Zero order methods. Gradient free optimization. Global optimization

- Шпаргалка по результатам в безградиентной оптимизации
- RL и эволюционные алгоритмы

Open in Colab
- Global optimization illustration

Open in Colab
- Nevergrad library

Open in Colab
- Optuna quickstart
- Демонстрация медленности методов нулевого порядка

Open in Colab
- Подбор гиперпараметров модели машинного обучения в Keras с помощью Optuna
- A Tutorial on Zero-Order Optimization

Case 1: 2-Point & Multi-Point Estimators

- A naïve approach:

$$\boxed{\mathsf{G}_f^{2n}(x;u) = \sum_{i=1}^{n} \frac{f(x+ue_i) - f(x-ue_i)}{2u} e_i}$$

- When f is L-smooth, we have

$$\boxed{\|\mathsf{G}_f^{2n}(x;u) - \nabla f(x)\| \le \frac{1}{2} uL\sqrt{n}}$$

where $f^u(x) = \mathbb{E}_{y\sim\tilde{\lambda}}[f(x+uy)]$ is a smooth version of f

$\lambda = \mathrm{Uni}(\mathbb{B}_n)$

Case 1: 2-Point & Multi-Point Estimators

- **2-point** gradient estimator:

$$\mathsf{G}_f^{(2)}(x;u,z) = n\,\frac{f(x+uz) - f(x-uz)}{2u}\,z \qquad z \sim \lambda$$

where λ is **spherically symmetric** with $\mathbb{E}_{z\sim\lambda}\left[\|z\|^2\right] = 1$

$$f^u = \mathbb{E}_z\, G_f^{(2)}$$

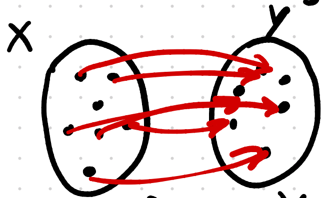
- Some facts for L-smooth / convex / μ-strongly convex function f:
  - f^u is L-smooth / convex / μ-strongly convex
  - $$|f^u(x) - f(x)| \le \frac{1}{2}u^2 L \cdot \frac{n}{n+2}\mathbb{E}_{z\sim\lambda}\left[\|z\|^4\right]$$
    $$\|\nabla f^u(x) - \nabla f(x)\| \le uL \cdot \frac{n}{n+1}\mathbb{E}_{z\sim\lambda}\left[\|z\|^3\right]$$

$$\begin{aligned}&\mathbb{E}_{z\sim\lambda}\left[\left\|\mathsf{G}_f^{(2)}(x;u,z)\right\|^2\right] \\ &\le (1+\kappa)\mathbb{E}_{z\sim\lambda}\left[\|z\|^4\right]\cdot n\|\nabla f(x)\|^2 \\ &\quad + \frac{1+\kappa}{4\kappa}\mathbb{E}_{z\sim\lambda}\left[\|z\|^6\right]\cdot n^2u^2L^2\end{aligned}$$

Методы нулевого порядка.

① Примеры:

- $\nabla f(x)$ вычислить невозможно или очень дорого
- оптимизация гиперпараметров моделей ML



$f: X \to Y$

по наборам (x_i, y_i) восстановить $f(x)$

$f(x_i) = y_i$

$$f(x) = MODEL(\Theta, X)$$

$f(x) = kx + b$

$k=2, b=0$

$x_i = 2 \quad y_i = 4$

$x_i = 3 \quad y_i = 6$

$= \Theta^T \cdot x + \Theta_0$

параметры

$\Theta \in \mathbb{R}^n, \Theta_0 \in \mathbb{R}$

гиперпараметры:

число шагов алгоритма GD, learning rate α } гиперпарам

$$Loss(\text{пар}, \text{гиперп}, \text{вход}) \to \min_{\text{гипер пар.}}$$

- обучение агентов RL
- симуляторы, в которых есть задаваемые параметры

$$x \rightarrow \boxed{} \rightarrow f(x)$$

② TSP
travelling
salesman
PROBLEM

N домов
$(N-1)!$ решений

Решение X — вектор длины N

3	5	7		...	9

$$f(x) = \sum_{i=1}^{N} \| x_{i+1} - \bar{x}_i \|$$

$x_{N+1} = x_1$

$\bar{x}_i$ = координаты домика, который в x_i

Эволюционные алгоритмы
генетический алгоритм

① Генерируем популяцию (набор особей) размера $|P|$
(x)

считаем $f(x)$ $\forall x \in$ популяции

② Скрещивание

$$x_1 \in P \rightarrow x_{12}$$
$$x_2 \in P \rightarrow x_{12}$$

получили $\frac{|P|}{2}$ особей

③ Мутация

$\forall x \in P$

получили $\frac{|P|}{2}$ мутантов

④ Селекция

выбираем $|P|$

<u>Идея:</u>