# (I) Матричное дифференцирование

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

Матрицей Якоби называется матрица

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

В частности, если $m = 1$ (т.е. $g(x)$ - скалярная функция векторного аргумента $x$), то

$\frac{\partial g}{\partial x}$ - градиент функции $g$.

Доказать, что:

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

$$[a_1 \dots a_n] \times \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = (a_1 x_1 \dots a_n x_n) = \sum x_i a_i$$

$$\frac{\partial \sum x_i a_i}{\partial x_k} = a_k$$

$$\frac{\partial (a^T x)}{\partial x} = \frac{\partial (\sum x_i a_i)}{\partial x} = \begin{bmatrix} \frac{\partial \sum x_i a_i}{\partial x_1} \\ \dots \\ \frac{\partial \sum x_i a_i}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = a$$

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

$$Ax = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \times \begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \sum_j^n a_{1j} x_j \\ \dots \\ \sum_j^n a_{mj} x_j \end{bmatrix}$$

$$\frac{\partial Ax}{\partial x} = \begin{bmatrix} \frac{\sum_j^n a_{1j} x_j}{\partial x_1} & \dots & \frac{\sum_j^n a_{1j} x_j}{\partial x_n} \\ \dots & \dots & \dots \\ \frac{\sum_j^n a_{mj} x_j}{\partial x_1} & \dots & \frac{\sum_j^n a_{mj} x_j}{\partial x_n} \end{bmatrix} =$$

$$= \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$$

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

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad x^T = [x_1 \dots x_n], \quad A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$$

$$x^T A = [x_1 \dots x_n] \times \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} =$$

$$= \left[ (x_1 a_{11} + \dots + x_n a_{n1}) \quad \dots \quad (x_1 a_{1n} + \dots + x_n a_{nn}) \right]$$

$$x^T A x = \left[ (x_1 a_{11} + \ldots + x_n a_{n1}) \ldots \ (x_1 a_{1n} + \ldots + x_n a_{nn}) \right] \times \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} =$$

$$= (\underline{x_1} a_{11} + \ldots + x_n a_{n1}) \underline{x_1} + \ldots + (\underline{x_1} a_{1n} + \ldots + x_n a_{nn}) x_n$$

$$= x_1^2 a_{11} + x_1 (x_2 a_{21} + \ldots) + x_1 (x_2 a_{12} + \ldots)$$

$$\frac{\partial (x^T A x)}{\partial x} = \begin{bmatrix} \frac{\partial (x^T A x)}{\partial x_1} \\ \vdots \\ \frac{\partial (x^T A x)}{\partial x_n} \end{bmatrix}$$

$$\frac{\partial (x^T A x)}{\partial x_k} = \frac{\partial \left( x_k^2 a_{kk} + x_k \sum_{j \neq k} x_j a_{jk} + x_k \sum_{j \neq k} a_{kj} x_j + \ldots \right)}{\partial x_k} =$$

$$= 2 x_k a_{kk} + \sum_{j \neq k} x_k a_{jk} + \sum_{j \neq k} a_{kj} x_j =$$

$$= \sum_{j=1}^{n} x_j (a_{jk} + a_{kj})$$

$$\frac{\partial (x^T A x)}{\partial x} = \begin{bmatrix} \sum_{j=1}^{n} x_j (a_{j1} + a_{1j}) \\ \vdots \\ \sum_{j=1}^{n} x_j (a_{jn} + a_{nj}) \end{bmatrix} = \begin{bmatrix} a_{11} + a_{11} & a_{21} + a_{12} & \cdots \\ \vdots & \vdots & \cdots \\ a_{1n} + a_{n1} & \cdots & \cdots \end{bmatrix} \times \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} =$$

$$= (A + A^T) \times \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = (A + A^T) x$$

в частности, если $A^T = A$, то

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

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

$$\frac{\partial \|x\|^2}{\partial x} = \frac{\partial x^Tx}{\partial x}$$

$$x^Tx = [x_1 \ldots x_n]\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1^2 + \ldots + x_n^2$$

$$\frac{\partial x^Tx}{\partial x_k} = 2x_k$$

$$\frac{\partial x^Tx}{\partial x} = \begin{bmatrix} 2x_1 \\ \vdots \\ 2x_n \end{bmatrix} = 2x$$

**⑤ если $g$ – скалярная функция и под $g(x)$ понимается применение функции $g$ 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 $a$.

$$\frac{\partial g(x)}{\partial x} = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \overset{=0}{\frac{\partial g_1}{\partial x_2}} & \overset{=0}{\ldots} & \overset{=0}{\frac{\partial g_1}{\partial x_n}} \\ \vdots & & & \vdots \\ \overset{=0}{\frac{\partial g_n}{\partial x_1}} & \overset{=0}{\frac{\partial g_n}{\partial x_2}} & \overset{=0}{\ldots} & \frac{\partial g_n}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 \ldots & \frac{\partial g_n}{\partial x_n} \end{pmatrix}$$

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

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

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

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

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

$$= \begin{pmatrix} \frac{\partial g_1}{\partial h_1} \cdot \frac{\partial h_1}{\partial x_1} + \frac{\partial g_1}{\partial h_2} \cdot \frac{\partial h_2}{\partial x_1} + \dots & \dots & \frac{\partial g_1}{\partial h_1} \cdot \frac{\partial h_1}{\partial x_n} + \dots + \frac{\partial g_1}{\partial h_m} \cdot \frac{\partial h_m}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial g_p}{\partial h_1} \cdot \frac{\partial h_1}{\partial x_1} + \dots + \frac{\partial g_p}{\partial h_m} \cdot \frac{\partial h_m}{\partial x_1} & \dots & \frac{\partial g_p}{\partial h_1} \cdot \frac{\partial h_1}{\partial x_n} + \dots + \frac{\partial g_p}{\partial h_m} \cdot \frac{\partial h_m}{\partial x_n} \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial g_p}{\partial x_1} & \dots & \frac{\partial g_p}{\partial x_n} \end{pmatrix}$$

(II) Пользуясь №I, найдите градиент $\frac{\partial g(\beta)}{\partial \beta}$ и гессиан $\frac{\partial^2 g(\beta)}{\partial \beta^T \partial \beta}$ функции $g(\beta) = \|X\beta - y\|^2$. Выведите отсюда, что решение линейной задачи наименьших квадратов $\hat{\beta} = \arg\min \|X\beta - y\|^2$ является решением нормальной системы линейных уравнений $X^T X \beta = X^T y$.

Решение:

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

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

= используем результаты задач 2, 3 =

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

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

$$= 2X^T X$$

Обозначим $x_0, x_1, \dots x_d$ столбцы матрицы $X$.
Если $X_0, X_1, \dots X_d$ - линейно независимые,
то матрица $X^T X$ невырождена и положительно
определена, поэтому минимум $\|X\beta - y\|^2$
достигается, когда первая производная по $\beta$
обращается в 0:

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

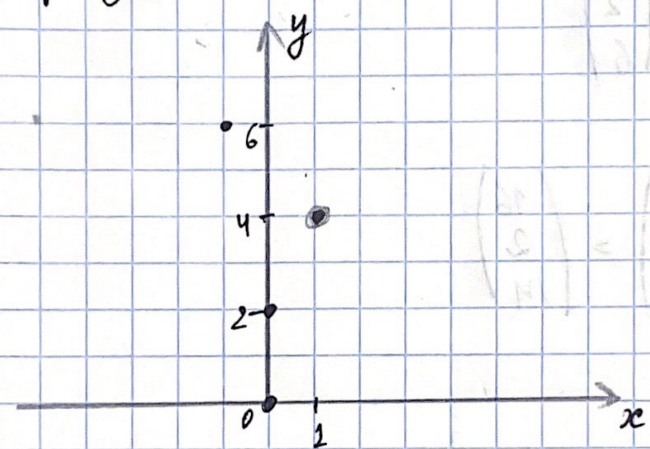
Единственным решением является вектор

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

(III) Дана обучающая выборка

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

1) изобразить точки




2) методом наименьших квадратов построить модель вида $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$; построить график этой функции;

Составим систему $X^T X \beta = X^T y$ и решим её.
Составляем матрицу $X$ и вектор $y$.

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

Имеем

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

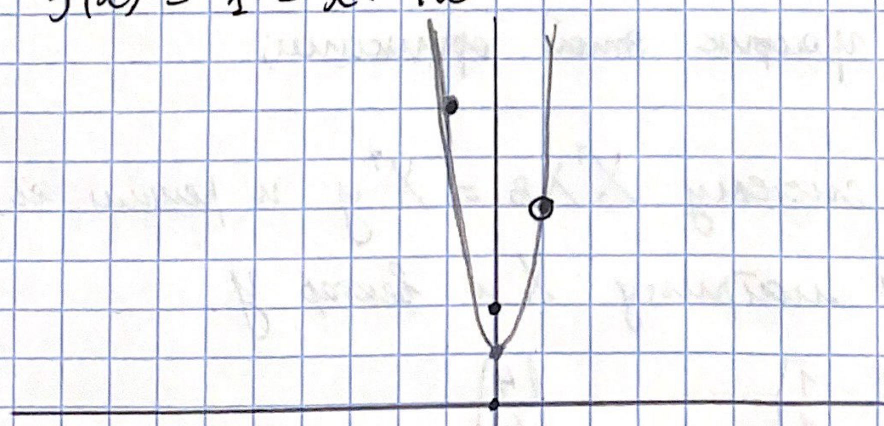
$$X^T y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 4 \\ 0 \\ 2 \\ 6 \end{pmatrix} = \begin{pmatrix} 16 \\ 2 \\ 14 \end{pmatrix}$$

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

$$\beta = \begin{pmatrix} 1 \\ -1 \\ 4 \end{pmatrix}$$

Таким образом, нашли модель

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



3) построить модель того же вида методом ридж-регрессии с параметром регуляризации $\lambda = 1$;

построить график этой функции.

Регуляризованная система $(X^T X + \lambda I)\beta = X^T y$,

где $I$ – единичная матрица

Для $\lambda = 1$ получаем

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

Решая регуляризованную систему нормальных уравнений

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

$$\begin{pmatrix} 6 & 1 & 3 \\ 1 & 4 & 1 \\ 3 & 1 & 4 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 16 \\ 2 \\ 14 \end{pmatrix},$$

находим

$$\beta = \begin{pmatrix} 3/2 \\ -1/2 \\ 5/2 \end{pmatrix}$$

Таким образом, наша модель

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