Оптимизация

$f(x) \to \min\limits_x$

$f'(x) = 0 \Rightarrow x^*$

$f''(x) \quad f''(x^*) > 0$

$f(x_1, \dots, x_n) \to \min\limits_{\vec{x}}$

$\nabla_x f = 0 \Rightarrow x^*$

$\nabla^2_x f = H$

$H > 0$ min

$H < 0$ max

H ??? седловая точка

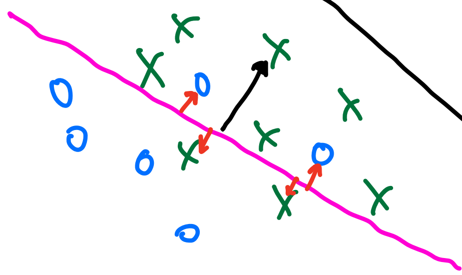
Критерий Сильвестра

$\begin{array}{ll} \Delta_1 > 0 & \Delta_1 > 0 \\ \Delta_2 > 0 & \Delta_2 < 0 \\ \Delta_3 > 0 & \Delta_3 > 0 \\ \dots & \dots \\ \min & \max \end{array}$

$\begin{cases} f(x) \to \min\limits_x \\ h(x) = a \end{cases}$

$x = h^{-1}(a)$

$f(h^{-1}(a)) \to \min\limits_a$

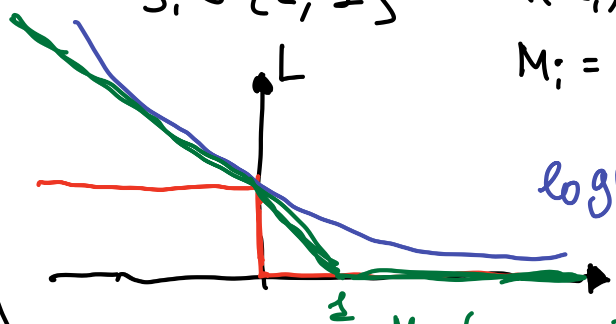


$[M_i > 0]$

$L(y_i, a(x_i)) = [y_i = a(x_i)]$

$y_i \in \{-1, 1\}$ $\quad a(x_i)$ – модель

$M_i = y_i \cdot a(x_i)$



$\log(1 + \exp(-M_i))$ логистич. регрессия

$\max(1 - M_i, 0)$ SVM

$y_i \ln p_i + (1 - y_i) \ln(1 - p_i) \quad \{0, 1\}$

Лагранжиан

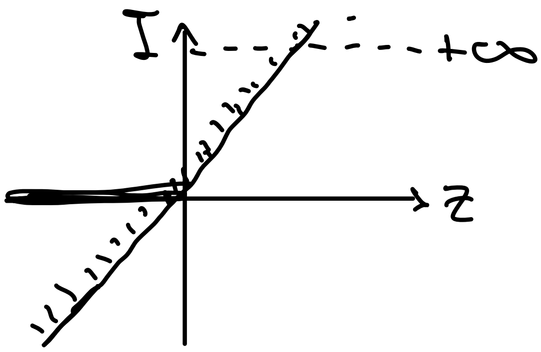
$\begin{cases} f(x) \to \min\limits_{x \in \mathbb{R}^d} \\ g_i(x) \leqslant 0 & i = 1, \dots, m \\ h_j(x) = 0 & j = 1, \dots, p \end{cases}$

$I_-(z) = \begin{cases} 0 & z \leq 0 \\ \infty & z > 0 \end{cases} \qquad \lambda \cdot z$

$I_0(z) = \begin{cases} 0, & z = 0 \\ \infty, & z \neq 0 \end{cases} \qquad \mu \cdot z$

$f(x) + \sum\limits_{i=1}^{m} I_-(g_i(x)) + \sum\limits_{j=1}^{p} I_0(h_j(x)) \to \min\limits_x$

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x) \to \min_x$$



! $\lambda_i \geq 0$

Двойственная ф. даёт оценку на минимум исх. функции

Двойственные функции

x – прямые переменные (primary)

λ, μ – двойственные переменных (dual)

$g(\lambda, \mu) = \inf\limits_x \mathcal{L}(x, \mu, \lambda)$ – двойственная ф.

→ всегда вогнутая

→ может при некоторых λ, μ равняться $-\infty$

Мотивация:

курсы по оптим.
- → теория двойственности
- → алгоритмы
- → когда работает

SGD:

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

- Если ф. выпуклая
- Липшевость градиента
- Условие Роббинса-Монро

$$\sum_{k=1}^{\infty} \eta_k = \infty \qquad \sum_{k=1}^{\infty} \eta_k^2 < \infty$$

$\Rightarrow$ Работает

DL: у нас все предпосылки нарушены, но мы всё равно используем SGD

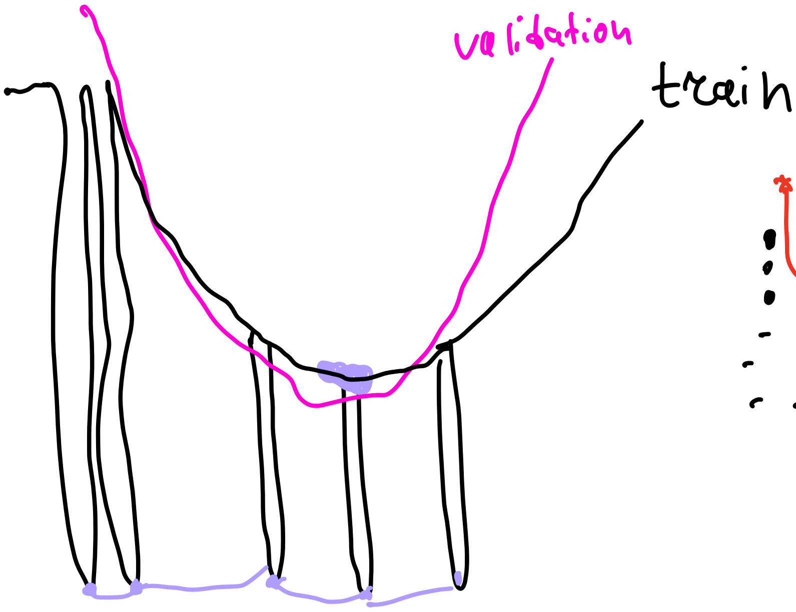
$\Rightarrow$ СПЕЦЭФФЕКТЫ

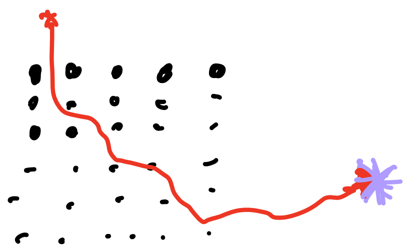
Фишер $\frac{n}{d} \to \infty$ ММП

$d >> n$:(

игрушки дьявола:
- регуляризация
- дропауты
- аугментация

....





---

x' – допустимая точка $g_i(x') \le 0$ $h_j(x') = 0$

$$\mathcal{L}(x',\lambda,\mu) = f(x') + \sum_{i=1}^{m} \underbrace{\lambda_i}_{\ge 0} \underbrace{g_i(x')}_{\le 0} + \sum_{j=1}^{p} \mu_j \overbrace{h_j(x')}^{=0} \le f(x')$$

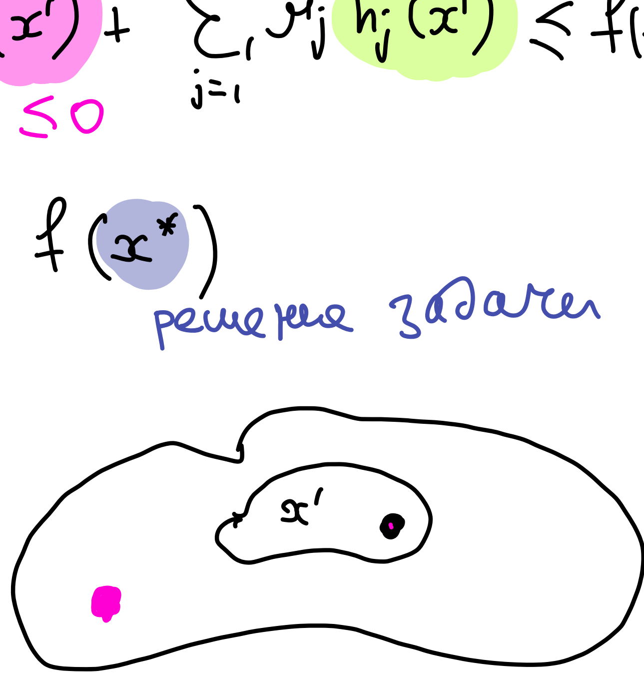
$$\inf_{x'} \mathcal{L}(x',\lambda,\mu) \le \inf_{x'} f(x') = f(x^*)$$

решение задачи

$$\geq$$

$$\inf_{x} \mathcal{L}(x,\lambda,\mu) = g(\lambda,\mu)$$

$$\Rightarrow g(\lambda,\mu) \le f(x^*)$$



Как приблизить $g(\lambda, \mu)$ к $f(x^*)$?

$$\begin{cases} g(\lambda, \mu) \to \max\limits_{\lambda, \mu} \\ \lambda_i \geq 0 \end{cases}$$

Двойственная задача - поиск самой хорошей нижней оценки

Исходная задача $\to$ Двойственная задача $\to$ Решаем её

x^* - решение прямой задачи

λ^*, μ^* - решение дв. задачи

$g(\lambda^*, \mu^*) \leq f(x^*)$ слабая двойственность

$=$ сильная двойственность (Условие Слейтера)

Упражнение

$$\begin{cases} \frac{1}{2} \|x\|_2^2 = \frac{1}{2} x^T x \to \min\limits_x \\ Ax = b \end{cases}$$

$\mathcal{L}: \mathbb{R}^n \to \mathbb{R}$

$d\mathcal{L} = \nabla \mathcal{L}^T dx$

$$\mathcal{L}(x, \mu) = \frac{1}{2} x^T x + \mu^T (Ax - b)$$

$$d\mathcal{L} = \frac{1}{2} \underbrace{dx^T \cdot x}_{[1\times 1]} + \frac{1}{2} x^T \cdot dx + \mu^T A dx = (x^T + \mu^T A) dx$$

$S_{1\times 1} \quad S^T = S$

$$\frac{\partial \mathcal{L}}{\partial x} = x + A^T \mu = 0$$

$$x^* = -A^T \mu$$

$$g(\mu) = \inf_x \mathcal{L}(x, \mu) = \frac{1}{2} (A^T \mu)^T A^T \mu + \mu^T (-AA^T \mu - b) =$$

$\mu^T A A^T \mu$

$= -\frac{1}{2}\mu^T A A^T \mu - \mu^T b \to \max\limits_{\mu}$

Задание: $\mu^* - ?$ $g(\mu^*) \leq f(x^*)$ – Правда ли это?

Упражнение

$$\begin{cases} c^T x \to \min\limits_x \\ Ax = b \\ x_j \geq 0 \quad \forall j \end{cases}$$

$-x_j \leq 0$

ф. линейная

$$\mathcal{L}(x, \lambda, \mu) = c^T x - \lambda^T (x - 0) + \mu^T (Ax - b)$$

$$d\mathcal{L} = c^T dx - \lambda^T dx + \mu^T A\, dx$$

$$c - \lambda + A^T \mu = 0 \quad \cancel{\ }$$

$$\mathcal{L}(x, \lambda, \mu) = (c - \lambda + A^T \mu)^T x - \mu^T b \to \min\limits_x$$

$c - \lambda + A^T \mu = 0$ $\quad \inf\limits_x \mathcal{L}(x, \lambda, \mu) = -\mu^T b$

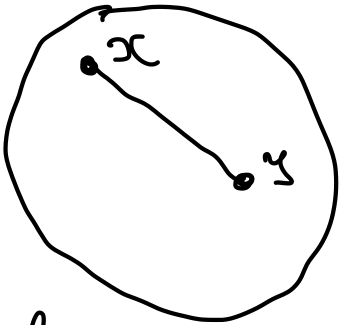
$\neq 0$ $\quad x \to -\infty$

$$g(\lambda, \mu) = \begin{cases} -\mu^T b, & c - \lambda + A^T \mu = 0 \\ -\infty & c - \lambda + A^T \mu \neq 0 \end{cases}$$

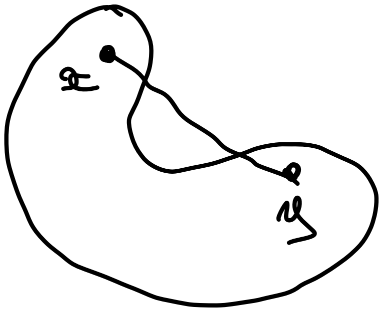
Я хочу $g(\lambda, \mu) \to \max\limits_{\lambda, \mu}$ => вторая строчка отваливается

$$\begin{cases} -\mu^T b \to \max\limits_{\mu, \lambda} \\ c - \lambda + A^T \mu = 0 \\ \lambda \geq 0 \end{cases}$$

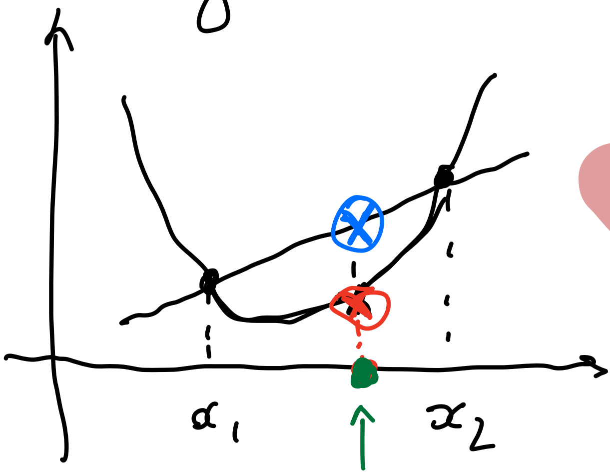
Двойственная задача.



выпуклое



не выпуклое



$tx_1+(1-t)x_2$

$$f(tx_1+(1-t)x_2) \le t f(x_1)+(1-t)f(x_2)$$

$t\in(0;1)$

выпукла вниз ф.

Условия Каруша-Куна-Такера

1939 диплом

1951

Сильная двойственность $g(\lambda^*,\mu^*) = f(x^*)$

Условия Слейтера:

① Выпуклая опт. задача

② $\exists$ допустимая точка x': $g_i(x')<0$

Внутренняя допустимая точка

Достаточные условия сильной двойственности

$$f(x^*) = g(\lambda^*,\mu^*) = \inf_x \left[f(x) + \sum_i \lambda_i^* g_i(x) + \sum_j \mu_j^* h_j(x)\right]$$

$$\le f(x^*) + \boxed{\sum_i \lambda_i^* g_i(x^*)} + \sum_j \mu_j^* h_j(x^*) \le f(x^*)$$

≤ 0 $\qquad = 0$

$= 0$

1) Если решить прямую задачу

$$\mathcal{L}(x, \lambda^*, \mu^*) \to \min_x \Rightarrow x^*$$

2) $\sum_i \underbrace{\lambda_i^*}_{\geq 0} \underbrace{g_i(x^*)}_{\leq 0} = 0 \qquad \lambda_i^* g_i(x^*) = 0 \quad \forall i$

Условие дополнительной нежесткости

Т. (Каруша - Куна - Такера)

Пусть x^* - решение прямой задачи, тогда

$\exists\ \mu^*, \lambda^*$:

$$\begin{cases} \nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0 \\ g_i(x^*) \leq 0 \quad \forall i \\ h_j(x^*) = 0 \quad \forall j \\ \lambda_i^* \geq 0 \quad \forall i \\ \lambda_i^* g_i(x^*) = 0 \quad \forall i \end{cases}$$

- необходимость доказали выше
- Достаточность забили

Упражнение

$$\begin{cases} (x-4)^2 + (y-4)^2 \to \min\limits_{x,y} \\ x+y \leq 4 \qquad y \leq 4 - x \\ x+3y \leq 9 \qquad y \leq 3 - \frac{1}{3}x \end{cases}$$

$$\mathcal{L} = (x-4)^2 + (y-4)^2 + \lambda_1(x+y-4) + \lambda_2(x+3y-9)$$

$$\begin{cases} 2(x-4) + \lambda_1 + \lambda_2 = 0 \\ 2(y-4) + \lambda_1 + 3\lambda_2 = 0 \\ x+y \le 4 \qquad \lambda_1 \ge 0 \\ x+3y \le 9 \qquad \lambda_2 \ge 0 \end{cases}$$

$$\nabla \mathcal{L} = 0$$

$$\lambda_1(x+y-4) = 0$$

$$\lambda_2(x+3y-9) = 0$$

Решаем (4 случая)

- $x+y-4=0 \quad x+3y=9 \quad \lambda_1 > 0 \quad \lambda_2 > 0$

$$\begin{cases} x+y=4 \\ x+3y=9 \end{cases}$$

$$x + 3(4-x) = 9$$

$$x = 3/2 \quad y = 5/2$$

$$\begin{cases} \cancel{2\left(\tfrac{3}{2}-4\right)} + \lambda_1 + \lambda_2 = 5 \\ \cancel{2\left(\tfrac{5}{2}-4\right)} + \lambda_1 + 3\lambda_2 = 3 \end{cases}$$

$$-2\lambda_2 = 2$$

$$\lambda_2 = -1$$

$$\cancel{\lambda_2 > 0}$$

$\cancel{\varnothing}$

- $x+y=4$
  $x+3y<9$
  $\lambda_1 > 0$
  $\lambda_2 = 0$

$$\begin{cases} 2(x-4) + \lambda_1 = 0 \\ 2(y-4) + \lambda_1 = 0 \\ x+y=4 \end{cases}$$

$$x = y = 2$$

$$\lambda_1 = 4$$

√

$$x+3y<9$$

$$2+6=8<9$$

- Оставшиеся 2 случая дадут $\varnothing$

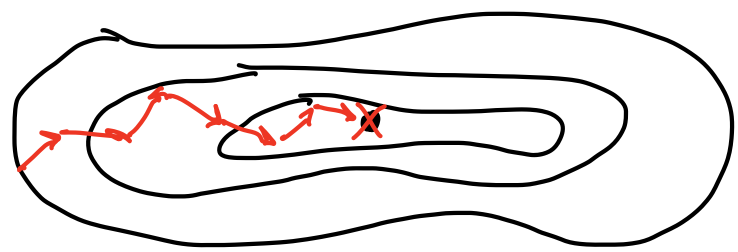
Итеративные методы

$$\begin{cases} f(x) \to \min\limits_{x \in \mathbb{R}^n} \\ g_i(x) \le 0 \quad i = 1, \dots, m \\ h_j(x) = 0 \quad j = 1, \dots, p \end{cases}$$

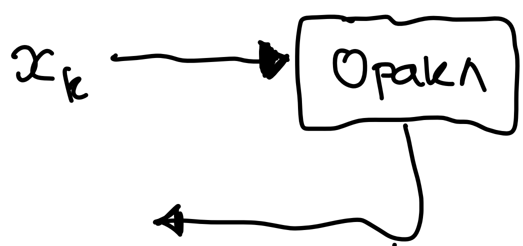
Не всегда понятно как это вообще решать

$\Rightarrow$ iterative methods

$(x_k)_{k=1}^{\infty}$

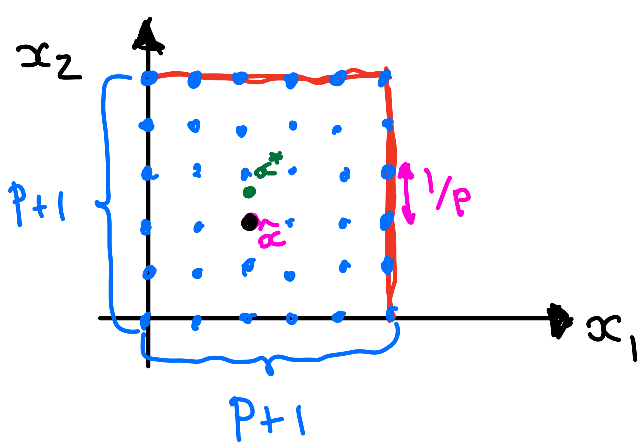


Oracle conception



$f(x_k)$	$f'(x_k)$	$f''(x_k)$	...
нулевого порядка	первого порядка	второго порядка	...

Упражнение



Перебор по решётке

$(p+1)^n$ вариантов

$$\begin{cases} f(x) \to \min\limits_{x \in \mathbb{R}^n} \\ x \in B^n \\ B^n = \{x \in \mathbb{R}^n \mid 0 \le x_i \le 1 \ \forall i\} \end{cases}$$

Цель: найти оптимальное x^* (не сможем)

можем найти $\tilde{x}$

$$|f(\tilde{x}) - f(x^*)| \le \varepsilon$$

$\|\tilde{x} - x^*\|_\infty \le \frac{1}{2p}$

$|f(\tilde{x}) - f(x^*)| < ?$

Для оценки этого расстояния на f накладывают огранич.

Липшецевость

$|f(x) - f(y)| \le L \cdot \|x - y\|_\infty$ $\quad \forall x, y$

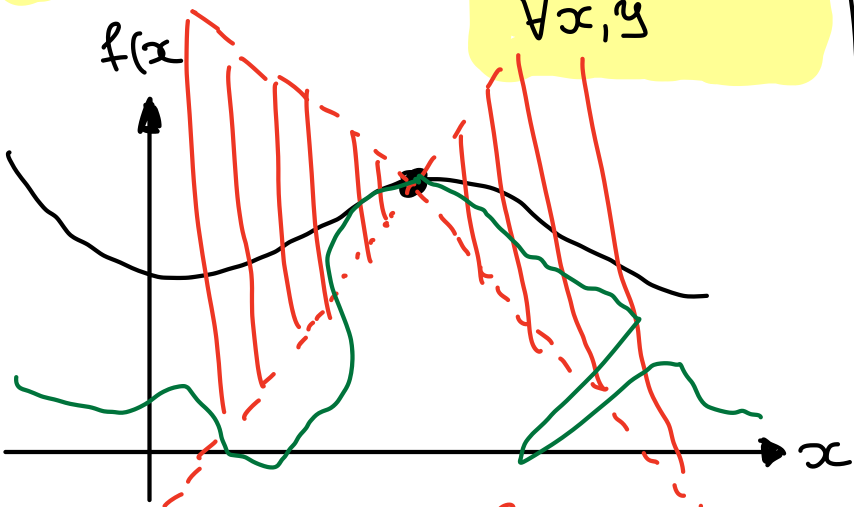


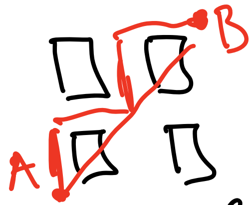
график внутри конуса

норма минковского:

$\|x\|_p = \sqrt[p]{\sum_i |x_i|^p}$

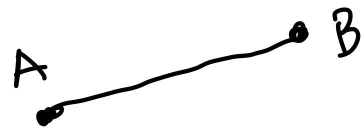
$p = 1$ манхэтенская норма

$\sum_i |x_i|$



$p = 2$ евклидова норма

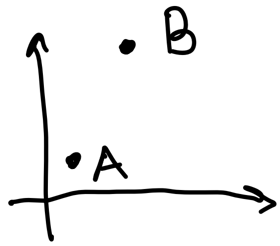
$\sqrt{\sum_i x_i^2}$



$p = \infty$

$\max_i |x_i|$

$\max(1,2) = 2$



$$|f(\tilde{x}) - f(x^*)| \le L \cdot \|x - y\|_\infty \le \frac{L}{2p} = O\left(\frac{1}{p}\right)$$

$= \varepsilon$

$O\left(\frac{1}{\sqrt[n]{K}}\right)$

$p = \frac{L}{2\varepsilon}$

$p + 1 = \frac{L}{2\varepsilon} + 1$

$(p+1)^n = \left(\frac{L}{2\varepsilon} + 1\right)^n$

$\left(\frac{1}{\varepsilon} + 1\right)^n$

$K = (p+1)^n$

$L = 2$ ф. может распечататься не более чем в два раза

$n = 2$ $\quad \lambda$ learning rate

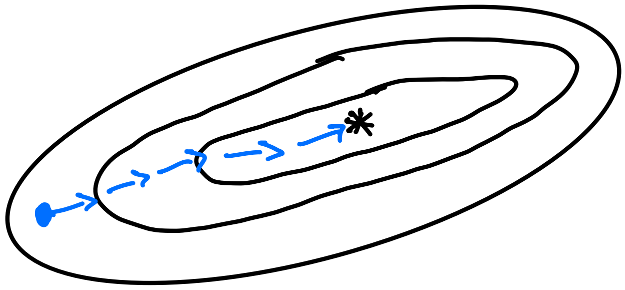
$\varepsilon = 1$ $\quad$ 4 варианта

k - кол-во итераций до сходимость $\qquad \varepsilon = 0.01 \qquad 101^2 > 10000$

Градиентный спуск

$f(x)$ – гладкая т.е. градиент липшецев

$$Q(w) = \sum_i (y_i - x_i^T w)^2 + \|w\|_2^2 \to \min_w$$



init: w^0

$$w^t = w^{t-1} - \eta_t \cdot \nabla_w Q(w^{t-1})$$

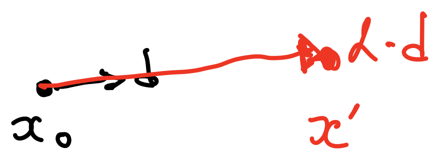
Gradient Descent

Направление антиградиента – направл. наиск. убывания

$x_0 + \alpha \cdot d \qquad \|d\| \leq 1 \qquad \alpha \in \mathbb{R}$

$x' = x_0 + \alpha \cdot d$

$x' - x_0 = \alpha \cdot d$



$f(x') < f(x_0)$

$$f(x') \approx f(x_0) + \nabla^T f(x_0) \cdot (x' - x_0)$$

$$\cancel{f(x_0)} + \nabla^T f(x_0) \underbrace{(x' - x_0)}_{\alpha \cdot d} < \cancel{f(x_0)}$$

$$\nabla^T f(x_0) \cdot \alpha \cdot d < 0 \quad | : \alpha$$

$$\nabla^T f(x_0)\cdot d < 0$$

$$f(x') \longrightarrow \min_d$$

$$f(x_0) + \nabla^T f(x_0)\cdot \alpha \cdot d \longrightarrow \min_d$$

$$\nabla^T f(x_0)\cdot \cancel{\alpha}\cdot d \longrightarrow \min_d \qquad \|d\|^2 = 1$$

$$\mathcal{L} = \nabla^T f(x_0)\cdot d + \lambda\cdot(d^T d - 1)$$

$$\frac{\partial \mathcal{L}}{\partial d} = \nabla f(x_0) + 2\lambda d = 0 \Rightarrow d = -\frac{1}{2\lambda}\cdot\nabla f(x_0)$$

$$\|d\|^2 = \frac{1}{4\lambda^2}\cdot\|\nabla f(x_0)\|^2 = 1$$

$$\frac{1}{2\lambda} = \frac{1}{\|\nabla f(x_0)\|}$$

$$d = -\frac{\nabla f(x_0)}{\|\nabla f(x_0)\|}$$

$$\underset{x'}{x^{t+1}} = \underset{x_0}{x^t} - \underbrace{\frac{\alpha}{\|\nabla f(x_0)\|}}_{\eta}\cdot\nabla f(x^t)$$

Упражнение

$$Q(w) = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i^T w)^2 \to \min_w$$

$$\nabla_w Q = -2 \cdot \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i^T w)\,x_i$$

$$w^t = w^{t-1} - 0.01 \cdot \left(-2 \cdot \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i^T w^{t-1})\,x_i\right)$$

n большое
=> долго считать

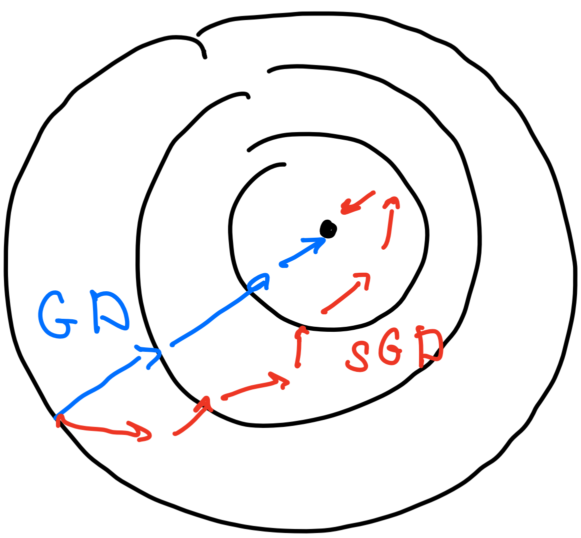
SGD

$1, \dots, n$ shuffle

$$\hat{\nabla}_w Q = -2 \cdot (y_i - x_i^T w)\,x_i$$

learning rate

$$w^t = w^{t-1} - 0.01 \cdot \left(-2 \cdot (y_i - x_i^T w^{t-1})\,x_i\right)$$



Mini-Batch SGD

$1, 2, 3, \dots\dots n$ shuffle

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

$$\sum_{i=1}^{m}$$

$m \ll n$

lr – надо подбирать

$(y_i - x_i^T w)$

$$Q(w) = \frac{1}{n}\sum_i L(w, x_i, y_i)$$

модификации:

① Momentum

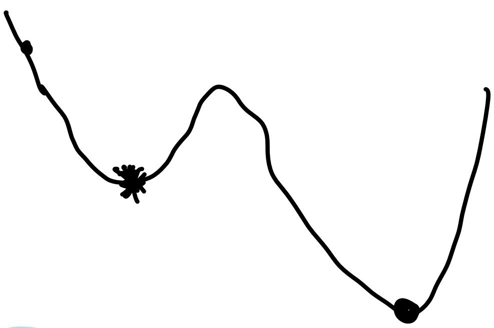
$$g_t = \frac{1}{m}\sum_{i=1}^{m} \nabla_w L(w^{t-1}, x_i, y_i)$$

$$w_t = w_{t-1} - \eta \cdot g_t$$

$$h_t = \alpha \cdot h_{t-1} + \eta \cdot g_t$$

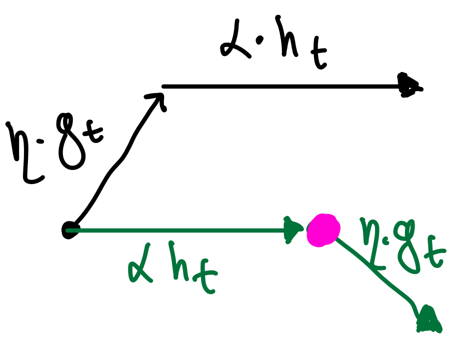
$$w_t = w_{t-1} - h_t$$

$$w_{t-1} - \alpha h_{t-1} - \eta g_t$$



мультистарт

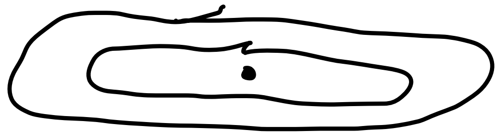
② Nesterov Momentum



$$h_t = \alpha \cdot h_{t-1} + \eta \cdot \nabla_w L(w_{t-1} - \alpha h_{t-1})$$

$$w_t = w_{t-1} - h_t$$

③ Rmsprop / Adagrad



$$\boxed{g_t = \frac{1}{m}\sum_{i=1} \nabla_w L(w^{t-1})}$$

оценка градиента

$$w_t = w_{t-1} - \eta \cdot g_t$$

$$- \eta \cdot \mathbb{E}[g_t]$$

θ $\quad \hat{\theta}$ $\quad \mathbb{E}(\hat{\theta}) = \theta$

$\mathrm{Var}(\hat{\theta})$

$$\mathrm{Var}(g_t) = \mathbb{E}(g_t^2) - \mathbb{E}(g_t)^2$$

g_t^2

$$G_t^j = G_{t-1}^j + (g_t^j)^2$$

$$w_t^j = w_{t-1}^j - \frac{\eta}{\sqrt{G_t^j + 10^{-6}}} \cdot g_t^j$$

Adagrad

Rmsprop

$$G_t^j = \alpha G_{t-1}^j + (1-\alpha)(g_t^j)^2$$

$$w_t^j = w_{t-1}^j - \frac{\eta}{\sqrt{G_t^j + 10^{-6}}} \cdot g_t^j$$

④ Adam

$$h_t^j = \frac{\beta_1 h_{t-1}^j + (1-\beta_1) g_t^j}{1-(\beta_1)^t}$$

$$G_t^j = \frac{\beta_2 G_{t-1}^j + (1-\beta_2)(g_t^j)^2}{1-(\beta_2)^t}$$

$$w_t^j = w_{t-1}^j - \frac{\eta_t}{\sqrt{G_t^j + 10^{-5}}} \cdot h_t^j$$

AdaBelief

$$h_t^j = \frac{\beta_1 h_{t-1}^j + (1-\beta_1) g_t^j}{1-(\beta_1)^t}$$

$$G_t^j = \frac{\beta_2 G_{t-1}^j + (1-\beta_2)(g_t^j - h_t^j)^2}{1-(\beta_2)^t}$$

$$w_t^j = w_{t-1}^j - \frac{\eta_t}{\sqrt{G_t^j + 10^{-5}}} \cdot h_t^j$$

Упражнение

Откуда в Adam $\frac{1}{1-\beta_1^2}$ 1 параметр

$$h_t = \beta_1 \cdot h_{t-1} + (1-\beta_1)\, g_t \qquad \mathbb{E}(h_t) = \mathbb{E}(g_t)$$

$$G_t = \beta_2 G_{t-1} + (1-\beta_2)\, g_t^2 \qquad \mathbb{E}(G_t) = \mathbb{E}(g_t^2)$$

$$w_t = w_{t-1} - \frac{\eta_t}{\sqrt{G_t+\varepsilon}}\, h_t$$

$$h_0 = 0$$

$$h_1 = \beta_1 h_0 + (1-\beta_1)\, g_1 = (1-\beta_1)\, g_1$$

$$h_2 = \beta_1 h_1 + (1-\beta_1)\, g_2 = \beta_1(1-\beta_1)\, g_1 + (1-\beta_1)\, g_2$$

$$h_3 = \beta_1 h_2 + (1-\beta_1)\, g_3 = \beta_1^2(1-\beta_1)\, g_1 + \beta_1(1-\beta_1)\, g_2 + (1-\beta_1)\, g_3$$

....

$$h_k = (1-\beta_1) \sum_{i=1}^{k} \beta_1^{k-i} \cdot g_i$$

$$\mathbb{E}(h_k) = \mathbb{E}\left[(1-\beta_1) \sum_{i=1}^{k} \beta_1^{k-i} \cdot g_i\right] =$$

$$= (1-\beta_1) \sum_{i=1}^{k} \beta_1^{k-i} \cdot \mathbb{E}(g_i) = (1-\beta_1) \cdot \mathbb{E}(g_i) \sum_{i=1}^{k} \beta^{k-i} =$$

$\mathbb{E}(g)$ смещение

$$= \mathbb{E}(g_i) \cdot (1-\beta_1) \cdot \frac{1-\beta_1^k}{1-\beta_1} = (1-\beta_1^k) \cdot \mathbb{E}(g_i)$$

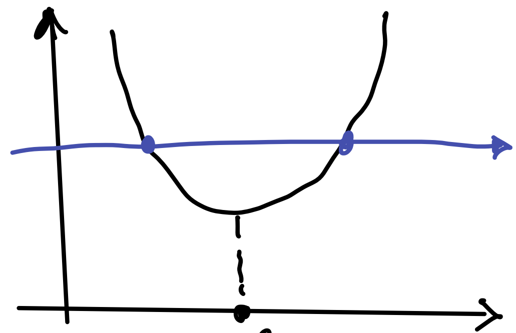
Методы второго порядка

Упражнение

$$f(x) = ax^2 + bx + c$$

$$x_t = x_{t-1} - \gamma \cdot \nabla_x f(x_{t-1})$$

Какое γ надо взять, чтобы попасть в оптимум за шаг



$$x^* = -\frac{b}{2a}$$

$$x_0 \qquad \nabla_x f(x_0) = 2ax_0 + b$$

$$x^* = x_0 - \gamma(2ax_0 + b)$$

$$-\frac{b}{2a} = x_0 - \gamma(2ax_0 + b)$$

$$\gamma(\cancel{2ax_0 + b}) = x_0 + \frac{b}{2a} = \frac{\cancel{2ax_0 + b}}{2a}$$

$$\gamma = \frac{1}{2a}$$

$$f'(x) = 2ax + b$$

$$f''(x) = 2a \qquad \gamma = \frac{1}{f''(x)}$$

$$x_t = x_{t-1} - \frac{\eta_t}{f''(x_{t-1})} \cdot f'(x_{t-1})$$

$\boxed{\text{Метод Ньютона}}$

Упражнение

$$f(x) = x^T A x + b^T x + c$$

$x \in \mathbb{R}^n$ $\quad f: \mathbb{R}^n \to \mathbb{R}$ $\quad$ $A^T = A$ ⊕ полож. опр.

$$x_t = x_{t-1} - \Gamma \nabla_x f(x_{t-1})$$

Найти Γ так, чтобы экстр. нашелся за один шаг

$$df = dx^T \cdot Ax + x^T A\, dx + b^T dx = (2x^T A + b^T)\, dx$$

$$x^T A^T dx$$

$$x^T A\, dx$$

$$\nabla_x f = 2A^T x + b = 2Ax + b$$

$$H = \nabla_x^2 f = 2A \qquad \boxed{df = \nabla^T f\, dx}$$

$$x^* = ? \qquad 2Ax + b = 0 \qquad x^* = -\frac{1}{2} A^{-1} b$$

$$-\frac{1}{2} A^{-1} b = x_0 - \Gamma \cdot (2Ax_0 + b)$$

$$\underbrace{\Gamma(2Ax_0 + b)} = \underbrace{x_0 + \frac{1}{2} A^{-1} b}$$

$$2A \cdot$$

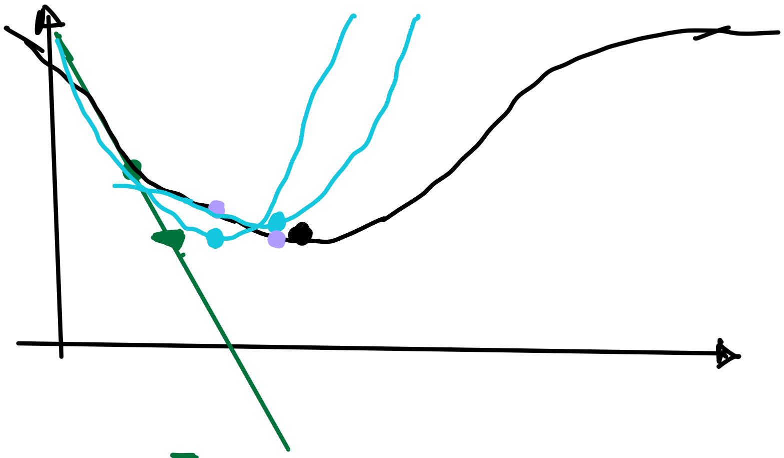
$$2 \cdot A \cdot \Gamma(2Ax_0 + b) = (2Ax_0 + b) \quad \Big| \cdot (2Ax_0 + b)^{-1}$$

$$2A \cdot \Gamma = I$$

$$\Gamma = \frac{1}{2} \cdot A^{-1} = H^{-1} = (\nabla_x^2 f)^{-1}$$

$$x_t = x_{t-1} - H^{-1}(x_{t-1}) \cdot \nabla f(x_{t-1})$$

$$\overset{II}{f}(x') \approx f(x_0) + \nabla^T f(x_0) \cdot (x' - x_0) + \frac{1}{2}(x' - x_0)^T H(x_0)(x' - x_0)$$



$$\overset{I}{f}(x) = f(x_0) + \nabla^T f(x_0)(x - x_0)$$

Упражнение

$$f(x) = x^T A x + b^T x + c \qquad A^T = A \qquad A \succ 0$$

Как связаны:

$(x_1 - x_0)$ и $\nabla f(x_1) - \nabla f(x_0)$

$$x_1 = x_0 - \frac{1}{2} A^{-1} \nabla f(x_0)$$

$$x_1 - x_0 = \frac{1}{2} A^{-1} \big(\underbrace{0}_{= \nabla f(x_1)} - \nabla f(x_0) \big)$$

$\nabla f(x_1)$ т.к. x_1 - оптимум

$$x_1 - x_0 \approx \gamma \cdot (\nabla f(x_1) - \nabla f(x_0))$$

γ - ? чтобы равенство не сильно ухудшилось

$y = wx \qquad MSE \to \min_w$

$$\hat{w} = \frac{\sum_i y_i x_i}{\sum_i x_i^2} = \frac{y^T x}{x^T x}$$

$$\hat{\gamma} = \frac{(x_1 - x_0)^T (\nabla f(x_1) - \nabla f(x_0))}{(\nabla f(x_1) - \nabla f(x_0))^T \cdot (\nabla f(x_1) - \nabla f(x_0))}$$

Апроксимация:

$$x_t = x_{t-1} - \frac{(x_1 - x_0)^T (\nabla f(x_1) - \nabla f(x_0))}{(\nabla f(x_1) - \nabla f(x_0))^T \cdot (\nabla f(x_1) - \nabla f(x_0))} \cdot \nabla f(x_{t-1})$$

$$\boxed{\begin{aligned} x_1 &= x_{t-1} \\ x_0 &= x_{t-2} \end{aligned}}$$

Сходимость градиентного спуска в гладком выпуклом случае

1) $f(x): \mathbb{R}^n \to \mathbb{R}^m$

f – L-липшецева если

$$\forall x \in \mathbb{R}^n \qquad \| \nabla f(x) \| \le L$$

2) f L-гладкая

(∇f – липшевицев с конст L)

т.е. $\| \nabla^2 f(x) \| \le L$

Если 2) тогда $\forall x, y \in \mathbb{R}^n$

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T H (y-x)$$

$$f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\|y-x\|^2$$

$$\left\| \frac{1}{2}(y-x)^T H (y-x) \right\| \leq \frac{1}{2}\cdot\|y-x\|\cdot\underbrace{\|H\|}_{\leq L}\cdot\|y-x\| \leq \frac{L}{2}\|y-x\|^2$$

$$\|x\cdot y\| \leq \|x\|\cdot\|y\|$$

$$x_t = x_{t-1} - \alpha \cdot \nabla f(x_{t-1})$$

$$-\alpha \nabla f(x_{t-1})$$

$$f(x_t) \leq f(x_{t-1}) + \langle \nabla f(x_{t-1}); x_t - x_{t-1} \rangle + \frac{L}{2}\|x_t - x_{t-1}\|^2$$

$$f(x_t) \leq f(x_{t-1}) - \alpha \langle \nabla f(x_{t-1}), \nabla f(x_{t-1}) \rangle + \frac{L\cdot\alpha^2}{2}\|\nabla f(x_{t-1})\|^2$$

$$\|\nabla f(x_{t-1})\|^2$$

$$f(x_t) \leq f(x_{t-1}) + \left(\frac{\alpha^2\cdot L}{2} - \alpha\right)\|\nabla f(x_{t-1})\|^2$$

хочу сделать её маленькой

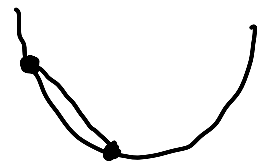
$$\frac{L}{2}\alpha^2 - \alpha \to \min_\alpha \qquad \alpha^* = \frac{1}{L}$$

$$f(x_t) \leq f(x_{t-1}) - \frac{1}{2L}\|\nabla f(x_{t-1})\|^2$$

$$f(x_{t-1}) - f(x_t) \geq \frac{1}{2L} \|\nabla f(x_{t-1})\|^2$$

Заметаю на каждом шаге такую величину

$f(x)$ - выпуклая



$$f(y) \geq f(x) + \nabla^T f(x)(y - x)$$

$$f(x_t) \geq f(x^*) + \nabla^T f(x^*)(x_t - x^*)$$

$$f(x_t) \geq f(x^*)$$

$$\boxed{f(x_t) - f(x^*) \leq \varepsilon} \qquad \varepsilon(t) - ?$$

$$f(x_t) - f(x^*) \leq \nabla^T f(x_{t-1}) \cdot (x_{t-1} - x^*)$$

$$f(x_t) \leq f(x_{t-1}) + \left(\frac{\alpha^2 \cdot L}{2} - \alpha\right) \|\nabla f(x_{t-1})\|^2 \leq$$

$$\leq f(x^*) + \nabla^T f(x_{t-1})(x_{t-1} - x^*)$$

$$\leq f(x^*) +$$