Гимаев Виктор,
3824М1ПМвм

Домашняя работа №2

№1

Докажем 1-6:

① если $a \in \mathbb{R}^n$, $x \in \mathbb{R}^n$, то $\frac{\partial(a^T x)}{\partial x} = a$

$a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$, $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\frac{\partial(a^T x)}{\partial x_i} = a_i$, $i = \overline{1,n}$,

$a^T x = \sum_{i=1}^{n} a_i x_i$

Представим $\frac{\partial(a^T x)}{\partial x}$ как вектор:

$$\frac{\partial(a^T x)}{\partial x} = \Big(\underbrace{\frac{\partial(a^T x)}{\partial x_1}}_{= a_1}, \underbrace{\frac{\partial(a^T x)}{\partial x_2}}_{= a_2}, \dots, \underbrace{\frac{\partial(a^T x)}{\partial x_n}}_{= a_n}\Big) =$$

$$= (a_1, a_2, \dots, a_n) = a^T$$

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

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad Ax = \begin{pmatrix} \sum_{i=1}^{n} a_{1i} x_i \\ \vdots \\ \sum a_{mi} x_i \end{pmatrix},$$

$$\frac{\partial(Ax)}{\partial x_i} = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{pmatrix}, \quad i = \overline{1,n}$$

Тогда:

$$\frac{\partial(Ax)}{\partial x} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = A$$

③ если $A \in \mathbb{R}^{n \times n}$, $x \in \mathbb{R}^n$, то $\frac{\partial(x^T A x)}{\partial x} = (A + A^T)$;
в частности, если $A^T = A$, то $\frac{\partial(x^T A x)}{\partial x} = 2Ax$.

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},\ A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix},\ x^T A x = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j$$

$$\frac{\partial (x^T A x)}{\partial x_1} = \sum_{i=2}^{n} a_{i1} x_i + 2 a_{11} x_1 + \sum_{j=2}^{n} a_{1j} x_j$$

$$\frac{\partial (x^T A x)}{\partial x_k} = \sum_{i \neq k} a_{ik} x_i + 2 a_{kk} x_k + \sum_{j \neq k} a_{kj} x_j = \sum_{i=1}^{n} a_{ik} x_i + \sum_{j=1}^{n} a_{kj} x_j$$

$$\frac{\partial (x^T A x)}{\partial x} = \begin{pmatrix} \sum_{i=1}^{n} a_{i1} x_i + \sum_{j=1}^{n} a_{1j} x_j \\ \dots \\ \sum_{i=1}^{n} a_{in} x_i + \sum_{j=1}^{n} a_{nj} x_j \end{pmatrix}^T$$

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

$$x^T A^T = (x_1, \dots, x_n) \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{pmatrix} = \left(\sum_{i=1}^{n} x_i a_{1i}, \dots, \sum_{i=1}^{n} x_i a_{ni}\right)$$

$$x^T A = (x_1, \dots, x_n) \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} = \left(\sum_{i=1}^{n} x_i a_{i1}, \dots, \sum_{i=1}^{n} x_i a_{in}\right)$$

$$x^T (A^T + A) = \left(\sum_{i=1}^{n} x_i a_{i1} + \sum_{j=1}^{n} x_j a_{1j}, \dots, \sum_{i=1}^{n} x_i a_{in} + \sum_{j=1}^{n} x_j a_{nj}\right) = \frac{\partial (x^T A x)}{\partial x}$$

Т.к. данная формула справедлива, то $A^T = A$:

$$\frac{\partial (x^T A x)}{\partial x} = x^T (A^T + A) = x^T (2A) = 2 x^T A$$

④ Если $x \in \mathbb{R}^n$, то $\frac{\partial \|x\|^2}{\partial x} = 2x$

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},\ \|x\|^2 = (x, x) = x_1^2 + x_2^2 + \dots + x_n^2$$

$$\frac{\partial \|x\|^2}{\partial x_i} = 2x_i,\ \frac{\partial \|x\|^2}{\partial x} = (2x_1, 2x_2, \dots, 2x_n) = 2(x_1, \dots, x_n) = 2x^T.$$

⑤ Если g - скалярная ф-ия и под $g(x)$ понимается применение ф-ии g к каждой компоненте вектора $x \in \mathbb{R}^n$, то

$$\frac{\partial g(x)}{\partial x} = diag(g'(x)),$$

где $diag(a)$ - диагональная матрица с диагональю a.

$$g(x) = \begin{pmatrix} g(x_1) \\ g(x_2) \\ \vdots \\ g(x_n) \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \frac{\partial g(x)}{\partial x} = \begin{pmatrix} \frac{\partial g(x_1)}{\partial x_1} & \frac{\partial g(x_1)}{\partial x_2} & \cdots & \frac{\partial g(x_1)}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial g(x_n)}{\partial x_1} & \frac{\partial g(x_n)}{\partial x_2} & \cdots & \frac{\partial g(x_n)}{\partial x_n} \end{pmatrix} =$$

$$= \begin{pmatrix} g'(x_1) & 0 & \cdots & 0 \\ 0 & g'(x_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g'(x_n) \end{pmatrix} = diag(g'(x))$$

т.к. $g(x_i)$ не зависит от x_j, где $i \neq j$, то получаем нули везде, кроме главной диагонали.

⑥ Если $h: \mathbb{R}^n \to \mathbb{R}^m$, $g: \mathbb{R}^m \to \mathbb{R}^p$, $x \in \mathbb{R}^n$, то

$$\frac{\partial g(h(x))}{\partial x} = \frac{\partial g(h(x))}{\partial h} \frac{\partial h(x)}{\partial x}$$

$$h = \begin{pmatrix} h_1(x_1, \ldots, x_n) \\ h_2(x_1, \ldots, x_n) \\ \vdots \\ h_m(x_1, \ldots, x_n) \end{pmatrix}, \quad g = \begin{pmatrix} g_1(h_1, \ldots, h_m) \\ g_2(h_1, \ldots, h_m) \\ \vdots \\ g_p(h_1, \ldots, h_m) \end{pmatrix}$$

$$\frac{\partial g(h(x))}{\partial h} = \begin{pmatrix} \frac{\partial g_1(h(x))}{\partial h_1} & \frac{\partial g_1(h(x))}{\partial h_2} & \cdots & \frac{\partial g_1(h(x))}{\partial h_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_p(h(x))}{\partial h_1} & \frac{\partial g_p(h(x))}{\partial h_2} & \cdots & \frac{\partial g_p(h(x))}{\partial h_m} \end{pmatrix}$$

$$\frac{\partial h(x)}{\partial x} = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1} & \frac{\partial h_m}{\partial x_2} & \cdots & \frac{\partial h_m}{\partial x_n} \end{pmatrix}$$

$$\frac{\partial g(h(x))}{\partial x} = \begin{pmatrix} \sum\limits_{i=1}^{m} \frac{\partial g_1}{\partial h_i}\frac{\partial h_i}{\partial x_1} & \dots & \sum\limits_{i=1}^{m} \frac{\partial g_1}{\partial h_i}\frac{\partial h_i}{\partial x_n} \\ \vdots & & \\ \sum\limits_{i=1}^{m} \frac{\partial g_p}{\partial h_i}\frac{\partial h_i}{\partial x_1} & \dots & \sum\limits_{i=1}^{m} \frac{\partial g_p}{\partial h_i}\frac{\partial h_i}{\partial x_n} \end{pmatrix}$$

$$\frac{\partial g(h(x))}{\partial h}\frac{\partial h(x)}{\partial x} = \begin{pmatrix} \sum\limits_{i=1}^{m} \frac{\partial g_1}{\partial h_i}\frac{\partial h_i}{\partial x_1} & \dots & \sum\limits_{i=1}^{m} \frac{\partial g_1}{\partial h_i}\frac{\partial h_i}{\partial x_n} \\ \vdots & & \\ \sum\limits_{i=1}^{m} \frac{\partial g_p}{\partial h_i}\frac{\partial h_i}{\partial x_1} & \dots & \sum\limits_{i=1}^{m} \frac{\partial g_p}{\partial h_i}\frac{\partial h_i}{\partial x_n} \end{pmatrix}$$

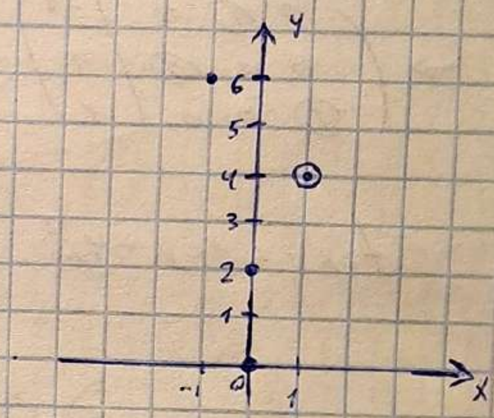
Получим ту же самую матрицу. $\Rightarrow$

$$\Rightarrow \frac{\partial g(h(x))}{\partial h}\frac{\partial h(x)}{\partial x} = \frac{\partial g(h(x))}{\partial x}.$$

№3

①

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



② $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \quad y = \begin{pmatrix} 4 \\ 4 \\ 0 \\ 2 \\ 6 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 3 \end{pmatrix}$$

$$X^T y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 4 \\ 0 \\ 2 \\ 6 \end{pmatrix} = \begin{pmatrix} 16 \\ 2 \\ 14 \end{pmatrix}$$

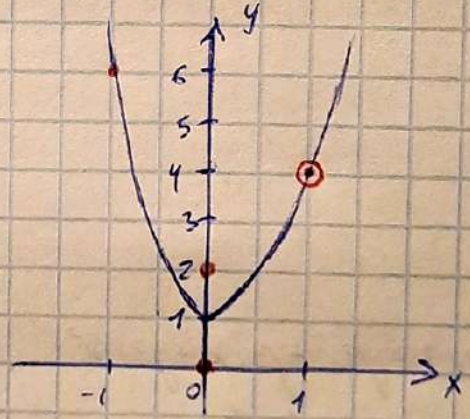
$$X^T X \beta = X^T y \qquad \begin{pmatrix} 5 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 3 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 16 \\ 2 \\ 14 \end{pmatrix}$$

$$\left(\begin{array}{ccc|c} 5 & 1 & 3 & 16 \\ 1 & 3 & 1 & 2 \\ 3 & 1 & 3 & 14 \end{array}\right) \xrightarrow[(3)-(1)]{(2)-3(1)} \left(\begin{array}{ccc|c} 5 & 1 & 3 & 16 \\ -14 & 0 & -8 & -46 \\ -2 & 0 & 0 & -2 \end{array}\right) \xrightarrow[-\frac{1}{2}(3)]{\frac{1}{2}(2)}$$

$$\rightarrow \left(\begin{array}{ccc|c} 5 & 1 & 3 & 16 \\ -7 & 0 & -4 & -23 \\ 1 & 0 & 0 & 1 \end{array}\right) \xrightarrow[(1)-5(3)]{(2)+7(3)} \left(\begin{array}{ccc|c} 0 & 1 & 3 & 11 \\ 0 & 0 & -4 & -16 \\ 1 & 0 & 0 & 1 \end{array}\right) \xrightarrow[(1)+3(2)]{-\frac{1}{4}(2)}$$

$$\rightarrow \left(\begin{array}{ccc|c} 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 4 \\ 1 & 0 & 0 & 1 \end{array}\right) \qquad \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 4 \end{pmatrix}$$

$$f(x) = 1 - x + 4x^2$$



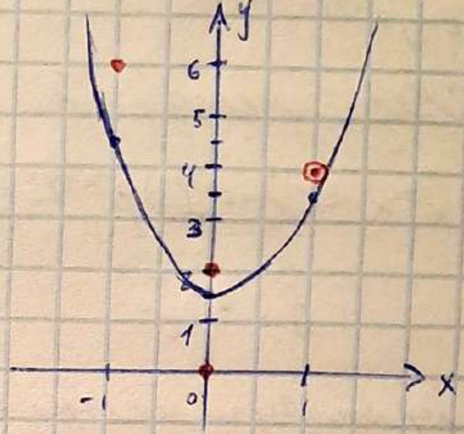
③ $\lambda = 1$, $(X^T X + \lambda I)\beta = X^T y$

$$X^T X + \lambda I = \begin{pmatrix} 5 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 3 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 1 & 3 \\ 1 & 4 & 1 \\ 3 & 1 & 4 \end{pmatrix}$$

$$\left(\begin{array}{ccc|c} 6 & 1 & 3 & 16 \\ 1 & 4 & 1 & 2 \\ 3 & 1 & 4 & 14 \end{array}\right) \xrightarrow[(3)-(1)]{(2)-4(1)} \left(\begin{array}{ccc|c} 6 & 1 & 3 & 16 \\ -23 & 0 & -11 & -62 \\ -3 & 0 & 1 & -2 \end{array}\right) \xrightarrow[(1)-3(3)]{(2)+11(3)} \left(\begin{array}{ccc|c} 15 & 1 & 0 & 22 \\ -56 & 0 & 0 & -84 \\ -3 & 0 & 1 & -2 \end{array}\right) \rightarrow$$

$$\to \left(\begin{array}{ccc|c} 15 & 1 & 0 & 22 \\ 2 & 0 & 0 & 3 \\ -3 & 0 & 1 & -2 \end{array}\right) \xrightarrow[(3)+3(2)]{\frac{1}{2}(2)\;(1)-15(2)} \left(\begin{array}{ccc|c} 0 & 1 & 0 & -1/2 \\ 1 & 0 & 0 & 3/2 \\ 0 & 0 & 1 & 3/2 \end{array}\right) \quad \begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\end{pmatrix} = \begin{pmatrix}3/2\\ -1/2\\ 5/2\end{pmatrix}$$

$$f(x) = 1{,}5 - 0{,}5x + 2{,}5x^2$$



№ 4

$y \sim N(X\beta, \sigma^2 I)$ – норм. распр.
$\beta \sim N(0, \tau I)$ – априорное распр.

① Апостериорное распр-е для β.

$$P(\beta|y) \sim P(y|\beta)\pi(\beta) = \exp\left(-\tfrac{1}{2}(y-X\beta)^T \tfrac{1}{\sigma^2}(y-X\beta)\right) \cdot \exp\left(-\tfrac{1}{2}(\beta-0)^T \tfrac{1}{\tau} I(\beta-0)\right) = \underline{\exp\left(-\tfrac{1}{2}\sigma^2\|y-X\beta\|^2 - \tfrac{1}{2}\tau^2\|\beta\|^2\right)}$$

A

Получили апостер. распр. β.

② Мат. ожид-е апост. распр. β ≤ β^{ridge}

$$E = \arg\max_\beta P(y|\beta)\pi(\beta) = \arg\max_\beta A =$$

$$= \arg\max(\log A) = \arg\max\left(-\tfrac{1}{2}\sigma^2\|y-X\beta\|^2 - \tfrac{1}{2}\tau^2\|\beta\|^2\right) =$$

$$= \arg\max -\left(\tfrac{1}{2}\sigma^2\|y-X\beta\|^2 + \tfrac{1}{2}\tau^{-2}\|\beta\|^2\right) =$$

$$= \arg\min\left(\tfrac{1}{2}\sigma^2\|y-X\beta\|^2 + \tfrac{1}{2}\tau^2\|\beta\|^2\right) =$$

$$= \arg\min \tfrac{1}{\sigma^2}(\beta^T X^T X\beta - 2y^T X\beta) + \tfrac{1}{\tau^2}\beta^T\beta =$$

$$= \arg\min \beta^T(X^T X + \sigma^2/\tau^2 I)\beta - 2\beta^T X^T y$$

Приравняем градиенты к 0:

$$2(X^TX + \sigma^2/\tau^2 I)\beta^{ridge} = X^Ty$$

$E = \beta^{ridge}$

③ Связь между λ и τ, σ^2.

$$(X^TX + \underbrace{\sigma^2/\tau^2}_{\Downarrow} I)\beta^{ridge} = X^Ty$$

$$\lambda = \sigma^2/\tau^2$$

λ пропорционально σ^2 и обратно пропорц. τ^2.

№5

$$\beta^{ridge} = (X^TX + \lambda I)^{-1}X^Ty$$

Расшир. матр. X и расшир. вектор y:

$\tilde{X} = \begin{pmatrix} X \\ \sqrt{\lambda} I \end{pmatrix}$, $\tilde{y} = \begin{pmatrix} y \\ 0 \end{pmatrix}$, где $\tilde{X} \in R^{(n+d)\times d}$, $\tilde{y} \in R^{n+d}$

Применяем метод наим. квадратов:

$$\tilde{X}^T\tilde{X}\hat{\beta} = \tilde{X}^T\tilde{y}$$

1) $\tilde{X}^T\tilde{X} = (X^T \ \sqrt{\lambda} I)\begin{pmatrix} X \\ \sqrt{\lambda} I \end{pmatrix} = X^TX + \lambda I$

2) $\tilde{X}^T\tilde{y} = (X^T \ \sqrt{\lambda} I)\begin{pmatrix} y \\ 0 \end{pmatrix} = X^Ty$

Тогда $\tilde{X}^T\tilde{X}\hat{\beta} = \tilde{X}^T\tilde{y} \Leftrightarrow (X^TX + \lambda I)\hat{\beta} = X^Ty$ и:

$$\hat{\beta} = (X^TX + \lambda I)^{-1}X^Ty$$

$$\hat{\beta} = \beta^{ridge}$$