№1 $\exists\, g: R^n \to R^m,\ x \in R^n$

$$\frac{\partial g}{\partial x} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \dots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \dots & \frac{\partial g_2}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \dots & \frac{\partial g_m}{\partial x_n} \end{bmatrix}$$ — матрица Якоби

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

$\triangle$ $a^T x = a_1 x_1 + a_2 x_2 + \dots + a_n x_n$

$$\frac{\partial(a_1 x_1 + \dots + a_n x_n)}{\partial x} = (a_1, a_2, \dots, a_n) = a$$

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

$$\triangle\ Ax = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \dots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \dots + a_{1n}x_n \\ \dots \\ a_{m1}x_1 + \dots + a_{mn}x_n \end{bmatrix}$$

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

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

J $A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$

$$\Delta \ x^T A x = \begin{aligned} & x_1(a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n) + \\ & + x_2(a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n) + \\ & + \dots \dots \dots \dots \dots \dots + \\ & + x_n(a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n) \end{aligned}$$

$$\Delta \ \frac{\partial(x^T A x)}{\partial x_1} = \begin{aligned} & (a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n) + \\ & + (a_{11}x_1 + a_{21}x_1 + \dots + a_{n1}x_n) \end{aligned}$$

т.е. $\frac{\partial(x^T A x)}{\partial x_1} = \sum_{i=1}^{n} (A_{1i} + A^T_{1i}) x_i$, таким образом:

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

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

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

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

5) если f – скалярная ф-я и под $f(x)$ понимается применение ф-ии f к каждой компоненте вектора $x \in R^n$, то

$$\frac{\partial g(x)}{\partial x} = diag(g'(x)),$$ где $diag(a)$ - диагональная матрица с диагональю

$$\text{Д} \quad g(x) = \begin{bmatrix} g(x_1) \\ g(x_2) \\ \cdots \\ g(x_n) \end{bmatrix}$$

$$\frac{\partial g(x)}{\partial x} = \begin{bmatrix} \frac{\partial g(x_1)}{\partial x_1} & \frac{\partial g(x_1)}{\partial x_2} & \cdots & \frac{\partial g(x_1)}{\partial x_n} \\ \frac{\partial g(x_2)}{\partial x_1} & \frac{\partial g(x_2)}{\partial x_2} & \cdots & \frac{\partial g(x_2)}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial g(x_n)}{\partial x_1} & \frac{\partial g(x_n)}{\partial x_2} & \cdots & \frac{\partial g(x_n)}{\partial x_n} \end{bmatrix}$$

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

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

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

$$h(x) = \begin{bmatrix} h_1(x) \\ h_2(x) \\ \cdots \\ h_m(x) \end{bmatrix}$$

$$\text{Д} \quad \frac{\partial g_1(h(x))}{\partial x_1} = \frac{\partial g_1}{\partial h_1} \frac{\partial h_1}{\partial x_1} + \frac{\partial g_1}{\partial h_2} \frac{\partial h_2}{\partial x_2} + \ldots$$

$$4 \quad \frac{\partial g_1(h(x))}{\partial x_2} = \frac{\partial g_1}{\partial h_1}\frac{\partial h_1}{\partial x_2} + \frac{\partial g_1}{\partial h_2}\frac{\partial h_2}{\partial x_2} + \dots$$

$$\frac{\partial g(h(x))}{\partial x} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_n} \\ \dots & \dots & \dots \\ \frac{\partial g_p}{\partial x_1} & \dots & \frac{\partial g_p}{\partial x_n} \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{\partial g_1}{\partial h_1} & \dots & \frac{\partial g_1}{\partial h_m} \\ \frac{\partial g_2}{\partial h_1} & \dots & \frac{\partial g_2}{h_m} \\ \dots & \dots & \dots \\ \frac{\partial g_p}{\partial h_1} & \dots & \frac{\partial g_p}{\partial h_m} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \dots & \frac{\partial h_1}{\partial x_n} \\ \frac{\partial h_2}{\partial x_1} & \dots & \frac{\partial h_2}{\partial x_n} \\ \dots & \dots & \dots \\ \frac{\partial h_m}{\partial x_1} & \dots & \frac{\partial h_m}{\partial x_n} \end{bmatrix} =$$

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

№2

$$g(\beta) = \| X\beta - y \|^2$$

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

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

Вывести, что решение $\hat{\beta} = \arg\min_{\beta} g(\beta)$ явл-ся решением нормальной системы лин-х ур-й:

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

т.к $\frac{\partial^2 g(\beta)}{\partial \beta^T \partial \beta}$ положительно опр-на, то минимум $g(\beta)$ будет достигаться в точке, в которой:

$$\frac{\partial g(\beta)}{\partial \beta} = 0 \to X\beta - y = 0$$

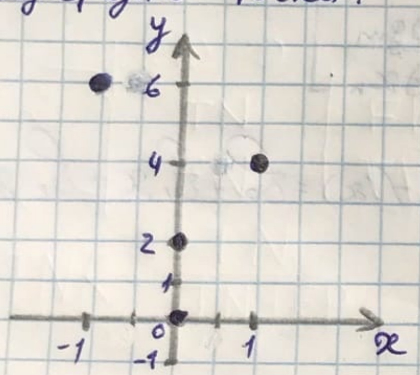
ч.т.д.

3.

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 ф-ц

Составляем матрицу X и вектор y

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

Имеем:

$$X^T \cdot X = \begin{bmatrix} 5 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 3 \end{bmatrix}, \quad X^T \cdot y = \begin{bmatrix} 16 \\ 2 \\ 14 \end{bmatrix}$$

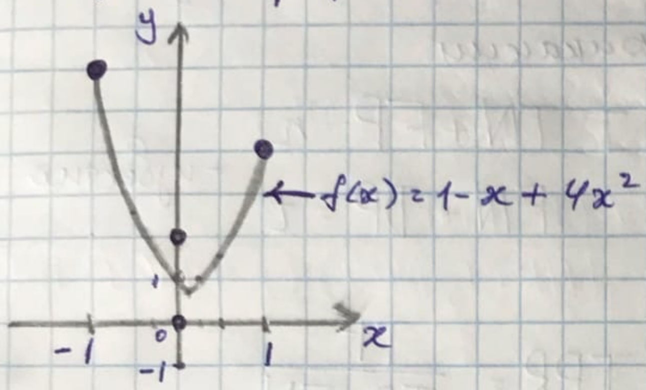
Решая систему нормальных ур-й $X^T \cdot X \cdot \beta = X^T y$, где:

$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$, находим, что: $\beta = \begin{bmatrix} 1 \\ -1 \\ 4 \end{bmatrix}$

Таким образом, модель имеет вид:

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

Построим график:



3) Построить модель ридж-регрессии с параметром регуляризации $\lambda = 1$; построить график этой ф-ц.

Учитывая вычисления в предыдущем пункте, имеем:

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

Решаем регулярную систему $(X^T X + \lambda I)\beta = X^T y$,

получаем: $\beta = \begin{bmatrix} 3/2 \\ -1/2 \\ 5/2 \end{bmatrix}$

Таким образом, модель ридж-регрессии имеет вид:

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