ICEF. Data Science

Home assignment 3

Zaykovskaya
Alexandra

3

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

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

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

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

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:

$\underline{\lambda_1 = 7}$:

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

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

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

$\lambda_2 = 5$:

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

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

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

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

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

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

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

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

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

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

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

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:

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

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

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

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

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

Finding eigenvalues:

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

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

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

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

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

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

Finding eigenvectors:

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

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

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

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

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

$\underline{\lambda_2 = 5:}$

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

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

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

$\underline{\lambda_3 = 0:}$

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

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

$\boxed{4}$

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

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:

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

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

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

$= y^Ty - y^TX\hat\beta - \hat\beta^TX^Ty + \hat\beta^TX^TX\hat\beta + \lambda\hat\beta^T\hat\beta$

$L(\hat\beta) \longrightarrow \min\limits_{\hat\beta}$

<u>FOC:</u>

$dL(\hat\beta) = 0$

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

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

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

$$L(\hat\beta) = (y-\hat y)^T(y-\hat y) + \lambda\hat\beta^T\hat\beta = (y^T-\hat y^T)(y-\hat y) +$$

$$+\lambda\hat\beta^T\hat\beta = y^Ty - y^T\hat y - (\hat y)^T y + (\hat y)^T\hat y + \lambda\hat\beta^T\hat\beta =$$

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

$$+\lambda\hat\beta^T\hat\beta$$

$$\left[\hat y = \hat\beta_1 x + \hat\beta_2 x = X\hat\beta\right]$$

Thus:

$$L(\hat\beta) = y^Ty - y^T X\begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix} - (\hat\beta_1\ \hat\beta_2)X^T y +$$

$$+(\hat\beta_1\ \hat\beta_2)X^TX\begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix} + \lambda(\hat\beta_1\ \hat\beta_2)\begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix} \rightarrow \min_{\hat\beta_1,\hat\beta_2}$$

FOC:

$$\frac{\partial L(\hat\beta)}{\partial\hat\beta_1} = -y^TX\begin{pmatrix}1\\ 0\end{pmatrix} - (1\ 0)X^Ty + (1\ 0)X^TX$$

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

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

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

Thus tower property

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

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

b) $Var(y|X) = Var(X\beta + u|X) = Var(u|X) =$

$= \sigma^2 W$

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

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

comm. transposed

Cont.

$$= X(X^TX)^{-1}X^T\,\overbrace{\mathrm{Var}(\beta|X)}^{e}\,\big(\underbrace{X(X^TX)^{-1}X^T}_{=X(X^TX)^{-1}X^T}\big)^T =$$

$$=$$

~~Machine Learning~~ Data Science

Home assignment 3

Aleksandra Zaykovskaya

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

a) $L(\hat\beta) = y - (\hat\beta_1 x + \hat\beta_2 x)^T (y - \hat\beta_1 x + \hat\beta_2 x) + \lambda \hat\beta^T \hat\beta \to \min_{\hat\beta}$

$\hat\beta = \begin{pmatrix} \hat\beta_1 \\ \hat\beta_2 \end{pmatrix}$ $[2\times 1]$; $X = \begin{pmatrix} | & | \\ x_1 & x_2 \\ | & | \end{pmatrix} \to \hat y = X\hat\beta$ $[n\times 2][2\times 1]$

$[n\times 2] \leftarrow$ since only 2 betas

~~FOC: $\frac{dL(\hat\beta)}{d\hat\beta}$~~

Thus, we can express the loss function as:

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

Let us take the differential: $d(A+B) = dA + dB$

$$dL(\hat\beta) = d\left((y - X\hat\beta)^T (y - X\hat\beta) + \lambda \hat\beta^T \hat\beta\right) = d\left((y - X\hat\beta)^T (y - X\hat\beta)\right) + d(\lambda \hat\beta^T \hat\beta)$$

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

$$d\left((y - X\hat\beta)^T (y - X\hat\beta)\right) = \overbrace{(y - X\hat\beta)^T d(y - X\hat\beta)}^{=\Sigma_i (y_i - \hat y_i) d(y_i - \hat y_i)} + \underbrace{d(y - X\hat\beta)^T \cdot (y - X\hat\beta)}_{=\Sigma_i (y_i - \hat y_i) d(y_i - \hat y_i)} = 2(y - X\hat\beta)^T d(y - X\hat\beta) =$$

$\leftarrow$ proven at the lecture

$$= -2(y - X\hat\beta)^T \cdot X\,d\hat\beta = -2(y - X\hat\beta)^T \cdot X \begin{pmatrix} d\hat\beta_1 \\ d\hat\beta_2 \end{pmatrix}$$

$$d(\lambda \hat\beta^T \hat\beta) = \lambda\, d(\hat\beta^T \hat\beta) = \lambda(\hat\beta^T d\hat\beta + d\hat\beta^T \cdot \hat\beta) = 2\lambda \hat\beta^T d\hat\beta = 2\lambda (\hat\beta_1\ \hat\beta_2) \begin{pmatrix} d\hat\beta_1 \\ d\hat\beta_2 \end{pmatrix}$$

constants

$\boxed{1}$

FOC:

$$dL(\beta) = 0 \iff -2(y - X\beta)^T \cdot X\,d\beta + 2\lambda\beta^T d\beta = 0 \iff$$

$$\iff (2\lambda\beta^T - 2(y - X\beta)^T \cdot X)\,d\beta = 0 \iff \lambda\beta^T - (y - X\beta)^T \cdot X = 0 \iff$$

$$\iff (\lambda\beta^T - (y - X\beta)^T \cdot X)^T = 0^T \iff$$

$$\iff \iff \beta\lambda^T - X^T(y - X\beta) = 0^T \iff$$

$$\iff \beta\lambda^T - X^T y - X^T X\beta = 0 \iff \beta\lambda^T - X^T X\beta = X^T y$$