Шлев А.Д 3824М1ПМав

№1) 1) $a \in \mathbb{R}^n, x \in \mathbb{R}^n$ $\frac{\partial(a^T x)}{\partial x} = a$?

$\exists\ g(x) = a^T x \Rightarrow \frac{\partial g}{\partial x} = grad(g) = grad(a^T x) = a^T$
$(g : \mathbb{R}^n \to \mathbb{R}^1)$

2) $A \in \mathbb{R}^{m \times n}, x \in \mathbb{R}^n$ $\frac{\partial(Ax)}{\partial x} = A$?

$\exists\ \begin{matrix} g(x) = Ax \\ g : \mathbb{R}^n \to \mathbb{R}^m \end{matrix} \Rightarrow g_i = \sum_{j=1}^{n} a_{ij} x_j \Rightarrow \frac{\partial g_i}{\partial x_k} = a_{ik} \Rightarrow \frac{\partial g}{\partial x} = A = \frac{\partial(Ax)}{\partial x}$

3) $A \in \mathbb{R}^{n \times n}, x \in \mathbb{R}^n$ $\frac{\partial(x^T A x)}{\partial x} = (A + A^T)x$? $\left(A^T = A \Rightarrow \frac{\partial(x^T A x)}{\partial x} = 2Ax\ ?\right)$

$\exists\ \begin{matrix} g(x) = x^T A x \\ g : \mathbb{R}^n \to \mathbb{R}^1 \end{matrix} \Rightarrow \frac{\partial g}{\partial x} = grad(g)$

$g(x) = x^T A x = \sum_{j=1}^{n} \sum_{i=1}^{n} x_j x_i a_{ij} \Rightarrow \frac{\partial g}{\partial x_k} = \sum_{i=1}^{n} x_i a_{ik} + \sum_{j=1}^{n} x_j a_{kj} \Rightarrow$

$\Rightarrow grad(g) = \left(\sum_{i=1}^{n} x_i a_{i1} \ldots \sum_{i=1}^{n} x_i a_{in}\right) + \left(\sum_{j=1}^{n} x_j a_{1j} \ldots \sum_{j=1}^{n} x_j a_{nj}\right) =$

$= x^T A^T + x^T A = x^T(A^T + A)$

$\Rightarrow$, если $A^T = A$ $\frac{\partial(x^T A x)}{\partial x} = 2x^T A$

4) $x \in \mathbb{R}^n$ $\frac{\partial \|x\|^2}{\partial x} = 2x$?

$\exists\ \begin{matrix} g(x) = \|x\|^2 \\ g : \mathbb{R}^n \to \mathbb{R}^1 \end{matrix} \Rightarrow \frac{\partial g}{\partial x} = grad(g) = grad(\|x\|^2) = grad\left(\sum_{i=1}^{n} x_i^2\right) =$
$= 2x^T$

5) g - скалярная ф-я, $x \in \mathbb{R}^n$ $\frac{\partial g(x)}{\partial x} = diag(g'(x))$?

$g(x) = \begin{pmatrix} g(x_1) \\ \vdots \\ g(x_n) \end{pmatrix} \Rightarrow g_i = g(x_i) \Rightarrow$

$g : \mathbb{R}^n \to \mathbb{R}^n$ $\frac{\partial g_i}{\partial x_k} = \frac{\partial g(x_i)}{\partial x_k} = \begin{cases} 0 & i \neq k \\ g'(x_i) & i = k \end{cases}$

$\Rightarrow \frac{\partial \varphi}{\partial x} = (g_{ij})$, где $g_{ij} = \begin{cases} 0 & i \neq j \\ g'(x_i) & i = j \end{cases} = diag(g'(x))$

б) $h : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^m \to \mathbb{R}^p$, $x \in \mathbb{R}^n$ $\quad \frac{\partial g(h(x))}{\partial x} = \frac{\partial g(h(x))}{\partial h} \frac{\partial h(x)}{\partial x}$?

$G(x) = g(h(x))$

Г $G : \mathbb{R}^n \to \mathbb{R}^p$ $\quad \Rightarrow \cancel{\frac{\partial G}{\partial x} = \partial(g(h(x)))}$

$$\Rightarrow \frac{\partial G_i}{\partial x_k} = \frac{\partial g_i(h(x))}{\partial x_k} = \frac{\partial g_i(h(x))}{\partial h} \frac{\partial h(x)}{\partial x_k} \Rightarrow \frac{\partial g(h(x))}{\partial x} = \left(\frac{\partial G_i}{\partial x_j}\right) \in \mathbb{R}^{p \times n}$$

$$\frac{\partial h(x)}{\partial x} = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \dots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1} & \dots & \frac{\partial h_m}{\partial x_n} \end{pmatrix} = \left(\frac{\partial h(x)}{\partial x_1} \dots \frac{\partial h(x)}{\partial x_n} \right)_{m \times n}$$

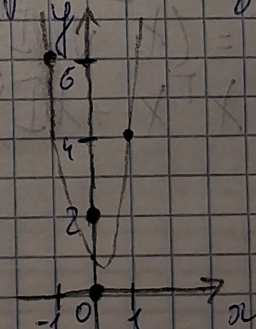
$$\frac{\partial g(h(x))}{\partial h} = \begin{pmatrix} \frac{\partial g_1}{\partial h_1} & \dots & \frac{\partial g_1}{\partial h_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_p}{\partial h_1} & \dots & \frac{\partial g_p}{\partial h_m} \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1(h(x))}{\partial h} \\ \vdots \\ \frac{\partial g_p(h(x))}{\partial h} \end{pmatrix}_{p \times m}$$

$$\Rightarrow \frac{\partial g(h(x))}{\partial h} \frac{\partial h(x)}{\partial x} = \left(\frac{\partial g_i(h(x))}{\partial h} \frac{\partial h(x)}{\partial x_j} \right) = \left(\frac{\partial G_i}{\partial x_j} \right) \in \mathbb{R}^{p \times n}$$

№3

x	1	1	0	0	-1
y	4	4	0	2	6

1)



2) МНК $f(x) = b_0 + b_1 x + b_2 x^2$

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{pmatrix} \quad y = \begin{pmatrix} 4 \\ 4 \\ 0 \\ 2 \\ 6 \end{pmatrix} \quad X^T X b = X^T y$$

$$\Rightarrow \begin{pmatrix} 5 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 3 \end{pmatrix} b = \begin{pmatrix} 16 \\ 2 \\ 14 \end{pmatrix} \Rightarrow b_0 = 1 \quad b_1 = -1 \quad b_2 = 4$$

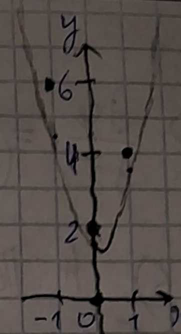
график

3) ридж-регрессия с $\lambda = 1$

$$(X^T X + \lambda I) b = X^T y$$

$$\begin{pmatrix} 6 & 1 & 3 \\ 1 & 4 & 1 \\ 3 & 1 & 4 \end{pmatrix}$$

$$\Rightarrow f(x) = \frac{3}{2} - \frac{1}{2}x + \frac{5}{2}x$$



№4 $y \sim N(Xb, \sigma^2 I)$ ~~$E\varepsilon_1 = 0$ $E\varepsilon_1 = \sigma^2 I$~~

$b \sim N(0, \tau I)$

~~$\Rightarrow y = X(b + \varepsilon_1) + \varepsilon_2$~~

$$\hat{b}^{ridge} = (X^T X + \lambda I)^{-1} X^T y$$

$\exists\, \varepsilon_1$: ~~$E\varepsilon_1 = 0$, $\operatorname{cov}\varepsilon_1 = \sigma^2 I$~~ $\varepsilon_1 \sim N(0, \sigma^2 I)$

$$\Rightarrow y = Xb + \varepsilon_1 \Rightarrow \hat{b} = (X^T X)^{-1} X^T (Xb + \varepsilon_1)$$

№5 МНК $\tilde{X} = \begin{pmatrix} X \\ \sqrt{\lambda} I \end{pmatrix}$ $\tilde{y} = \begin{pmatrix} y \\ 0 \end{pmatrix}$ $\lambda > 0$

$$\Rightarrow \tilde{y} = \tilde{X}\hat{b} \qquad \tilde{X}^T \tilde{X} = \begin{pmatrix} X^T & \sqrt{\lambda} I \end{pmatrix} \begin{pmatrix} X \\ \sqrt{\lambda} I \end{pmatrix} = X^T X + \lambda I -$$

– невырождена и положительно опр-на

$$\Rightarrow \tilde{X}^T \tilde{X} \hat{b} = \tilde{X}^T \tilde{y} \Rightarrow \hat{b} = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{y}$$

$$\tilde{X}^T \tilde{y} = \begin{pmatrix} X^T & \sqrt{\lambda} I \end{pmatrix} \begin{pmatrix} y \\ 0 \end{pmatrix} = X^T y \Rightarrow$$

$$\Rightarrow \hat{b} = (X^T X + \lambda I)^{-1} X^T y = \hat{b}^{ridge}$$

№4 $y \sim N(Xb, \sigma^2 I)$ $\quad b \sim N(0, \tau I)$

$$f_b(x) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|\tau I|}} e^{-\frac{1}{2}x^T(\tau I)^{-1}x} = \frac{1}{(2\pi)^{\frac{n}{2}}\tau^{\frac{n}{2}}} e^{-\frac{1}{2\tau}x^T x}$$

$$f_{y|b=x}(y) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|\sigma^2 I|}} e^{-\frac{1}{2}(y-Xx)^T(\sigma^2 I)^{-1}(y-Xx)} = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} e^{-\frac{1}{2\sigma^2}(y-Xx)^T(y-Xx)}$$

$$\Rightarrow f_{b|y=y}(x) = \frac{f_{y|b=x}(y) f_b(x)}{\int_{\mathbb{R}^n} f_{y|b=\xi}(y) f_b(\xi) d\xi}$$

$$\int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} e^{-\frac{1}{2\sigma^2}(y-X\xi)^T(y-X\xi)} \cdot \frac{1}{(2\pi)^{\frac{n}{2}}\tau^{\frac{n}{2}}} e^{-\frac{1}{2\tau}\xi^T\xi} d\xi$$

$\Rightarrow f_{b|y=y}(x)$ – нормальное распред $\Rightarrow$

$$f_{b|y=y}(\mu) = \frac{e^{-\frac{1}{2}\left(\frac{\mu^T\mu}{\tau} + \frac{1}{\sigma^2}(y-X\mu)^T(y-X\mu)\right)}}{\int_{\mathbb{R}^n} e^{-\frac{1}{2\sigma^2}(y-X\xi)^T(y-X\xi) - \frac{1}{2\tau}\xi^T\xi} d\xi} = \frac{1}{\int_{\mathbb{R}^n} e^{-\frac{1}{2}\left(\frac{\xi^T\xi}{\tau} + \frac{1}{\sigma^2}(y-X\xi)^T(y-X\xi)\right)} d\xi}$$

μ – мат. ожид.

$\Rightarrow \frac{\mu^T\mu}{\tau} + \frac{1}{\sigma^2}(y-X\mu)^T(y-X\mu) = 0$ причём μ – т. макс.

$$\Rightarrow \frac{\partial}{\partial\mu}\left(\frac{\mu^T\mu}{\tau} + \frac{1}{\sigma^2}(y-X\mu)^T(y-X\mu)\right) = \frac{2}{\tau}\mu^T + \frac{2}{\sigma^2}(y-X\mu)^T(-X) = 0$$

$$\Rightarrow \frac{\sigma^2}{\tau}\mu + X^T(X\mu - y) = 0 \Rightarrow \left(X^TX + \frac{\sigma^2}{\tau}I\right)\mu = X^Ty$$

$\Rightarrow \mu = b^{ridge}$ с $\lambda = \frac{\sigma^2}{\tau}$