sum_{m=1}^{n} x_{mj}\left(\sum_{i=1}^{k} x_{mi} w_i\right)}{\lambda} \leftarrow \forall j \text{ due to the problem symmetry}$$

In matrix terms:

FOC:

$dL = 0 \iff \cancel{d((Xw)^T Xw) + d(w^T}$

$\iff d((Xw)^T Xw + \lambda(1 - w^T w)) = 0 \iff$

$\iff d((Xw)^T Xw) + d(\lambda(1 - w^T w)) = 0 \iff$

$\iff 2(Xw)^T d(Xw) - \lambda d(w^T w) = 0 \iff$

$\cancel{\iff 2Xw} \iff \cancel{2w^T X^T X dw - 2w^T dw = 0} \iff$

$\cancel{\iff w^T X^T X dw - \lambda w^T dw = 0 \iff (w^T X^T X - w^T) dw = 0 \iff}$

$\cancel{\iff w^T (X^T X - \lambda) dw = 0}$

Note:
$d(AB) = A \cdot dB + dA \cdot B$

$\iff 2(Xw)^T X dw + 2(Xw)^T dX \cdot w -$

$= 0$ since $X$ does not depend on $w$ and is const.

$- \lambda d(w^T w) = 0 \iff 2w^T X^T X dw - 2\lambda w^T dw = 0 \iff$

$\iff (w^T X^T X - \lambda w^T) dw = 0 \iff \cancel{w^T (X^T X - \lambda) = 0}$

$\iff w^T X^T X - \lambda w^T = 0. \iff \cancel{w^T = w^T X^T X \cdot \lambda^{-1}}.$

$\iff X^T X w - \lambda w = 0 \iff \cancel{w = (X^T X - \lambda I)^{-1}}$

$\iff w = X^T X w \cdot \underbrace{\lambda^{-1}}_{= \frac{1}{\lambda}}$

(1) $L(\hat\beta) = (y-\hat y)^T(y-\hat y) + \lambda\hat\beta^T\hat\beta$; $\hat y = \hat\beta_1 x + \hat\beta_2 x = X\hat\beta$

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

$= y^T y - y^T X\hat\beta - \hat\beta^T X^T y + \hat\beta^T X^T X\hat\beta + \lambda\hat\beta^T\hat\beta$

$\cancel{dL(\hat\beta) =}$ $L(\hat\beta) \longrightarrow \min_{\hat\beta}$

<u>FOC:</u>

$dL(\hat\beta) = 0$

const. $\to d(y^T X) = 0$. const $\to d(X^T y) = 0$

$dL(\hat\beta) = \underbrace{d(y^T y)}_{0} - d(y^T X\hat\beta) - d(\hat\beta^T X^T y) +$

$+ d(\hat\beta^T(X^T X + \lambda I)\hat\beta) = 0 \Leftrightarrow$

$\Leftrightarrow -y^T X d\hat\beta - \underbrace{d(\hat\beta^T) X^T y}_{= y^T X d\hat\beta} + d(\hat\beta^T(X^T X + \lambda I)\hat\beta) = 0 \Leftrightarrow$

$\Leftrightarrow -2y^T X d\hat\beta + 2\hat\beta^T(X^T X + \lambda I) d\hat\beta = 0 \Leftrightarrow$

$\Leftrightarrow (-y^T X + \hat\beta^T(X^T X + \lambda I)) d\hat\beta = 0 \Leftrightarrow$

$\Leftrightarrow \hat\beta^T(X^T X + \lambda I) - y^T X = 0 \Leftrightarrow \hat\beta = (X^T X + \lambda I)^{-1} \cdot X y^T$

optimal $\hat\beta = \begin{pmatrix}\hat\beta_1 \\ \hat\beta_2\end{pmatrix}$.

In scalar terms:

$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, $\hat y = \begin{pmatrix} \hat y_1 \\ \vdots \\ \hat y_n \end{pmatrix}$; $X = \begin{pmatrix} x_{11} & x_{12} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{pmatrix}$; $\hat\beta = \begin{pmatrix}\hat\beta_1 \\ \hat\beta_2\end{pmatrix}$; $\hat\beta^T = (\hat\beta_1 \ \hat\beta_2)$

Thus: $L = (y_1 \dots y_n)\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - (y_1 \dots y_n)\begin{pmatrix} x_{11} & x_{12} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{pmatrix}\begin{pmatrix}\hat\beta_1 \\ \hat\beta_2\end{pmatrix} -$

$$-(\hat\beta_1\ \hat\beta_2)\begin{pmatrix}x_{11} & x_{21} & \cdots & x_{n1}\\ x_{12} & x_{22} & \cdots & x_{n2}\end{pmatrix}\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix} + (\hat\beta_1\ \hat\beta_2)\begin{pmatrix}x_{11} & \cdots & x_{n1}\\ x_{12} & \cdots & x_{n2}\end{pmatrix}\cdot$$

$$\cdot\begin{pmatrix}x_{11} & x_{12}\\ \vdots & \vdots\\ x_{n1} & x_{n2}\end{pmatrix}\begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix} + \lambda(\hat\beta_1\ \hat\beta_2)\begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix} = \sum_{i=1}^n (y_i^2) - \hat\beta_1\sum_{i=1}^n (y_i x_{i1}) -$$

$$-\hat\beta_2\sum_{i=1}^n (y_i x_{i2}) - \sum_{i=1}^n y_i(\hat\beta_1 x_{i1} + \hat\beta_2 x_{i2}) +$$

$$+\underbrace{\hat\beta_1\sum_{i=1}^n x_{i1}(\hat\beta_1 x_{i1} + \hat\beta_2 x_{i2})}_{=\sum_i x_{i1}^2\hat\beta_1^2 + \sum_i x_{i1}x_{i2}\hat\beta_1\hat\beta_2} + \hat\beta_2\sum_{i=1}^n x_{i2}(\hat\beta_1 x_{i1} + \hat\beta_2 x_{i2}) + \lambda\hat\beta_1^2 + \lambda\hat\beta_2^2$$

FOC:

$$\frac{\partial L}{\partial \beta_i} = 0,\ i = 1,2.$$

For $i=1$:

$$\frac{\partial L}{\partial \hat\beta_1} = 0 \iff -\sum_{i=1}^n (y_i x_{i1}) - \sum_{i=1}^n (y_i x_{i1}) + \cancel{\sum_i x_{i1}\hat\beta_1} +$$

$$+2\sum_i \hat\beta_1 x_{i1}^2 + \sum_i x_{i1}x_{i2}\hat\beta_2 + \sum_i x_{i1}x_{i2}\hat\beta_2 + 2\lambda\hat\beta_1 = 0 \iff$$

$$\iff -2\sum_{i=1}^n (y_i x_{i1}) + 2\sum_{i=1}^n \hat\beta_1 x_{i1}^2 + 2\sum_{i=1}^n x_{i1}x_{i2}\hat\beta_2 + 2\lambda\hat\beta_1 = 0 \iff$$

$$\iff \hat\beta_1 = \frac{\sum_i (y_i x_{i1}) - \sum_i x_{i1}x_{i2}\hat\beta_2}{\sum_i (x_{i1}^2) + \lambda}$$

$$\frac{\partial L}{\partial \hat\beta_2} = 0 \iff -\sum_{i=1}^n (y_i x_{i2}) - \sum_{i=1}^n (y_i x_{i2}) + \sum_{i=1}^n x_{i1}x_{i2}\hat\beta_1 +$$

$$+2\hat\beta_2\sum_i (x_{i2}^2) \underbrace{\ }_{+2\lambda\hat\beta_2} = 0 \iff \hat\beta_2 = \frac{\sum_i y_i x_{i2} - \hat\beta_1\sum_i x_{i1}x_{i2}}{\sum_i (x_{i2})^2 + \lambda}$$

(1) b) As it can be seen, both in the matrix and in the scalar form $\lambda$ is placed in the denominator, which means that, if $\lambda$ tends to infinity both $\hat{\beta}_1$ and $\hat{\beta}_2$ will converge to ~~the~~ zero.