Домашнее задание № 4

№1

$g: \mathbb{R}^n \to \mathbb{R}^m$, $x \in \mathbb{R}^n$

$\frac{\partial g}{\partial x}$ - матрица Якоби

Докажем 1-6

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

$a = (a_1, \dots, a_n)^T$
$x = (x_1, \dots, x_n)^T$

$\frac{\partial(a^T x)}{\partial x_i} = a_i \quad i = \overline{1,n}$

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

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

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

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

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad Ax = \begin{pmatrix} \sum_{i=1}^{n} a_{1i} x_i \\ \vdots \\ \sum_{i=1}^{n} a_{mi} x_i \end{pmatrix}$$ тогда

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

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

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

$$x = (x_1, \dots, x_n)^T$$

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \qquad x^T A x = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j$$

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

$$\dots$$

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

$$\ominus \sum_{i=1}^{n} a_{ik} x_i + \sum_{j=1}^{n} a_{kj} x_j$$

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

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

$$x^T A^T = (x_1, \dots, x_n) \cdot \begin{pmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1m} & \cdots & a_{mn} \end{pmatrix} \ominus$$

$$\ominus \left(\sum_{i=1}^{n} x_i a_{1i}, \dots, \sum_{i=1}^{n} x_i a_{ni} \right)$$

$$x^T A = \dots = \left(\sum_{i=1}^{n} x_i a_{i1}, \dots, \sum_{i=1}^{n} x_i a_{in} \right)$$

$$x^T A^T + x^T A = Vec + Vec = \left(\sum_{i=1}^{n} x_i a_{i1} + \sum_{j=1}^{n} x_j a_{1j}, \dots, \right.$$

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

т.к. формула справедлива, то при $A^T = A$

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

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

$$\|x\|^2 = (x, x) = x_1^2 + x_2^2 + \ldots + x_n^2$$

$$\frac{\partial \|x\|^2}{\partial x_i} = 2x_i \Rightarrow \frac{\partial \|x\|^2}{\partial x} = (2x_1, \ldots, 2x_n) \Rightarrow$$

$$\Rightarrow 2(x_1, \ldots, x_n) = 2x^T$$

⑤ если g - скаляр. функц. ..., то $\frac{\partial g(x)}{\partial x} = \text{diag}(g'(x))$

диагон. матр.

$$g(x) = (g_1(x), \ldots, g_n(x))^T$$

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

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

$$\Leftrightarrow \begin{pmatrix} g'(x_1) & & 0 \\ & \ddots & \\ 0 & & g'(x_n) \end{pmatrix} = \text{diag}(g'(x))$$

⑥ если $h: R^n \to R^m$, $g: R^m \to R^p$, $x \in R^n$, то

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

Распишем h, g

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

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

$$\frac{\partial h}{\partial x} = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_n} \\ \frac{\partial h_m}{\partial x_1} & \frac{\partial h_m}{\partial x_n} \end{pmatrix}$$

$$\frac{\partial g(h)}{\partial x} = \begin{pmatrix} \sum_{i=1}^{m} g'_{1h_i} \cdot h'_{ix_1} & \dots & \sum_{i=1}^{m} g'_{1h_i} \cdot h'_{ix_n} \\ \vdots & & \vdots \\ \sum_{i=1}^{m} g'_{ph_i} \cdot h'_{ix_1} & \dots & \sum_{i=1}^{m} g'_{ph_i} \cdot h'_{ix_n} \end{pmatrix}$$

где $g'_{1h_i} \cdot h'_{ix_1} = \frac{\partial g_1}{\partial h_i} \cdot \frac{\partial h_i}{\partial x_1}$

При перемножении $\frac{\partial g(h)}{\partial h}$ и $\frac{\partial h(x)}{\partial x}$ получим такую же матрицу, а значит ⑥ - верно.

№2.

Надо найти градиент и гессиан $g(\beta) = \|X\beta - y\|^2$ и вывести что $\hat{\beta} = \arg\min \|X\beta - y\|^2$ является решением $X^T X \beta = X^T y$

Решение:

$$\frac{\partial g(\beta)}{\partial \beta} = 2(X\beta - y)^T X$$

$$\frac{\partial^2 g(\beta)}{\partial \beta^T \partial \beta} = \frac{\partial\left(\frac{\partial g(\beta)}{\partial \beta}\right)}{\partial \beta^T} = \frac{\partial\left(2(X\beta - y)^T X\right)}{\partial \beta^T} =$$

$$= \frac{\partial\left(2\beta^T X^T X - y^T X\right)}{\partial \beta^T} = 2X^T X$$

Подозрение на экстремум $g'_\beta = 0 \Rightarrow$

$$2(X\beta - y)^T X = 0$$

$$\cancel{X\beta X^T = yX}$$

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

$X^T X \beta = X^T y$ – норм. сис. лин. ур.

В случае если X состоит из лин. нез. столб, то опр. $X^T X > 0$ т.к. $X^T X$ – матрица Грама

И тогда н.с.л.ур. имеет единств. решение β^*
И тогда по гессиану можно понять какая это точка: max или min.

$g''_{\beta^T\beta} = 2X^T X > 0 \Rightarrow X^T X > 0$ – сим. матр. с.з. $> 0 \Rightarrow$

$\Rightarrow$ точка β^* – лок. min

Посмотрим разложение в ряд Тейлора

$$g(\beta) = g(\beta^*) + (\underbrace{g'|_{\beta=\beta^*}}_{=0}, h) + \frac{1}{2}(\underbrace{g''|_{\beta=\beta^*} h, h}_{>0}) + \underbrace{\ldots}_{\to 0}$$

Тогда обобщим что $g(\beta) > g(\beta^*) \Rightarrow$

β^* - лок. min.

№3

Это было сделано в Д/З №1.

№4.

$y \sim N(X_\beta, \sigma^2 I)$ - модель.

$\beta \sim N(0, \tau I)$ - априорн. распр.

Решение:

① Апостер. распр. β

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

$$\cdot \exp\left(-\tfrac{1}{2}(\beta - 0)^T \tfrac{1}{\tau} I (\beta - 0)\right) =$$

$$= \underbrace{\exp\left(-\tfrac{1}{2\sigma^2}\|y - X\beta\|^2 - \tfrac{1}{2\tau^2}\|\beta\|^2\right)}_{A}$$ - апост. распр. β

② E ?

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

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

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

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

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

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

Приравняем градиенты к 0.

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

$E = \beta^{ridge}$

③ $\beta^{ridge} = E.$

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

и есть ответ. $\lambda = \sigma^2/\tau^2$

λ - пропорционален σ^2 и обр. проп. τ^2.