Задание 1 Матричное дифференцирование

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

Матрица Якоби

$$\frac{\partial g}{\partial x} = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \cdots & \frac{\partial g_m}{\partial x_n} \end{pmatrix}$$

Если $m = 1$, то $\frac{\partial g}{\partial x}$ – градиент функции g

Доказать, что

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

$$a = \begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_n \end{pmatrix} \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}$$

$$\frac{\partial (a^T x)}{\partial x_i} = a_i \quad i = \overline{1,n}; \; a^T x = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n$$

$$\Rightarrow \frac{\partial (a^T x)}{\partial x} = \left(\frac{\partial (a^T x)}{\partial x_1}, \frac{\partial (a^T x)}{\partial x_2}, \cdots \frac{\partial (a^T x)}{\partial x_n} \right) =$$

$$= (a_1, a_2, a_3, \ldots a_n) = a^T$$

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

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}$$

$$Ax = \begin{pmatrix} a_{11}x_1 + \dots + a_{1n}x_n \\ a_{21}x_1 + \dots + a_{2n}x_n \\ \dots \\ a_{m1}x_1 + \dots + a_{mn}x_n \end{pmatrix}$$

$$\cancel{\frac{\partial(Ax)}{\partial x}}$$

т.е $$Ax = \begin{pmatrix} \sum_{i=1}^{n} a_{1i} x_i \\ \dots \\ \sum_{i=1}^{n} a_{mi} x_i \end{pmatrix}$$

$$\frac{\partial(Ax)}{\partial x_1} = \begin{pmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{m1} \end{pmatrix}$$

$$\frac{\partial(Ax)}{\partial x_2} = \begin{pmatrix} a_{12} \\ a_{22} \\ \dots \\ a_{m2} \end{pmatrix}$$ и т.д $\Rightarrow$

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

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

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

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} \quad A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

$$x^T A x = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j$$

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

$$\sum_{i=1}^{n} a_{i1} x_i + \sum_{j=1}^{n} a_{1j} x_j \Rightarrow$$

$$\Rightarrow \quad \frac{\partial(x^TAx)}{\partial x} = \begin{pmatrix} \sum_{i=1}^{n} a_{i1}x_i + \sum_{j=1}^{n} a_{1j}x_j \\ \sum_{i=1}^{n} a_{i2}x_i + \sum_{j=1}^{n} a_{2j}x_j \\ \dots \\ \sum_{i=1}^{n} a_{in}x_i + \sum_{j=1}^{n} a_{nj}x_j \end{pmatrix}$$

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

$$A\cancel{x = (x_1, x_2, \dots, x_n)}$$

$$Ax = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} a_{1i}x_i \\ \dots \\ \sum_{i=1}^{n} a_{ni}x_i \end{pmatrix}$$

$$A^Tx = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \cdot & \cdot & \cdot & \cdot \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} a_{i1}x_i \\ \dots \\ \sum_{i=1}^{n} a_{in}x_i \end{pmatrix}$$

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

Т.к данное выражение верно, то если $A^T = A$,

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

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

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$$

$$\|x\|^2 = x_1^2 + x_2^2 + \dots + x_n^2$$

В таком случае

$$\frac{\partial \|x\|^2}{\partial x} = (2x_1, 2x_2, \dots 2x_n)^T = 2\begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = 2x$$

5) Если g – скалярная функция и под $g(x)$ понимается применение ф-ии g к каждой компоненте вектора $x \in R^n$, то $\frac{\partial g(x)}{\partial x} = \operatorname{diag}(g'(x))$, где $\operatorname{diag}(a)$ – диагональная матрица с диагональю a

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad g(x) = \begin{pmatrix} g(x_1) \\ g(x_2) \\ \vdots \\ g(x_n) \end{pmatrix}$$

$$\frac{\partial g(x)}{\partial x} = \begin{pmatrix} \frac{\partial g(x_1)}{\partial x_1} & \dots & \frac{\partial g(x_1)}{\partial (x_n)} \\ \dots & \dots & \dots \\ \frac{\partial g(x_n)}{\partial x_1} & \dots & \frac{\partial g(x_n)}{\partial (x_n)} \end{pmatrix} \Rightarrow$$

$\Rightarrow$ Все элементы матрицы, кроме эл-ов главной диагонали будут равны 0, т.е:

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

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

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

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix} \quad h = \begin{pmatrix} h_1(x) \\ h_2(x) \\ \cdots \\ h_m(x) \end{pmatrix} \quad g = \begin{pmatrix} g_1(h) \\ g_2(h) \\ \cdots \\ g_p(h) \end{pmatrix}$$

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

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

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

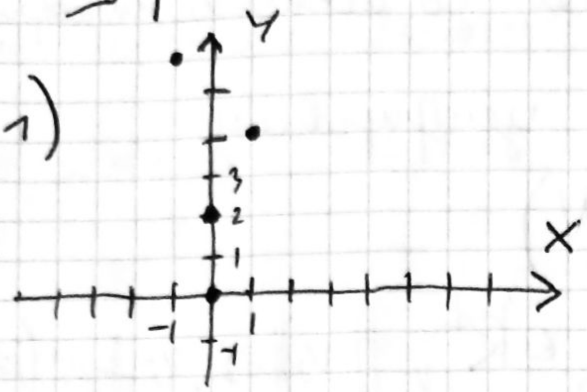
$$\frac{\partial g(h(x))}{\partial h} \cdot \frac{\partial h(x)}{\partial x} = \begin{pmatrix} \frac{\partial g_1(h(x)}{\partial h_1} & \cdots & \frac{\partial g_1(h(x))}{\partial h_m} \\ \cdots & \cdots & \cdots \\ \frac{\partial g_p(h(x)}{\partial h_1} & \cdots & \frac{\partial g_p(h(x))}{\partial h_m} \end{pmatrix} \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_n} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{m} \frac{\partial g_1}{\partial h_i} \frac{\partial h_i}{\partial x_1} & \cdots & \sum_{i=1}^{m} \frac{\partial g_1}{\partial h_i} \frac{\partial h_i}{\partial x_n} \\ \cdots & \cdots & \cdots \\ \sum_{i=1}^{m} \frac{\partial g_p}{\partial h_i} \frac{\partial h_i}{\partial x_i} & \cdots & \sum_{i=1}^{m} \frac{\partial g_p}{\partial h_i} \frac{\partial h_i}{\partial x_n} \end{pmatrix}$$

Задание 3

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

1)



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

Составим матрицу X

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{pmatrix} \qquad y = \begin{pmatrix} 4 \\ 4 \\ 0 \\ 2 \\ 6 \end{pmatrix}$$

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

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

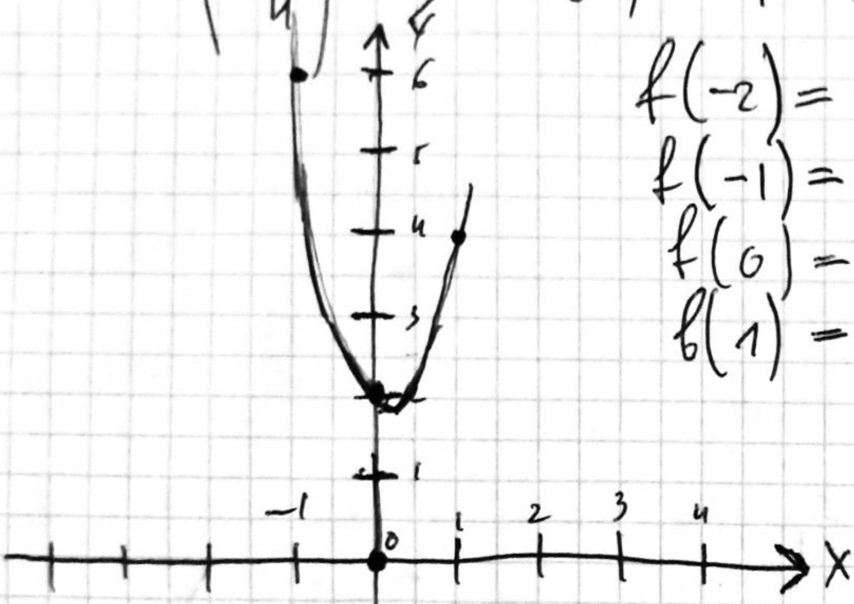
$$X^T X \beta = X^T y$$

$$\begin{pmatrix} 5 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 3 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 16 \\ 2 \\ 14 \end{pmatrix}$$

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

$$\Rightarrow \left(\begin{array}{ccc|c}0&1&3&11\\0&0&-8&-32\\1&0&0&1\end{array}\right) \Rightarrow \left(\begin{array}{ccc|c}0&1&3&11\\0&0&+1&4\\1&0&0&1\end{array}\right) \Rightarrow \left(\begin{array}{ccc|c}0&1&0&-1\\0&0&1&4\\1&0&0&1\end{array}\right)$$

$$\beta = \begin{pmatrix}1\\-1\\4\end{pmatrix} \Rightarrow f(x) = 1 - x + 4x^2$$



$f(-2) = 1 + 3 + 4 \cdot 4 = 20$

$f(-1) = 2 + 4 = 6$

$f(0) = 1$

$f(1) = 4$

Вершина парабали $x = -\frac{-1}{8} = \frac{1}{8}$

$y = -\frac{1-16}{16} = -\frac{-15}{16} = \frac{15}{16}$

3) При $\lambda = 1$ и $(X^T X + \lambda I)\beta = X^T y$, где I – ед. матр

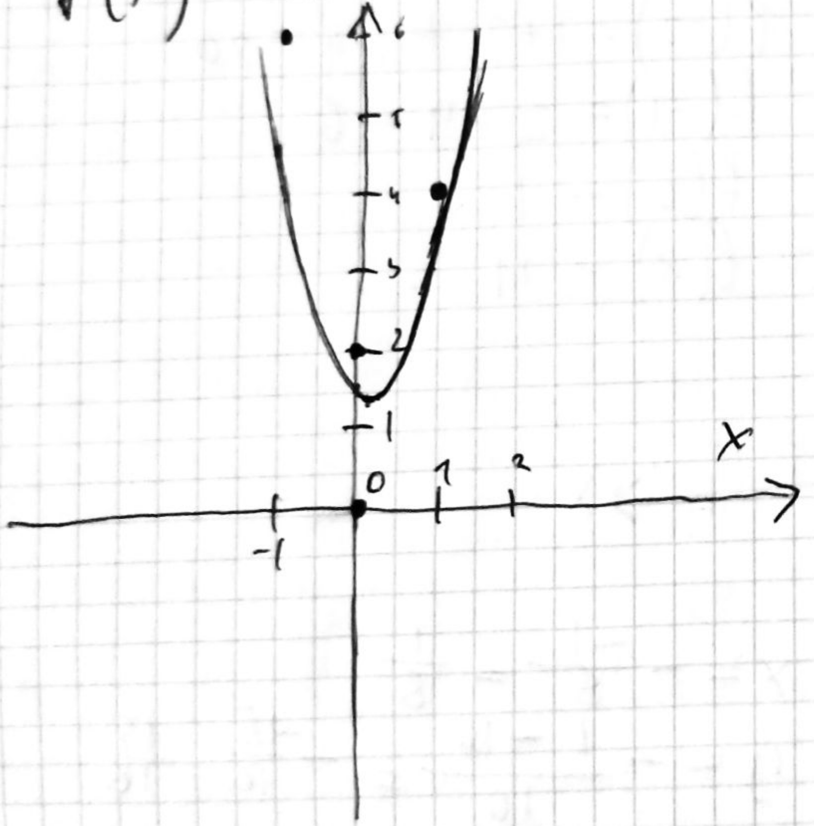
$$X^T X + \lambda I = \begin{pmatrix}5&1&3\\1&3&1\\3&1&3\end{pmatrix} + \begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} = \begin{pmatrix}6&1&3\\1&4&1\\3&1&4\end{pmatrix}$$

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

$$\Rightarrow \left(\begin{array}{ccc|c} 15 & 1 & 0 & 22 \\ 2 & 0 & 0 & 3 \\ -3 & 0 & 1 & -2 \end{array}\right) \Rightarrow \left(\begin{array}{ccc|c} 1 & 1 & 0 & 1 \\ 2 & 0 & 0 & 3 \\ -3 & 0 & 1 & -2 \end{array}\right) \Rightarrow \left(\begin{array}{ccc|c} 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & \frac{3}{2} \\ -3 & 0 & 1 & -2 \end{array}\right) \Rightarrow$$

$$\Rightarrow \left(\begin{array}{ccc|c} 0 & 1 & 0 & -\frac{1}{2} \\ 1 & 0 & 0 & \frac{3}{2} \\ 0 & 0 & 1 & \frac{5}{2} \end{array}\right) \quad \beta = \begin{pmatrix} \frac{3}{2} \\ -\frac{1}{2} \\ \frac{5}{2} \end{pmatrix} \Rightarrow$$

$$f(x) = 1{,}5 - 0.5x + 2.5x^2$$



Вершина параболи

$$x = -\frac{-0.5}{5} = \frac{1}{10}$$

$$y = -\frac{\frac{1}{4} - 15}{10} = \frac{14.75}{10}$$

$$f(-1) = 1{,}5 + 0.5 + 2.5 = 4.5$$

$$f(0) = 1{,}5$$

$$f(1) = 1 + 2.5 = 3.5$$

Задание 2

Пользуясь №1, найти градиент $\frac{\partial g(\beta)}{\partial \beta}$, гессиан $\frac{\partial^2 g(\beta)}{\partial \beta^T \partial \beta}$ функции $g(\beta) = \|X\beta - y\|^2$.

Вывести, что решение линейной задачи наименьших квадратов $\hat{\beta} = \arg\min \|X\beta - y\|^2$ является решением нормальной системы линейных уравнений $X^T X \beta = X^T y$

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

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

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

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

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

Рассмотрим $\frac{\partial g(\beta)}{\partial \beta} = 2(X\beta - y)^T X = 0$, тогда:

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

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

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

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

что соответствует норм. системе лин. урав. (НСЛУ)

В случае если X имеет линейно независимые столбцы, то $\det X > 0$, $\det X^T > 0 \Rightarrow \det X^T X > 0 \Rightarrow$

$\Rightarrow$ НСЛУ будет иметь единственное решение

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

т.к $\det X^T X > 0$ и $X^T X$ симметричная матрица, то собственные числа матрицы > 0 и матрица положительно определённая, а значит $\frac{\partial g(\beta)}{\partial \beta}$ — ~~..~~ соответствует локальным минимумам.