Пример з.о. в м.о.:

$\{a_i, b_i\}_{i=1}^n$ – обуч. выборка

объекты   лейбл

- мин. эмпирич. риска (ERM):

модель

$$\min_{x \in \mathbb{R}^d} \left\{ \hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} l\left(g(x, a_i), b_i \right) \right\}$$

ф. потерь   веса

- мин. ожид. ф. потерь: реальн. природы (мы не знаем)

a_i, b_i – приходят из $\mathcal{D}$

$$\min_{x \in \mathbb{R}^d} \left\{ f(x) = \mathbb{E}_{(a,b)\sim\mathcal{D}} \left[l\left(g(x,a), b \right) \right] \right\}$$

NB 1 постан. – это Монте-Карло оценка 2 постановки

$l(g(x, a_i), b_i)$ выпукла по x, М-липшиц. по x для всех a_i, b_i

$\operatorname{argmin} \hat{f}(x)$

$\Downarrow$

$$\left| f(\hat{x}^*) - \min f(x) \right| \sim \frac{1}{\sqrt{n}}$$

- Пусть

$$b(a_1 \ldots a_{d-1}) = x_0 + x_1 a_1 + \ldots + x_{d-1} a_{d-1}$$

неизвестны

$$b_i = x_0 + x_1 a_1 + \ldots + x_{d-1} a_{d-1} + \xi_i$$

$$\xi_i \sim N(0, \sigma^2)$$

i.i.d. $\rightarrow$ $b_i \sim N(x_0 + \ldots + x_{d-1} a_{i,d-1}, \sigma^2)$

n i.i.d. объектов с неиз. м.о. $\Rightarrow$ макс. правд.:

$$x^* = \arg\max_{x \in \mathbb{R}^d} \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(b_i - a_i^T x)^2\right)$$

$$= \arg\max_{x \in \mathbb{R}^d} \ln\left(\prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(b_i - a_i^T x)^2\right)\right)$$

$$= \arg\max_{x \in \mathbb{R}^d} \left\{ Const + \sum_{i=1}^n -\frac{1}{2\sigma^2}(b_i - a_i^T x)^2 \right\}$$

$$= \arg\min_{x \in \mathbb{R}^d} \left\{ \sum_{i=1}^n (b_i - a_i^T x)^2 \right\}$$

$$= \arg\min \left\{ \frac{1}{n} \sum_{i=1}^n (b_i - a_i^T x)^2 \right\}$$

Задача мин. эмп. риска с g - линейной

L - квадратичной

Задача оптимизации:

целев. ф.

$$\min_{\substack{x \in \mathbb{R}^d \\ x \in Q}} f(x)$$

размерность задачи

мн.-во огранич. („простое")

пример: – шар
– симплекс

$$g_i(x) \le 0$$
$$h_j(x) = 0$$

NB

- может не иметь решений $\min_{x \in \mathbb{R}} x$
- сколько решений либо нет, либо много
- сложность задачи зав. от f, мн.-ва Q, от разм. d

Оракул — процедура, дающая инф. о функции

Примеры оракулов

- **Оракул нулевого порядка** в запрашиваемой точке x возвращает значение целевой функции $f(x)$.
- **Оракул первого порядка** в запрашиваемой точке возвращает значение функции $f(x)$ и её градиент в данной точке $\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n}\right)$.
- **Оракул второго порядка** в запрашиваемой точке возвращает значение и градиент функции $f(x), \nabla f(x)$, а также её гессиан в данной точке $\left(\nabla f^2(x)\right)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$.

Пример метода

Входные данные: начальная точка x^0 (0 – верхний индекс), требуемая точность решения задачи $\varepsilon > 0$.
Настройка. Задать $k = 0$ (счётчик итераций) и $I_{-1} = \varnothing$ (накапливаемая информационная модель решаемой задачи).
Основной цикл

1. Задать вопрос к оракулу $\mathcal{O}$ в точке x^k.
2. Пересчитать информационную модель: $I_k = I_{k-1} \cup (x^k, \mathcal{O}(x^k))$.
3. Применить правило метода $\mathcal{M}$ для получения новой точки x^{k+1} по модели I_k.
4. Проверить критерий остановки $\mathcal{T}_\varepsilon$. Если критерий выполнен, то выдать ответ $\bar{x}$, иначе положить $k := k + 1$ и вернуться на шаг 1.

Пример:

$$\min_{x \in \mathbb{R}^d} f(x)$$

Алгоритм 1 Градиентный спуск с постоянным размером шага

Вход: размер шага $\gamma > 0$, стартовая точка $x^0 \in \mathbb{R}^d$, количество итераций K

```
for k = 0, 1, ..., K − 1 do
    Вычислить ∇f(x^k)
    x^{k+1} = x^k − γ∇f(x^k)
end for
```

Выход: x^K

Критерий остановки (точность решения):

- По аргументу

решение

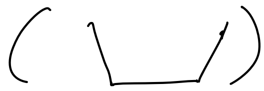
NB норма евклидова

$$\|x^k - x^*\| \le \varepsilon$$

неизв.

практ. вид: $\|x^{k+1} - x^k\| \le \varepsilon$

$$\|x^{k+1} - x^k\| \le \underbrace{\underset{\le \varepsilon'}{\underset{\text{сх} \to 0}{\|x^{k+1} - x^*\|}} + \underset{\le \varepsilon'}{\underset{\to 0}{\|x^k - x^*\|}}}_{\le 2\varepsilon'}$$

• x^* — не уникален ()

по функции:

$$f(x^k) - f^* \le \varepsilon$$

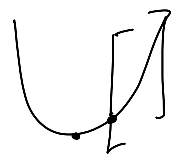
можем не знать

$\min f(x)$

$$|f(x^{k+1}) - f(x^k)|$$

• по норме градиента:

$$\|\nabla f(x^k)\| \le \varepsilon$$



! для безусловных задач (на $\mathbb{R}^d$)

Сложность методов оптим.:

• число итераций $\pm$

• аналит. / оракульная сложность — число обращений к оракулу для дост. точн. ε

• арифмет. / временная сложность — общее число вычислений для дост. точн. ε

Пример (класс задач липш. ф.)

$$\min_{x \in B_d^\infty(1)} f(x)$$

$$B_d^\infty(1) = \{x \in \mathbb{R}^d \mid 0 \le x_i \le 1 \quad i = 1..d\}$$

f — M-липшицева в L_∞-норме

$\forall x, y$

$$|f(x) - f(y)| \le M\|x-y\|_\infty = M \max_{i=1..d} |x_i - y_i|$$

NB: f - непр на B_d (компакт), значит $\exists\, f^* = \min\limits_{x \in B_d} f(x)$

Цель: $\bar{x} \in B_d^\infty(1)$: $f(\bar{x}) - f^* \le \varepsilon$

- Метод:

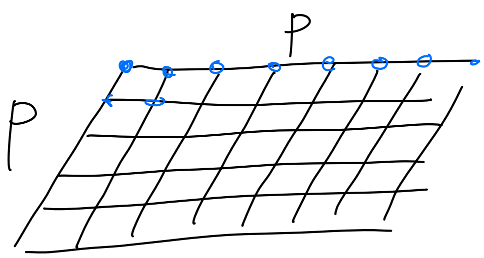
Алгоритм 2 Метод равномерного перебора

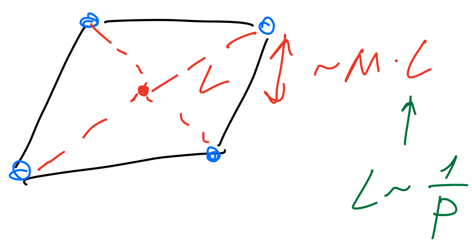
Вход: целочисленный параметр перебора $p \ge 1$

1: Сформировать $(p+1)^d$ точек вида $x_{(i_1,\dots,i_d)} = \left(\frac{i_1}{p}, \frac{i_2}{p}, \dots, \frac{i_d}{p}\right)^\top$, где $(i_1, \dots, i_d) \in \{0, 1, \dots, p\}^d$

2: Среди точек $x_{(i_1,\dots,i_d)}$ найти точку $\bar{x}$ с наименьшим значением целевой функции f.

Выход: $\bar{x}, f(\bar{x})$





Теорема 1

Алгоритм 2 с параметром p возвращает такую точку $\bar{x}$, что

$$f(\bar{x}) - f^* \le \frac{M}{2p},$$

откуда следует, что методу равномерного перебора нужно в худшем случае

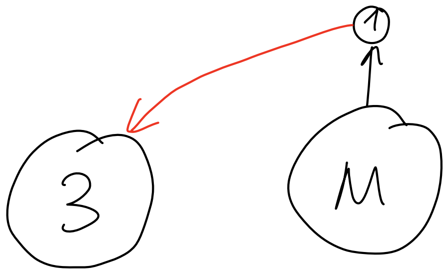
$$\left(\left\lfloor \frac{M}{2\varepsilon} \right\rfloor + 2\right)^d$$

обращений к оракулу, чтобы гарантировать $f(\bar{x}) - f^* \le \varepsilon$.

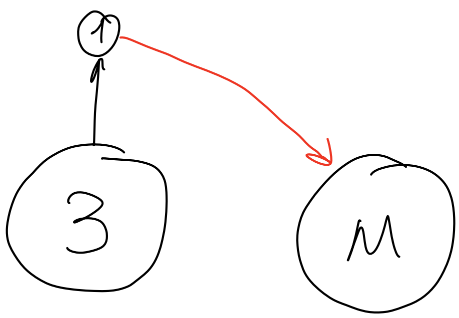
$M = 2$ $\quad d = 13$ $\quad \varepsilon = 10^{-2}$ $\quad \approx 10^{26}$ $\approx$ миллиарды лет

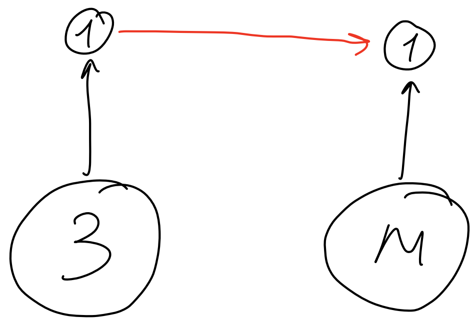
- Верхние и нижние оценки:

Верхняя оценка



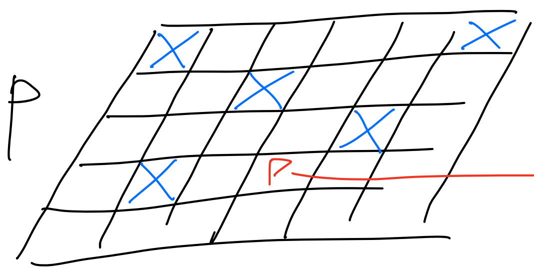
нижняя оценка





- Нижняя оценка

N — число входов функции

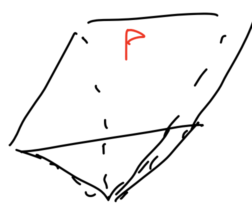


p: $p^d > N$

по принципу Дирихле

1) $f \equiv 0$ метод угадает и не заходит P

2) „запуск" в f ровно в P



$\sim \frac{M}{P}$

Теорема 2

Пусть $\varepsilon < \frac{M}{2}$. Тогда аналитическая сложность описанного класса задач, т.е. аналитическая сложность метода на «худшей» для него задаче из данного класса, составляет по крайней мере

$$\left(\left\lfloor \frac{M}{2\varepsilon} \right\rfloor\right)^d \quad \text{вызовов оракула.}$$

Скорости сходимости:

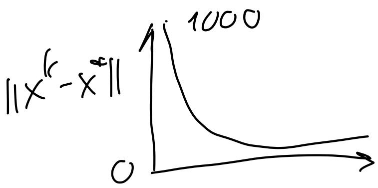
- Сублинейная

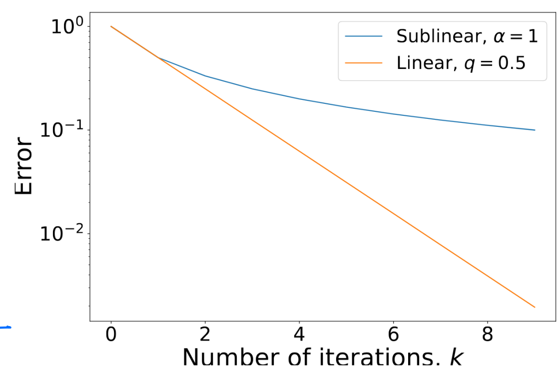
любой кр. сход. → $\|x^k - x^*\| \le \frac{C}{k^\alpha}$ $\quad C > 0 \quad \alpha > 0$

$$\|x^k - x^*\| \le \varepsilon \Rightarrow \frac{C}{k^\alpha} \le \varepsilon \Rightarrow \boxed{k \geq \sqrt[\alpha]{\frac{C}{\varepsilon}}}$$

- Линейная

$$\|x^k - x^*\| \le C q^k \qquad C > 0 \quad q \in (0;1)$$

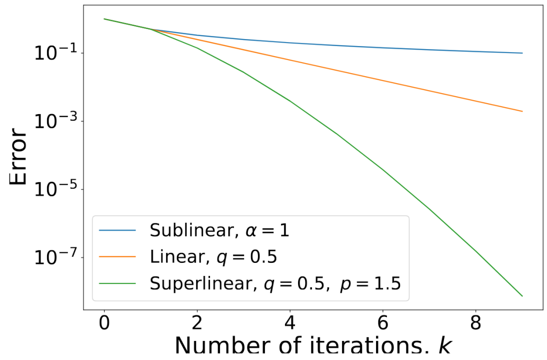




строить в log масштабе!

- Сверхлинейная

$$\|x^k - x^*\| \le C q^{k^p} \qquad C > 0 \quad q \in (0;1) \quad p > 1$$



- Квадратичная

$$\|x^k - x^*\| \le C q^{2^k} \qquad C > 0 \qquad q \in (0;1)$$

или

$$\|x^{k+1} - x^*\| \le C \|x^k - x^*\|^2$$