Оптимизация

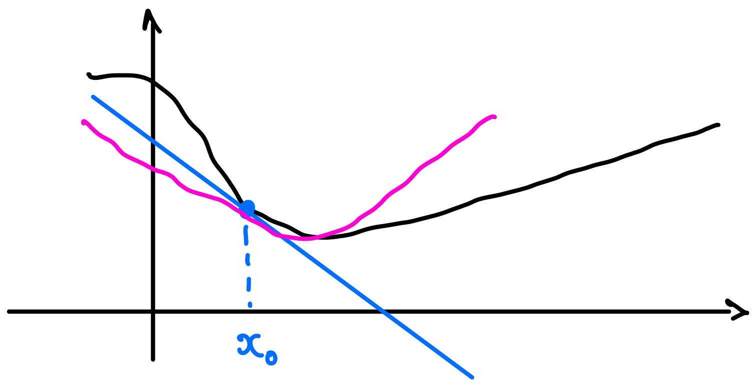
① Дифферинцеал и производная

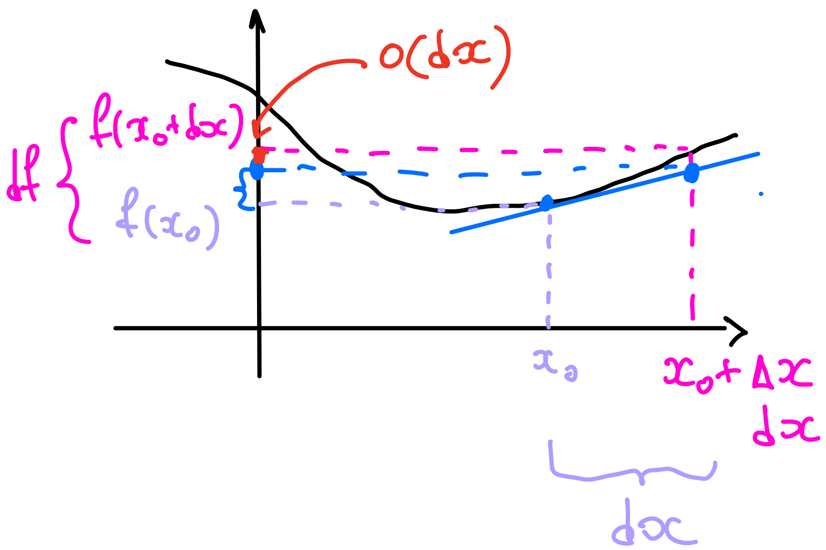
$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

$$f(x) - f(x_0) = f'(x_0)(x - x_0)$$
$$df = f'(x_0) \cdot dx$$

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \cancel{o(x)}$$

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \frac{f'''(x_0)}{3!}(x-x_0)^3 + \ldots$$





$$df = f'(x)\,dx + o(dx)$$

Опр. 1

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = 0 \qquad f(x) = \bar{o}(g(x)) \quad x \to x_0$$

g растёт быстрее f

$$e^x = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2}(x-x_0)^2 + \bar{o}(x^2)$$

$$= f(x_0) + f'(x_0)(x-x_0) + \bar{o}(x)$$

$x_0 = 0$ $\quad 1 + 1\cdot x + \frac{x^2}{2} + \underbrace{\bar{o}(x^2)}_{\frac{x^3}{3!} + \frac{x^4}{4!} + \dots}$

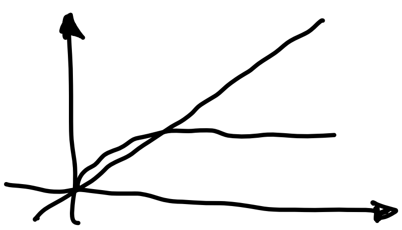
$1 + x + \bar{o}(x)$

$x^2 \in \bar{o}(x)$

$x^3 \in \bar{o}(x)$

$\ln x \notin \bar{o}(x)$

$$\lim_{x\to 0} \frac{x^2}{x} = 0 \qquad x^2 = f(x), \quad x = g(x)$$



$\ln x \ll x \ll x^k \ll a^x \ll x^x$

$k>1 \qquad a>1$

$x! \sim \sqrt{2\pi x}\cdot\left(\frac{x}{e}\right)^x$

ф-ла Стирлинга

$$\lim_{x\to\infty} \frac{x!}{x^x} = 0$$

$$\lim_{x\to\infty} \frac{e^x}{x!} = 0$$

$x \cdot \ln x \ll x^2$

Упражнение

$$o(42x) = o(x)$$

$$o(x) + o(x) = o(x)$$

$$o(x) + o(x^2) = o(x)$$

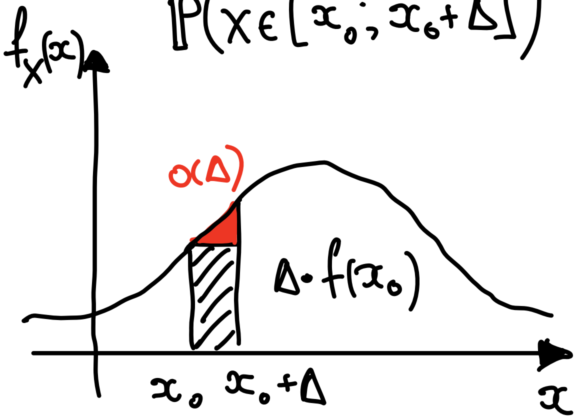
$$o(-x) = o(x)$$

X	0	1
$P(X=k)$	0.5	0.5

$f_X(x)$

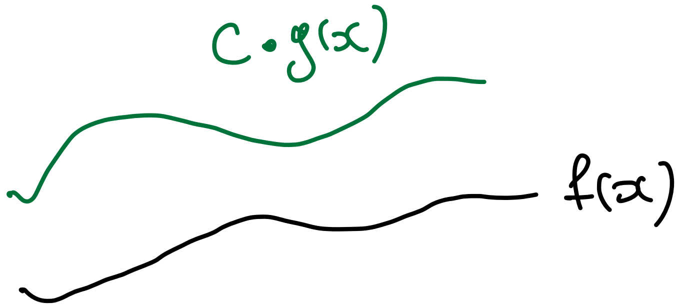
$X \sim N(\mu, \sigma^2)$

$$\Delta \cdot f(x_0) + o(\Delta) = P(X \in [x_0; x_0+\Delta])$$



Опр.

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = const > 0$$

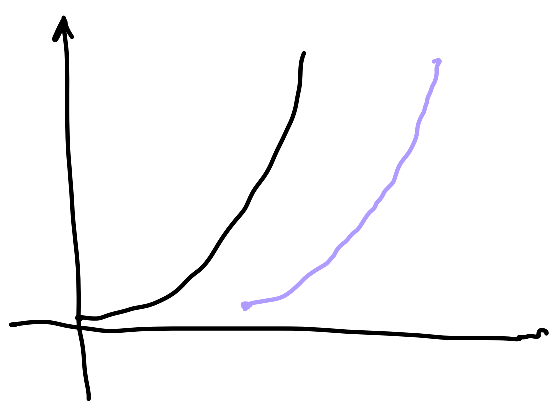


$$f(x) \leq c \cdot |g(x)|$$

$$f(x) = O(g(x))$$

$$n^2 = O(n^2)$$

$$\frac{n^2}{2} - 4n - 8 = O(n^2)$$



$\Omega(n^2)$ лучший случай

$\Theta(n^2)$ сразу Ω и O

$O(n^2)$ худший случай

$\Omega(n \log n)$

② Функции неск. переменных

$$f(x_1, \dots, x_n) : \mathbb{R}^n \to \mathbb{R}$$

$$\nabla_x f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} \qquad \frac{\partial f}{\partial x_i} = \lim_{\Delta \to 0} \frac{f(x_1, \dots, x_i + \Delta, \dots, x_n)}{\Delta}$$

градиент

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1} & \frac{\partial^2 f}{\partial x_n x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

матрица Гессе

Штольц

$$f(x) = f(a) + \left.\frac{\partial f}{\partial x_1}\right|_{x_1 = a_1} (x_1 - a_1) + \left.\frac{\partial f}{\partial x_2}\right|_{x_2 = a_2} (x_2 - a_2) + \dots + \left.\frac{\partial f}{\partial x_n}\right|_{x_n = a_n} (x_n - a_n) +$$

$$\boxed{df = \nabla f^T dx}$$

$$\nabla_x f^T (x_0) \cdot (x - x_0)$$

$$\langle \nabla_x f(a); (x - a) \rangle$$

$$+ \frac{1}{2} (x-a)^T H (x-a) + o(\|x\|^2)$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i x_j}(a) \cdot (x_i - a_i)(x_j - a_j)$$

$$\|x\|^2 = \langle x, x \rangle = x_1^2 + \dots + x_n^2$$

$$f(x) = f(a) + \nabla_x^T f(a)(x-a) + \frac{1}{2}(x-a)^T H (x-a) + o(\|x\|^2)$$

$$f: \mathbb{R}^n \to \mathbb{R}^m$$

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} f_1(x_1, \dots x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{pmatrix} \qquad \underset{\text{матр. Якоби}}{J} = \left(\frac{\partial f_i}{\partial x_j} \right)_{ij}$$

$$\boxed{df = J \cdot dx}$$

матрицы: $df = tr(\nabla_X^T f \, dX)$

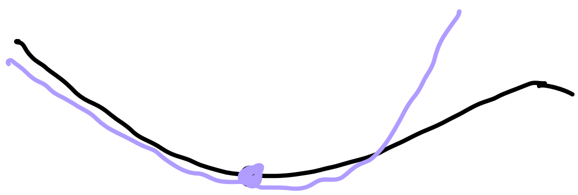
1 лекция advanced

Упражнение

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

$$f'(x) = -x^2 + 6x - 5 = 0$$

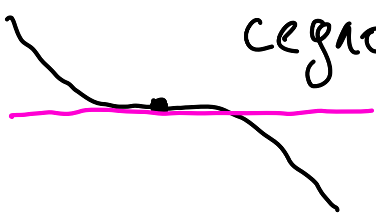
$$\left.\begin{matrix} x_1 = 1 \\ x_2 = 5 \end{matrix}\right\} \text{критические точки}$$



$$f''(x) = -2x + 6$$

$f''(1) = 4 > 0$ min

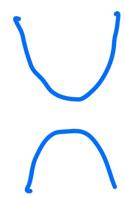
$f''(5) = -4 < 0$ max



$ax^2 + bx + c$

$a > 0$

$a < 0$



Задолжал случай $f''(x_0)=0$ (Теорема про чередование знаков)

Упражнение

$$f(x,y) = 3x^2 + xy + 2y^2 - x - 4y$$

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} 6x + y - 1 \\ x + 4y - 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{cases} 6x + y - 1 = 0 \\ x + 4y - 4 = 0 \end{cases}$$

$$\left(\begin{array}{cc|c} 6 & 1 & 1 \\ 1 & 4 & 4 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 1/6 & 1/6 \\ 1 & 4 & 4 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 1/6 & 1/6 \\ 0 & 23/24 & 23/24 \end{array}\right) \sim$$

$$\sim \left(\begin{array}{cc|c} 1 & 1/6 & 1/6 \\ 0 & 1 & 1 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 1 \end{array}\right) \quad \begin{cases} x = 0 \\ y = 1 \end{cases}$$

Крит. точка

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \partial x} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 6 & 1 \\ 1 & 4 \end{pmatrix}$$

$$H\big|_{(0,1)} = \begin{pmatrix} 6 & 1 \\ 1 & 4 \end{pmatrix}$$

Полож. определена или отрицательно опр. ?

$\Delta_1 > 0$ $\quad >$

$\Delta_2 > 0$ $\quad <$

$\Delta_3 > 0$ $\quad >$

$\dots\dots$ $\quad \dots$

$\Delta_n > 0$ $\quad >$

остальное

$6 > 0$

$\det\begin{pmatrix} 6 & 1 \\ 1 & 4 \end{pmatrix} = 24 - 1 > 0$

$\Rightarrow \min$

min max седло

$$\boxed{v^T \begin{pmatrix} 6 & 1 \\ 1 & 4 \end{pmatrix} v > 0}$$

$$d^2 f = dx^T \cdot H \cdot dx = d(df)$$

Упражнение

$$f(x, y) = e^{x-y} \qquad x_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\nabla f = \begin{pmatrix} e^{x-y} \\ -e^{x-y} \end{pmatrix} \qquad H = \begin{pmatrix} e^{x-y} & -e^{x-y} \\ -e^{x-y} & e^{x-y} \end{pmatrix}$$

$$\nabla f(0) = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \qquad H(0) = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

$$(x - y, \ -x + y)$$

$$f(x,y) \approx f(0,0) + (1, -1)\begin{pmatrix} x - 0 \\ y - 0 \end{pmatrix} + \frac{1}{2}(x - 0, \ y - 0)\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x - 0 \\ y - 0 \end{pmatrix}$$

$$1 + x - y + \frac{x^2}{2} - \frac{yx}{2} - \frac{xy}{2} + \frac{y^2}{2} = 1 + x - y + \frac{x^2}{2} + \frac{y^2}{2} - xy$$

③ Градиентные спуски

$$w \qquad Q(w) \to \min_w$$

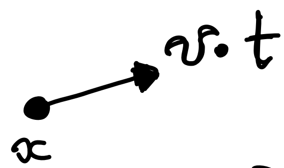
$$w_0 = 0 \qquad \text{random}$$

$$w_t = w_{t-1} - \eta_t \cdot \nabla_w Q(w_{t-1})$$

η_t: 0.01 learning rate

Производная по направлению

$$\lim_{t \to 0} \frac{f(x + v \cdot t) - f(x)}{t} = \nabla_v f(x)$$



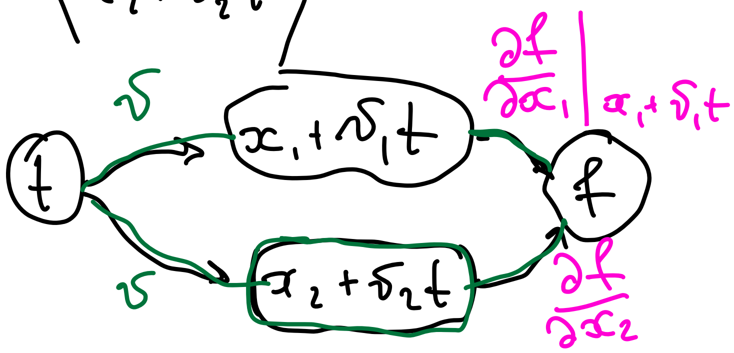
$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

$$v = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$ – частная производная

$$\frac{\partial}{\partial t} f(x + v \cdot t) \Big|_{t=0} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x) \cdot v_i = \langle \nabla f(x), v \rangle =$$

$$= \|\nabla f(x)\| \cdot \|v\| \cdot \cos\alpha$$

$$f\begin{pmatrix} x_1 + v_1 \cdot t \\ x_2 + v_2 t \end{pmatrix}$$



$$v \parallel \nabla f(x)$$

$$\cos\alpha = \cos 0 = 1$$

Направление max ↑ функции из x

Упражнение

$L = (y - wx)^2$

$y = w \cdot x$

$Q(w) = \frac{1}{n}\sum_{i=1}^{n}(y - wx)^2$

x	y
1	2
2	3

$\sum_{i=1}^{n} \frac{\partial L}{\partial w}(x_i, y_i)$

а) $\hat{w}^{OLS}$ – ? $\quad Q(w) = \frac{1}{2}\left((2-w)^2 + (3-2w)^2\right) \to \min_w$

$$\frac{\partial Q}{\partial w} = -(2-w) - 2(3-2w) = 0$$

$$\Rightarrow \hat{w} = 1.6$$

б) GD $\quad \eta = 0.1$

2 шага GD

$w_0 = 0$

-8

$w_1 = 0 - 0.1 \cdot \nabla_w Q(0) = 0.8$

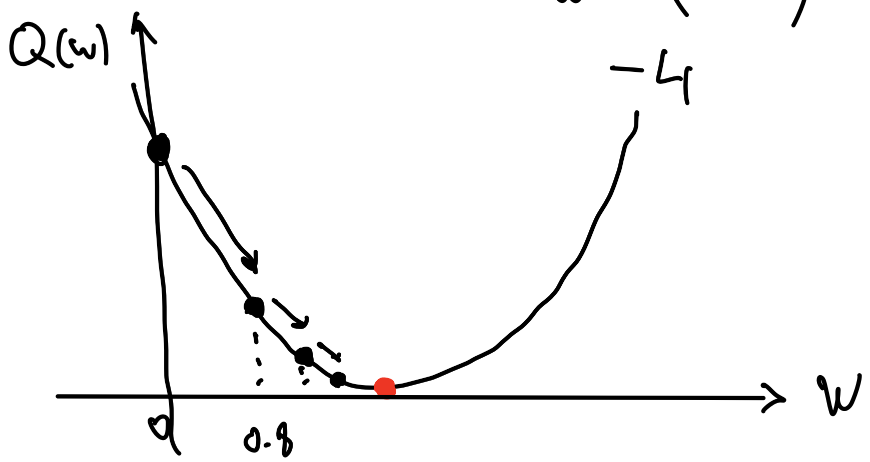
$-(2-w) - 2(3-2w)$

0 $\quad$ 0

0.8 $\quad$ 0.8

$w_2 = 0.8 - 0.1 \; \nabla_w Q(0.8) = 1.2$

-4

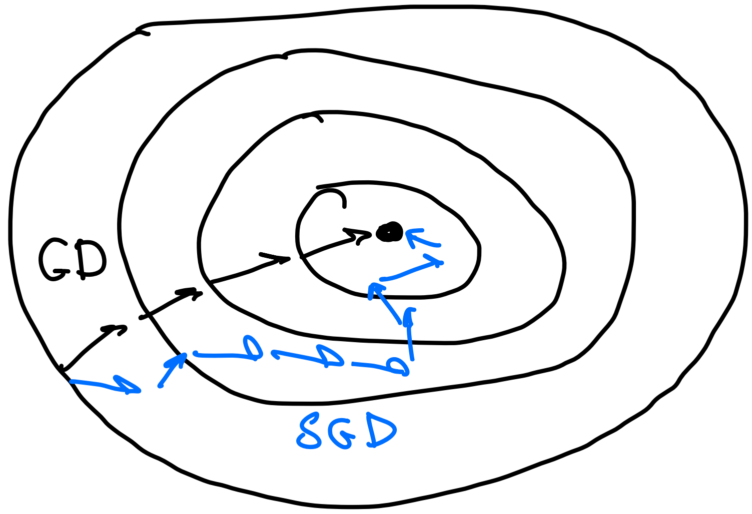


1 итерация:

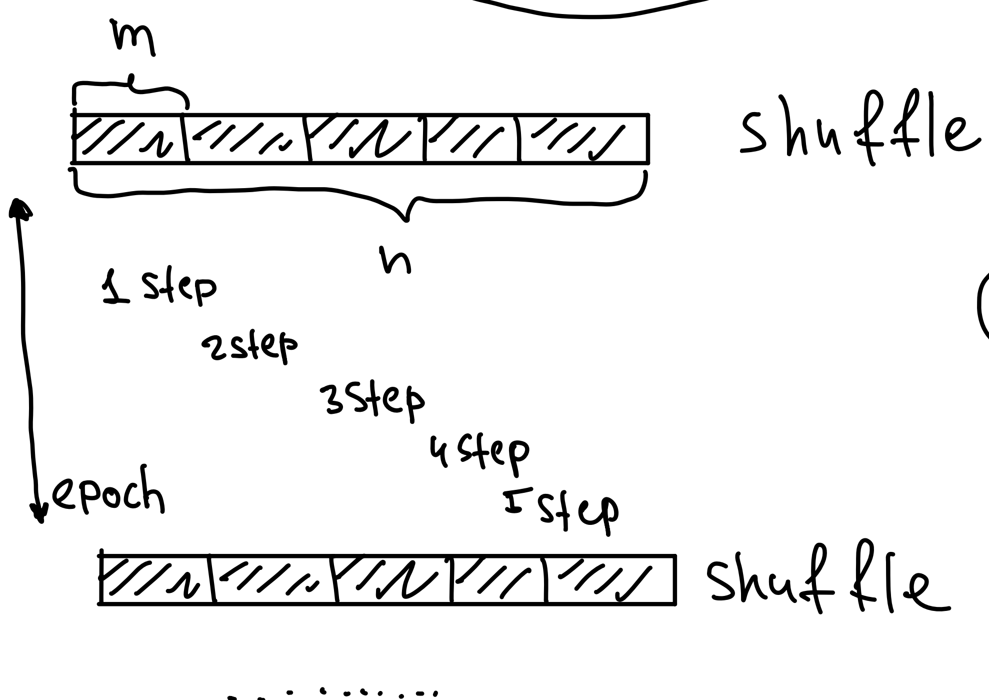
$O(n)$

в) SGD $O(1)$

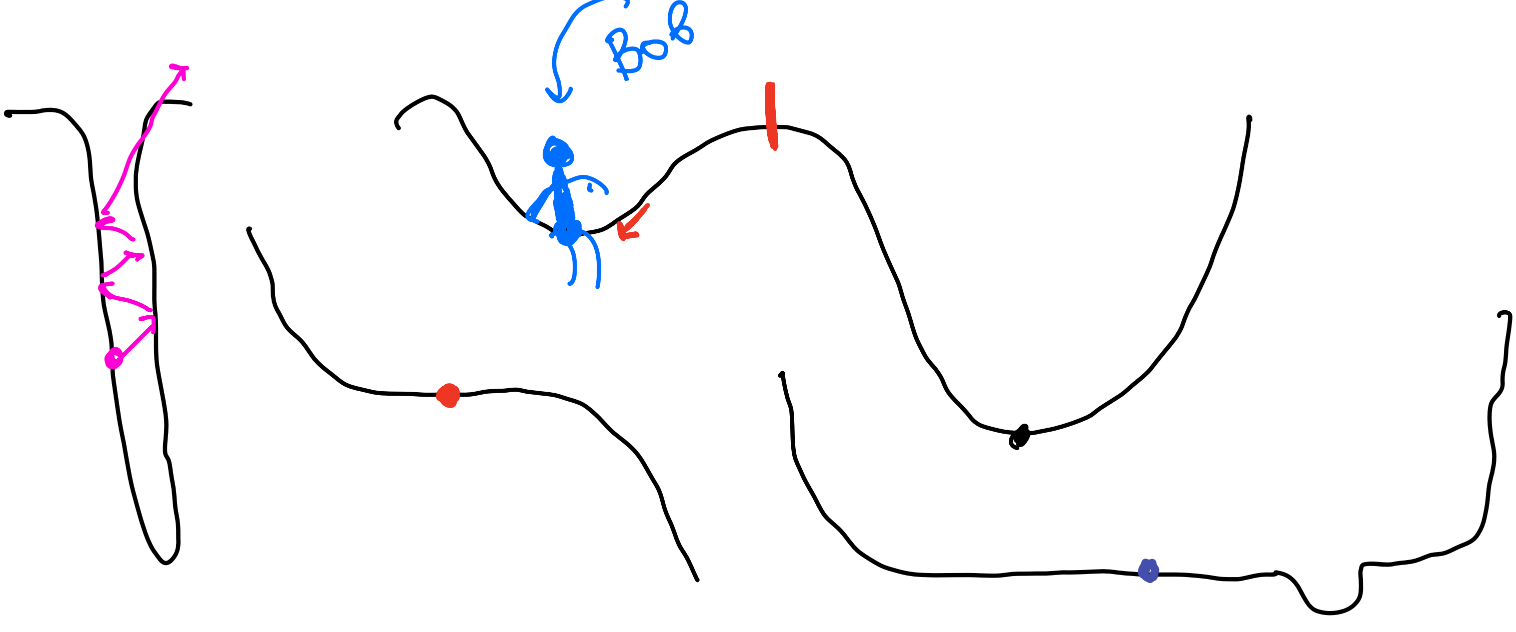
1 рандомное наблюдение

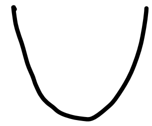


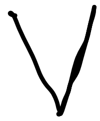
$O(m) = O(10) =$
$= O(1)$



mini-batch SGD



MSE $L(y, \hat{y}) = (y - \hat{y})^2$ 

MAE $L(y, \hat{y}) = |y - \hat{y}|$ 

4 Условная оптимизация

$$\begin{cases} f(x) \to \min\limits_{x} \\ g(x) = a \end{cases} \qquad x = g^{-1}(a) \qquad f(g^{-1}(a)) \to \min\limits_{a}$$

$$\begin{cases} f(x) \to \min\limits_{x} \\ g_1(x) = 0 \\ \dots\dots \\ g_m(x) = 0 \end{cases}$$

$$\mathcal{L} = f(x) - \lambda_1 g_1(x) - \ldots - \lambda_m g_m(x)$$

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x} = 0 \\ g_i(x) = 0 \end{cases} \Rightarrow$$ Набор крит. точек

Упражнение

$$\begin{cases} 5 - 3x - 4y \to \min\limits_{x,y} \\ x^2 + y^2 = 25 \end{cases} \qquad \begin{cases} f(x, y) \to \min\limits_{x,y} \\ g(x, y) = 0 \end{cases}$$

$$\mathcal{L} = 5 - 3x - 4y - \lambda(x^2 + y^2 - 25)$$

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x} = -3 - 2x\lambda = 0 \\ \dfrac{\partial \mathcal{L}}{\partial y} = -4 - 2y\lambda = 0 \end{cases}$$

$$2x\lambda = -3 \qquad \lambda = \frac{-3}{2x}$$

$$2y\lambda = -4 \qquad \lambda = -\frac{4}{2y}$$

$$8x = 6y \qquad 4x = 3y$$

$$x^2+y^2=25$$
$$\frac{9}{16}y^2+y^2=25$$
$$\frac{25}{16}y^2=25$$
$$y^2=16$$
$$y=\pm 4$$

$$x=\frac{3}{4}y=\pm 3$$

Критические точки:

$(3,4)$ $\lambda=-1/2$

$(-3,-4)$ $\lambda=1/2$

$$H=\begin{pmatrix} \frac{\partial^2 L}{\partial x^2} & \frac{\partial^2 L}{\partial x \partial y} \\ \frac{\partial^2 L}{\partial y \partial x} & \frac{\partial^2 L}{\partial y^2} \end{pmatrix}=\begin{pmatrix} -2\lambda & 0 \\ 0 & -2\lambda \end{pmatrix}$$

$\Delta_0>0$ 1 $(3,4)$

$\Delta_1>0$ 2 min

$\Delta_0<0$ −1 $(-3,-4)$

$\Delta_1>0$ 2 max

$$H=\begin{pmatrix} \Delta_0 & \\ & \Delta_1 \end{pmatrix}$$

Другой путь:

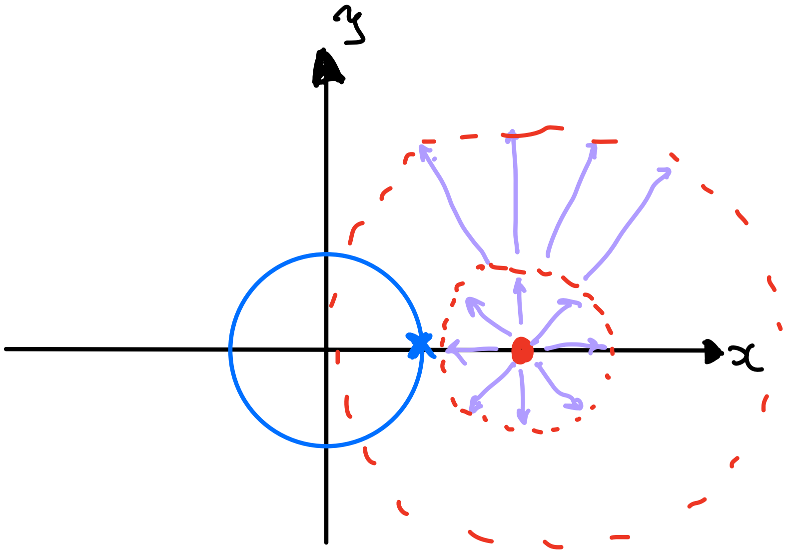
$$A=\begin{pmatrix} 0 & g'_x & g'_y \\ g'_x & \frac{\partial^2 L}{\partial x^2} & \frac{\partial^2 L}{\partial x \partial y} \\ g'_y & \frac{\partial^2 L}{\partial y \partial x} & \frac{\partial^2 L}{\partial y^2} \end{pmatrix}$$

Знаки !!!

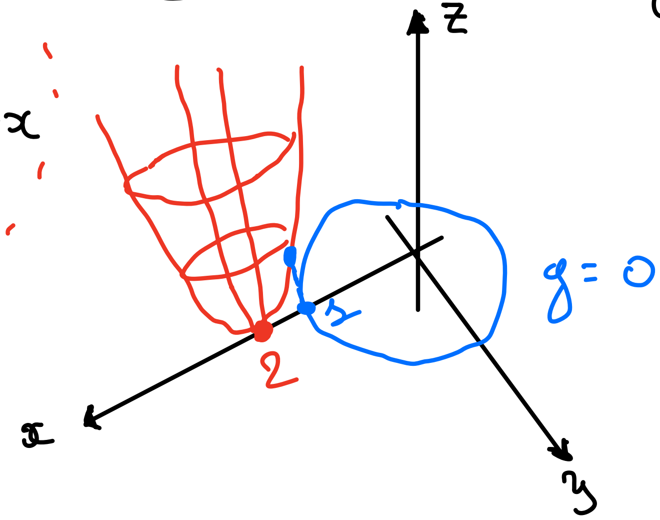
$\det A<0$ min

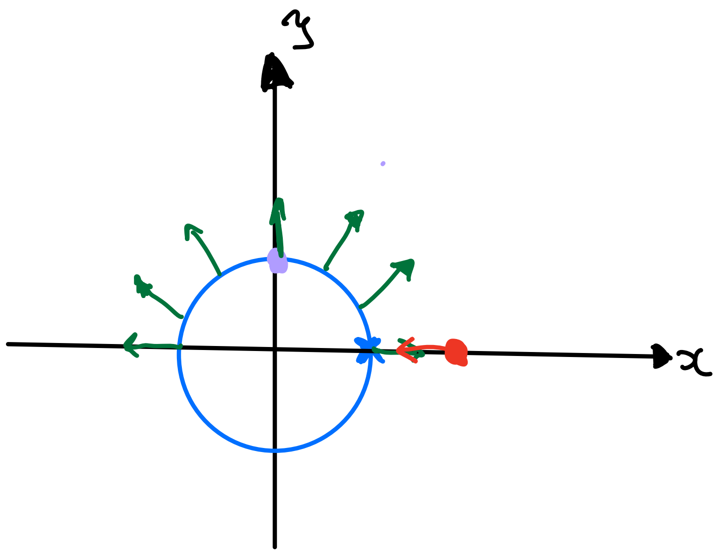
$\det A>0$ max

Почему это работает?



$$\begin{cases}(x-2)^2+y^2 \to \min & f(x,y)\\ x^2+y^2-1=0 & g(x,y)\end{cases}$$





$g(x,y)=0$

$\left[\ \nabla f \perp\right.$ поверхности $g(x,y)=0$

$$\nabla f \parallel \nabla g$$

$$\nabla f = \lambda \nabla g$$

$$\nabla\Big(\underbrace{f - \lambda \cdot \nabla g}_{\mathcal{L}}\Big)$$

Выписывая условие выше,

я предполагал $\nabla g \neq 0$

Надо рассматривать две ситуации

$$\mathcal{L} = f - \lambda g$$

$$\begin{cases} \nabla \mathcal{L} = 0 \\ g = 0 \end{cases} \qquad \qquad \begin{cases} \nabla g = 0 \\ g = 0 \end{cases}$$

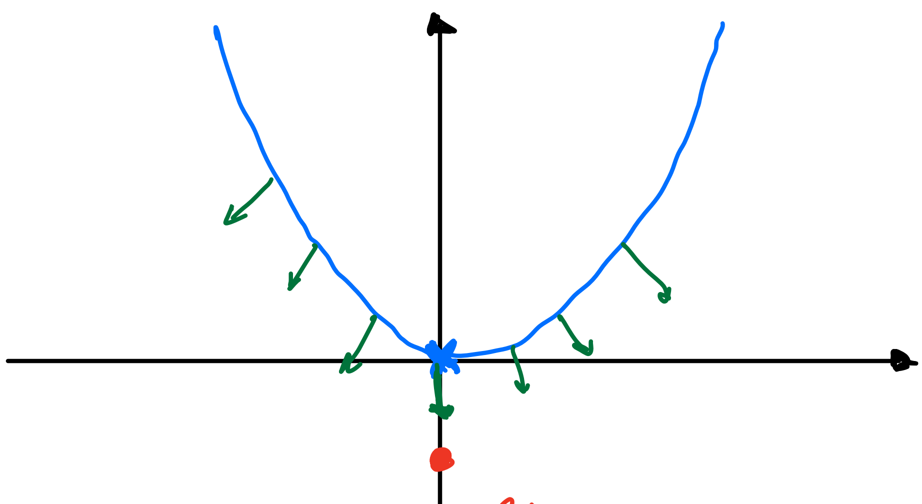
„хорошие" точки „плохие" точки

Упражнение

$$\begin{cases} f(x,y) = x^2 + (y+1)^2 \\ g_1(x,y) = x^2 - y = 0 \end{cases} \qquad \begin{cases} f(x,y) = x^2 + (y+1)^2 \\ g_2(x,y) = x^4 - y^2 \end{cases}$$

$$\nabla g_1(x,y) = (2x, -1) \neq 0 \qquad \nabla g_2(x,y) = (4x^3, -2y)$$

$\forall$ задача хорошо решается проблемы



чаша с минимумом в точке -1

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x} = 0 \\ \dfrac{\partial \mathcal{L}}{\partial y} = 0 \\ g_2 = 0 \end{cases} \qquad \varnothing$$

Однако:

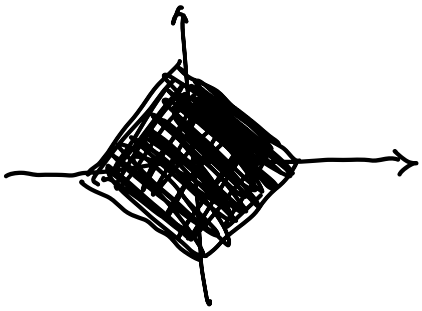
$$\begin{cases} \dfrac{\partial g_2}{\partial x} = 0 \\ \dfrac{\partial g_1}{\partial y} = 0 \\ g_2 = 0 \end{cases} \Rightarrow (0,0)$$

$l_1 \quad l_2$

$$\sum |w_i| \leq C$$

$$\sum w_i^2 \leq C$$

$$Q(w) \to \min_w$$



Условия с неравенствами

$$\begin{cases} f(x_1, \dots, x_n) \to \min\limits_{x_1, \dots, x_n} \\ \left.\begin{matrix} g_1 = 0 \\ \dots \\ g_m = 0 \end{matrix}\right| \lambda \\ \left.\begin{matrix} h_1 \leq 0 \\ \dots \\ h_k \leq 0 \end{matrix}\right| \mu \end{cases}$$

ККТ Теорема

$$\begin{cases} \nabla \mathcal{L} = 0 \\ h_i \leq 0 \qquad \mu_i \geq 0 \\ g_j = 0 \\ \mu_i \cdot h_i = 0 \end{cases}$$

Условие дополнит. нежесткости

Упражнение

$$\begin{cases} (x-4)^2 + (y-4)^2 \to \min\limits_{x,y} \\ x + y \le 4 \qquad x+y-4 \le 0 \\ x + 3y \le 9 \qquad x+3y-9 \le 0 \end{cases}$$

$$L = (x-4)^2 + (y-4)^2 + \mu_1 (x+y-4) + \mu_2 (x+3y-9)$$

$$2(x-4) + \mu_1 + \mu_2 = 0 \qquad L'_x$$

$$2(y-4) + \mu_1 + 3\mu_2 = 0 \qquad L'_y$$

$$x+y-4 \le 0 \qquad \mu_1 \ge 0 \qquad \mu_1 \cdot (x+y-4) = 0$$

$$x+3y-9 \le 0 \qquad \mu_2 \ge 0 \qquad \mu_2 \cdot (x+3y-9) = 0$$

① $$\begin{cases} x+y-4=0 \\ x+3y-9=0 \\ \mu_1 \ge 0 \\ \mu_2 \ge 0 \\ L'_x = 0 \\ L'_y = 0 \end{cases} \qquad x = 3/2 \quad y = 5/2 \quad \mu_2 = -1 \;\; ??!$$

$\varnothing$

② $$\begin{cases} x+y-4=0 \\ \mu_2 = 0 \\ x+3y-9 \le 0 \\ \mu_1 \ge 0 \\ L'_x = 0 \quad L'_y = 0 \end{cases} \qquad \begin{aligned} x &= 2 \\ y &= 2 \\ \mu_1 &= 4 \\ \mu_2 &= 0 \end{aligned}$$

✓

③ $\varnothing$ ④ $\varnothing$