HA#3 dse Kozakov Luka gr-1

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

(a) Find optimal $\hat\beta_1$ & $\hat\beta_2$ for fixed λ.

$$loss(\hat\beta) \to \min_{\hat\beta_1 \hat\beta_2}$$

$$loss(\hat\beta) = (y - \hat\beta_1 x - \hat\beta_2 x)^T (y - \hat\beta_1 x - \hat\beta_2 x) + \lambda(\hat\beta_1^2 + \hat\beta_2^2) \to \min_{\hat\beta_1, \hat\beta_2}$$

$$loss(\hat\beta) = y^2 - \hat\beta_1 xy - \hat\beta_2 xy - \hat\beta_1 xy + \hat\beta_1^2 x^2 + \hat\beta_1 \hat\beta_2 x^2 - \hat\beta_2 xy + \hat\beta_1 x^2 \hat\beta_2 + \hat\beta_2^2 x^2 + \lambda \hat\beta_1^2 + \lambda \hat\beta_2^2 \to \min_{\hat\beta_1 \hat\beta_2}$$

FOC:

$$\begin{cases} \dfrac{\partial\, loss(\hat\beta)}{\partial \hat\beta_1} = 0 \\ \dfrac{\partial\, loss(\hat\beta)}{\partial \hat\beta_2} = 0 \end{cases}$$

$$\begin{cases} -xy - xy + 2\hat\beta_1 x^2 + \hat\beta_2 x^2 + x^2 \hat\beta_2 + 2\lambda \hat\beta_1 = 0 \\ -xy + \hat\beta_1 x^2 - xy + \hat\beta_1 x^2 + 2\hat\beta_2 x^2 + 2\lambda \hat\beta_2 = 0 \end{cases}$$

$$\hat\beta_1 [2x^2 + 2\lambda] = xy - \hat\beta_2 x^2 - \hat\beta_2 x^2 \to \hat\beta_1 = \frac{xy - \hat\beta_2 x^2 - \hat\beta_2 x^2}{2x^2 + 2\lambda}$$

$$\hat\beta_2 [2x^2 + 2\lambda] = 2xy - 2\hat\beta_1 x^2 \to \hat\beta_2 = \frac{2xy - 2\hat\beta_1 x^2}{2x^2 + 2\lambda}$$

b) $\lambda \to \infty$; $\hat\beta_1$ & $\hat\beta_2 \to 0$

$\boxed{2}$

$y = X\beta + u$

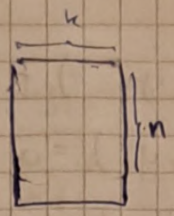
β - non-rnd

$\mathbb{E}(u|X) = 0$

$\mathrm{Var}(u|X) = \sigma^2 W$

$W \neq I$

$\hat{\beta}$ - ols estimator of β

$X \sim$ 

$\mathrm{rk}\, X = k$

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

$$\mathbb{E}(\hat{\beta}) = [(X^T X)^{-1} X^T]\, \mathbb{E}(Y) = [(X^T X)^{-1} X^T] X\beta =$$
$$= [\underbrace{(X^T X)^{-1}}_{\text{inverse}} X^T X]\beta = \beta$$

$\Rightarrow \hat{\beta}$ - unbiased estimator

(b) $$\mathrm{Var}(\hat{\beta}) = [(X^T X)^{-1} X^T][\mathrm{Var}(Y)][(X^T X)^{-1} X^T]^T =$$
$$= [(X^T X)^{-1} X^T]\, I\sigma^2\, [(X^T X)^{-1} X^T]^T$$

$$[(AB)C]^T = C^T (AB)^T = C^T B^T A^T$$

$$C(AB)^T = C B^T A$$

$$\mathrm{Var}(\hat{\beta})^T = [(X^T X)^{-1} X^T]$$

$$\sigma^2 \underbrace{(X^T X)^{-1} X^T X}_{I} (X^T X)^{-1} \rightarrow \mathrm{Var}(\hat{\beta}) = \underbrace{(X^T X)^{-1}}_{W} \sigma^2$$

(c) $\underset{n\times 1}{Y} = \underset{n\times k}{X}\,\underset{k\times 1}{\beta} + \underset{n\times 1}{u}$: $u \sim N(0, I\sigma^2)$ {(No intercept)}

assumption

$\hat{d}$ $\cancel{\;\;}$ $= \sqrt{\mathrm{tr}(X^T X)^{-1} s^2}$, s^2 - unbiased estimator of σ^2

$$CI(\hat{\beta}) = [\hat{\beta} - t_{\frac{\alpha}{2}}(n - (\#\text{parameters}))\cdot \hat{d}\,;\ \hat{\beta} + t_{\frac{\alpha}{2}}(n-p)\hat{d}]$$

k

$\boxed{4}$ $\|Xw\|^2$ subject to $\|w\|^2 = 1$

$$\|UDV^T w\|^2 = \left(\sqrt{[UDV^T w]^T [UDV^T w]}\right)^2$$

$$\rightarrow \left(\sqrt{w^T V D^T \underbrace{U^T U}_{I} D V^T w}\right)^2 = \left(\sqrt{w^T V D V^T w}\right)^2$$

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

$$X^T X = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix} \longrightarrow \begin{vmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{vmatrix} = 0 \rightarrow \chi(\lambda) = (6-\lambda)^2 - 1$$

$$\chi(\lambda) = 36 - 12\lambda + \lambda^2 - 1 = \lambda^2 - 12\lambda + 35 \longrightarrow \lambda_1 = 5;\ \lambda_2 = 7$$

$$V_{\lambda_1}: \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \begin{array}{c|c} v_1 & v_2 \\ \hline -1 & 1 \end{array}$$

$$v_1 = -v_2$$

$$V_{\lambda_1} = \text{span}\left\{\begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}\right\}$$

$$V_{\lambda_2}: \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad -v_1 + v_2 = 0 \rightarrow v_1 = v_2$$

$$\begin{array}{c|c} v_1 & v_2 \\ \hline 1 & 1 \end{array}$$

$$V_{\lambda_2} = \text{span}\left\{\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}\right\}$$

$$U = \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \qquad \sigma_1 = \sqrt{7};\ \sigma_2 = \sqrt{5}$$

$$X X^T = \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix} \longrightarrow \begin{vmatrix} 5-\lambda & 0 & 3 \\ 0 & 5-\lambda & 1 \\ 3 & 1 & 2-\lambda \end{vmatrix} = 0 \rightarrow \chi(\lambda) = -\lambda^3 + 12\lambda^2 - 35\lambda = 0$$

$$\lambda_1 = 0;\ \lambda_2 = 5;\ \lambda_3 = 7$$

$$V_{\lambda_1}: \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{cases} 5v_1 + 0 + 3v_3 = 0 \\ 0 + 5v_2 + v_3 = 0 \end{cases} \rightarrow \begin{cases} 5v_1 = -3v_3 \\ 5v_2 = -v_3 \end{cases} \qquad \begin{array}{c|c|c} v_1 & v_2 & v_3 \\ \hline 3 & 1 & -5 \end{array}$$

$$\hookrightarrow \begin{cases} v_1 = \frac{-3}{5} v_3 \\ v_2 = \frac{-1}{5} v_3 \end{cases} \qquad V_{\lambda_1} = \text{span}\left\{\begin{pmatrix} 3/\sqrt{35} \\ 1/\sqrt{35} \\ -5/\sqrt{35} \end{pmatrix}\right\}$$

$$V_{\lambda_2=5} = \operatorname{span}\left\{\begin{pmatrix} -1/\sqrt{10} \\ 3/\sqrt{10} \\ 0 \end{pmatrix}\right\} \; ; \quad V_{\lambda_3=7} = \operatorname{span}\left\{\begin{pmatrix} 3/\sqrt{14} \\ 1/\sqrt{14} \\ 2/\sqrt{14} \end{pmatrix}\right\}$$

$$V^T = \begin{pmatrix} 3/\sqrt{14} & 1/\sqrt{14} & 2/\sqrt{14} \\ -1/\sqrt{10} & 3/\sqrt{10} & 0 \\ 3/\sqrt{35} & 1/\sqrt{35} & -5/\sqrt{35} \end{pmatrix}$$

$$X = \left[\underbrace{\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}}_{2\times 2} \underbrace{\begin{pmatrix} \sqrt{7} & 0 & 0 \\ 0 & \sqrt{5} & 0 \end{pmatrix}}_{2\times 3} \underbrace{\begin{pmatrix} 3/\sqrt{14} & 1/\sqrt{14} & 2/\sqrt{14} \\ -1/\sqrt{10} & 3/\sqrt{10} & 0 \\ 3/\sqrt{35} & 1/\sqrt{35} & -5/\sqrt{35} \end{pmatrix}^T}_{3\times 3}\right]^T$$

(c) Best low rank approximation:

$$\sqrt{7} \cdot \underbrace{\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}}_{2\times 1} \underbrace{\left(\frac{3}{\sqrt{14}} \;\; \frac{1}{\sqrt{14}} \;\; \frac{2}{\sqrt{14}} \right)}_{1\times 3} = \begin{pmatrix} 3/2 & 1/2 & 1 \\ 3/2 & 1/2 & 1 \end{pmatrix}$$

$$\boxed{rk = 1}$$

$$\underbrace{\begin{pmatrix} 2{,}646 & 0 \\ 0 & 2{,}236 \\ 0 & 0 \end{pmatrix}}_{\Sigma} \qquad V^T = \begin{pmatrix} 0{,}707 & -0{,}707 \\ 0{,}707 & 0{,}707 \end{pmatrix}^T \qquad \begin{pmatrix} 3/2 & 3/2 \\ 1/2 & 1/2 \\ 1 & 1 \end{pmatrix}$$

$$V D V^T = \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix}$$

$$\cancel{\max_{\|w\|=1} \|Xw\|^2 \longrightarrow w^T \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix} w \longrightarrow \max_{\|w\|=1}}$$

$\langle w ; Ww \rangle$ - Quadratic form

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

$\boxed{4}$ $\begin{cases} \|Xw\|^2 \to \max\limits_w \\ \text{s.t. } \|w\|^2 = 1 \end{cases}$ $\quad X = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}$

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

$\|Xw\|^2 - \lambda\|w\|^2 + \lambda \longrightarrow FOC$

FOC: $df(x) = d(\|Xw\|^2 - \lambda\|w\|^2 + \lambda)$

$\|Xw\|^2 = \langle Xw; Xw \rangle$

$d\langle Xw; Xw \rangle = 2\langle Xw; dXw \rangle = 2\langle Xw; X\,dw \rangle$

$d\|Xw\|^2 = \cancel{\tfrac{1}{2}} \cdot 2\langle X^T X w; dw \rangle \to \nabla f(x) = 2X^T X w$

$d(-\lambda\|w\|^2) = -\lambda\langle w; dw \rangle \cdot 2 = -2\lambda\langle w; dw \rangle$

$2\langle X^T X w; dw \rangle - 2\lambda\langle w; dw \rangle$

$= 2\langle \underbrace{X^T X w - 2\lambda w}_{\nabla f}; dw \rangle \to \nabla f = 2X^T X w - 2\lambda w \overset{!}{=} 0$

$\nabla^2 f(x) > 0 \to x^*$ - strict local minimum

$\nabla^2 f(x) < 0 \to x^*$ - strict local maximum

$2X^T X w = 2\lambda w \to X^T X w = \lambda w$ eigenvalue problem.

$\begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \lambda\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \to \begin{pmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{pmatrix}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$\lambda_1 = 5;\ \lambda_2 = 7 \qquad 7 > 5$

$V_{\lambda_1} = \text{span}\begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix};\ V_{\lambda_2} = \text{span}\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$

$\nabla^2 f = 2(X^T X - 7I) = 2\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} \boxed{-2} & 2 \\ 2 & -2 \end{pmatrix}$

$\Delta_1 = -2 < 0 \qquad \Delta_2 = \det(\nabla^2 f) = 0 \leq 0$

$\Rightarrow \nabla^2 f$ is negative semi-definite $\to \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$ is a local maximum.