Alex Nikolaenko gr1

HA3 Datascience

1. $E(x) = 0;\ V(x) = 1;\ E(y) = 0$

$loss(\hat\beta) = (y-\hat y)^T(y-\hat y) + \lambda \hat\beta^T\hat\beta$, $\hat y = \hat\beta_1 x + \hat\beta_2 x \Rightarrow \hat\beta = \begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix}$ and $\hat y = \hat{x}\hat\beta$, $\hat{x} = \begin{pmatrix}x\\ x\end{pmatrix}$

a) we minimize $loss(\hat\beta)$ by $\hat\beta_1, \hat\beta_2$ s.t. $\lambda = \bar\lambda$;

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

$= y^Ty - \hat\beta^T\hat{x}^Ty - y^T\hat{x}\hat\beta + \hat\beta^T\hat{x}^T\hat{x}\hat\beta + \lambda\hat\beta^T\hat\beta$

F.O.C.: $d\,loss(\hat\beta) = 0$

$(d\hat\beta_1\ d\hat\beta_2)\begin{pmatrix}x_1 & \cdots & x_n\\ x_1 & \cdots & x_n\end{pmatrix}\cdot\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix} = [1\times2][2\times n][n\times1] = 1\times1 \Rightarrow d(\hat\beta^T)\hat{x}^Ty = (d(\hat\beta^T)\hat{x}^Ty)^T$

$\Rightarrow d(\overset{=0}{y^Ty}) - d(\hat\beta^T)\hat{x}^Ty - y^T\hat{x}\,d(\hat\beta) + d(\hat\beta^T)(\hat{x}^T\hat{x}\hat\beta) + \hat\beta^T(d(\hat{x}^T\hat{x}\hat\beta)) + \lambda d(\hat\beta^T)\hat\beta + \lambda\hat\beta^T d\hat\beta = 0$

$\Rightarrow -2y^T\hat{x}\,d(\hat\beta) + \hat\beta^T\hat{x}^T\hat{x}\,d\hat\beta + \hat\beta^T(d(\hat{x}^T\hat{x}\hat\beta)) + \lambda d(\hat\beta^T)\hat\beta +$

$(d\hat\beta_1\ d\hat\beta_2)\begin{pmatrix}x_1 & \cdots & x_n\\ x_1 & \cdots & x_n\end{pmatrix}\begin{pmatrix}x_1 & x_1\\ \vdots & \vdots\\ x_n & x_n\end{pmatrix}\cdot\begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix} =$

$+ \lambda\hat\beta^T d\hat\beta = 0$

$\lambda\cdot(\hat\beta_1, \hat\beta_2)\begin{pmatrix}d\hat\beta_1\\ d\hat\beta_2\end{pmatrix} = [1\times1][1\times2][2\times1] = 1\times1$

$\Rightarrow \lambda\hat\beta^T d\hat\beta = (\lambda\hat\beta^T d\hat\beta)^T$

$= [1\times2][2\times n][n\times2][2\times1] = 1 \Rightarrow$

$\Rightarrow d(\hat\beta^T)(\hat{x}^T\hat{x}\hat\beta) = [d(\hat\beta)^T(\hat{x}^T\hat{x}\hat\beta)]^T$

$\Rightarrow -2y^T\hat{x}\,d(\hat\beta) + 2\hat\beta^T\hat{x}^T\hat{x}\,d(\hat\beta) + 2\lambda\hat\beta^T d(\hat\beta) = 0 \Rightarrow [y^T\hat{x} - \beta^T(\hat{x}^T\hat{x} + I_{2\times2}\lambda)]d(\hat\beta) = 0$

$\Rightarrow y^T\hat{x} = \beta^T(\hat{x}^T\hat{x} + \lambda I_{2\times2}) \Rightarrow \hat{x}^Ty = (\hat{x}^T\hat{x} + \lambda I_{2\times2})^T\beta$

$(\hat{x}^T\hat{x})^T = \hat{x}^T\hat{x}$, $(\lambda I)^T = \lambda I \Rightarrow$

$\Rightarrow (\hat{x}^T\hat{x} + \lambda I)^T = \hat{x}^T\hat{x} + \lambda I$

$\Rightarrow$ since $(\hat{x}^T\hat{x} + \lambda I_{(2\times2)})$ is symmetric, inverse exists

$\Rightarrow \beta = (\hat{x}^T\hat{x} + \lambda I_{(2\times2)})^{-1}\hat{x}^Ty$; $\hat{x}^T\hat{x} = \begin{pmatrix}\Sigma x_i^2 & \Sigma x_i^2\\ \Sigma x_i^2 & \Sigma x_i^2\end{pmatrix}$; $\lambda I_{(2\times2)} = \begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix}$; $\hat{x}^Ty = \begin{pmatrix}\Sigma x_iy_i\\ \Sigma x_iy_i\end{pmatrix}$

$$\Rightarrow (\hat{x}^T\hat{x} + \lambda I_{(2\times2)})^{-1} = \frac{1}{(\lambda + \Sigma x_i^2)^2 - (\Sigma x_i^2)^2}\cdot\begin{pmatrix}\lambda+\Sigma x_i^2 & -\Sigma x_i^2\\ -\Sigma x_i^2 & \lambda + \Sigma x_i^2\end{pmatrix} = \frac{1}{\lambda^2 + 2\lambda\Sigma x_i^2}\begin{pmatrix}\lambda+\Sigma x_i^2 & -\Sigma x_i^2\\ -\Sigma x_i^2 & \lambda+\Sigma x_i^2\end{pmatrix}$$

$$\Rightarrow \beta = \frac{1}{\lambda^2 + 2\lambda\Sigma x_i^2}\begin{pmatrix}\lambda+\Sigma x_i^2 & -\Sigma x_i^2\\ -\lambda_i^2 & \lambda+\Sigma x_i^2\end{pmatrix}\cdot\Sigma xy\begin{pmatrix}1\\ 1\end{pmatrix} = \Sigma xy\cdot\frac{1}{\lambda^2 + 2\lambda\Sigma x_i^2}\begin{pmatrix}\lambda\\ \lambda\end{pmatrix} = \frac{\Sigma xy}{\lambda + 2\Sigma x_i^2}\begin{pmatrix}1\\ 1\end{pmatrix}$$

$$\Rightarrow \hat\beta_1 = \hat\beta_2 = \frac{\Sigma x_iy_i}{\lambda + 2\Sigma x_i^2}$$

b). as $\lambda \to \infty$ both estimates go to 0: $\hat\beta_1 \to 0$, $\hat\beta_2 \to 0$ as $\lambda \to \infty$

c) as $\lambda \to 0$: $\hat\beta_1 \to \dfrac{\Sigma x_iy_i}{2\Sigma x_i^2}$; $\hat\beta_2 \to \dfrac{\Sigma x_iy_i}{2\Sigma x_i^2} \Rightarrow \hat\beta_1 + \hat\beta_2 \to \dfrac{\Sigma x_iy_i}{\Sigma x_i^2}$

2) $y = X\beta + u$, β is non-random; $E(u|X) = 0$ $X = [n\times k]$, $rank(X) = k$

$V(u|X) = \sigma^2 W$, $W \neq I$; $\hat\beta = (X^TX)^{-1}X^Ty$

a) $E(\hat\beta|X) = E((X^TX)^{-1}X^Ty|X) = (X^TX)^{-1}X^TE(y|X) = (X^TX)^{-1}X^TE(X\beta + u|X) = (X^TX)^{-1}X^T[X\beta + E(u|X)] =$

$= (X^TX)^{-1}X^T[X\beta + 0] = (X^TX)^{-1}X^TX\beta = \beta$

$E(\hat\beta) = E(E(\hat\beta|X)) = E(\beta) = \beta$

b) $V(\hat\beta|X) = V((X^TX)^{-1}X^Ty|X) = (X^TX)^{-1}X^T V(y|X)X(X^TX)^{-1} = (X^TX)^{-1}X^TV(X\beta+u|X)X(X^TX)^{-1} =$

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

c) no, since for normal CI $V(\hat\beta|X)$ should be $\sigma^2(X^TX)^{-1} \Rightarrow$

$CI = \hat\beta \pm se\hat\beta \cdot z_{\alpha/2} \neq \hat\beta \pm \sigma\sqrt{(X^TX)^{-1}}\, z_{\alpha/2}$ as $\sqrt{(X^TX)^{-1}} \neq \sqrt{(X^TX)^{-1}X^TWX(X^TX)^{-1}}$

unless X^TWX is some kind of pseudo-inverse matrix of $(X^TX)^{-1}$

d) $Cov(y,\hat\beta|X) = Cov(y,(X^TX)^{-1}X^Ty|X) = Cov(y,y|X)\,X(X^TX)^{-1} = V(y|X)\,X(X^TX)^{-1} = V(\beta X+u|X)X(X^TX)^{-1}$

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

3) $X = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}$ a) $X^TX = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$; $|X^TX - \lambda I| = (6-\lambda)^2 - 1 = \lambda^2 - 12\lambda + 35 =$
$= (\lambda-7)(\lambda-5)$

$\lambda=5$: $v_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$; $\lambda = 7$: $v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \Rightarrow X^TX = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 5 & 0 \\ 0 & 7 \end{pmatrix}\begin{pmatrix} -\frac12 & \frac12 \\ \frac12 & \frac12 \end{pmatrix}$

b) $X = UDV^T$; $X^TX = VD^T\underbrace{\cancel{U^TU}}_{I}DV^T = VD^TDV^T \Rightarrow D = D^T = \begin{pmatrix} \sqrt7 & 0 \\ 0 & \sqrt5 \end{pmatrix}$;

to make V orthogonal: $\begin{pmatrix} -1 \\ 1 \end{pmatrix} : \|v_1\| = \begin{pmatrix} -1/\sqrt2 \\ 1/\sqrt2 \end{pmatrix}$; $\begin{pmatrix} 1 \\ 1 \end{pmatrix} : \|v_2\| = \begin{pmatrix} 1/\sqrt2 \\ 1/\sqrt2 \end{pmatrix}$; $V = \begin{pmatrix} 1/\sqrt2 & -1/\sqrt2 \\ 1/\sqrt2 & 1/\sqrt2 \end{pmatrix}$ since $5 \overset{?}{<} 7$; $V^T = \begin{pmatrix} 1/\sqrt2 & 1/\sqrt2 \\ -1/\sqrt2 & 1/\sqrt2 \end{pmatrix}$

to find U: $XX^T = UDV^TVD^TU^T = UDD^TU^T$

$XX^T = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix}$; $|XX^T - \lambda I| = (5-\lambda)^2(2-\lambda) - (5-\lambda) - 9(5-\lambda) =$

$= (5-\lambda)(\lambda^2 - 7\lambda + 10 - 1 - 9) = (5-\lambda)(7-\lambda)\lambda$

$\Rightarrow \lambda = 7$: $\begin{pmatrix} -2 & 0 & 3 \\ 0 & -2 & 1 \\ 3 & 1 & -5 \end{pmatrix} \to \begin{pmatrix} -2 & 0 & 3 \\ 0 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix} \to \begin{pmatrix} 0 & 2 & -1 \\ 0 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix} \to \begin{pmatrix} 0 & 2 & -1 \\ 1 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}$; $u_1 = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} \cdot \frac{1}{\sqrt{14}}$

$\lambda = 5$: $\begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 1 \\ 3 & 1 & -3 \end{pmatrix}$; $u_2 = \begin{pmatrix} 1 \\ -3 \\ 0 \end{pmatrix} \cdot \frac{1}{\sqrt{10}}$; $\lambda = 0$: $\begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix} \to \begin{pmatrix} 2 & -1 & 1 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix} \to \begin{pmatrix} 2 & -1 & 1 \\ 0 & 5 & 1 \\ 1 & 2 & 1 \end{pmatrix} \to \begin{pmatrix} 0 & 5 & 1 \\ 1 & 2 & 1 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow$

$\Rightarrow u_3 = \begin{pmatrix} 3 \\ 1 \\ -5 \end{pmatrix} \cdot \frac{1}{\sqrt{35}} \Rightarrow X = \begin{pmatrix} \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{35}} \\ \frac{1}{\sqrt{14}} & \frac{-3}{\sqrt{10}} & \frac{1}{\sqrt{35}} \\ \frac{2}{\sqrt{14}} & 0 & \frac{-5}{\sqrt{35}} \end{pmatrix} \cdot \begin{pmatrix} \sqrt7 & 0 \\ 0 & \sqrt5 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1/\sqrt2 & \frac{1}{\sqrt2} \\ \frac{-1}{\sqrt2} & \frac{1}{\sqrt2} \end{pmatrix}$

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

4) X - standartized; $X = UDV^T$ $\begin{cases} \|Xw\|^2 \to \max\limits_{w} \\ \text{s.t.} \\ \|w\|^2 = 1 \end{cases}$

d) $L = \|Xw\|^2 + \lambda(1 - \|w\|^2)$; since $\|a\|^2 = \langle a, a \rangle$

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

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

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

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

$\Rightarrow w^T (X^T X) = \lambda w^T \Rightarrow (X^T X) w = \lambda w$ F.O.C. are $\begin{cases} X^T X w = \lambda w \\ w^T w = 1 \end{cases}$

f). it is known from e) that w is an eigenvector of $X^T X$, since

$X^T X = U D^T U^T U D V^T = V \underbrace{D^T D}_{\text{Diagonal}} V^T$; columns of V are eigenvectors of $X^T X$, while

$d_{11}, d_{22} \ldots$ are eigenvalues;

From the useful properties and applications of Rayleigh quotient the best eigenvector should correspond to the highest eigenvalue $\Rightarrow$ 1_{st} column of V is w^*.