Лекция 11. Стохастическая оптимиз.

Оценка макс. правд.

$$\underbrace{\xi_1, \dots, \xi_n}_{\text{простая выборка}} \quad (\xi_i \ \text{i.i.d.}) \qquad \xi_i = \begin{cases} 1, & x \\ 0, & 1-x \end{cases}$$

x - не изв. $(x \in [0,1])$

$$\ell(x, \xi) = x^{\xi}(1-x)^{1-\xi} = \begin{cases} x, & \xi = 1 \\ 1-x, & \xi = 0 \end{cases}$$

↑ вер

$$L(x, \{\xi_i\}_{i=1}^n) = \prod_{i=1}^n x^{\xi_i}(1-x)^{1-\xi_i} = x^{\sum_{i=1}^n \xi_i}(1-x)^{n-\sum_{i=1}^n \xi_i}$$

↑ ф-ция правдоподобия.

$x_0 \to \xi_1, \dots, \xi_n$

↗ настоящее значение параметра

$$-\ln L(x, \{\xi_i\}_i^n) = -\sum_{i=1}^n \ln \ell(x, \xi_i)$$

$$x_* = \arg\min_{x \in [0,1]} \mathbb{E}_{\xi \in Be(x_0)} \left[-\ln \ell(x, \xi) \right]$$

$$f(x, \xi) = -\ln(x, \xi)$$

$$\min_x f(x) := \mathbb{E} f(x, \xi) \longrightarrow \text{реш. } x_*$$

не можем посчитать $\mathbb{E}$,

но можем сэмплить:

$\xi_1, \dots, \xi_n$

Два подхода

онлайн

SGD

$$\underbrace{x^{k+1} = x^k - h\nabla f(x^k, \xi^k)}$$

$h \sim \varepsilon / M^2$

$$\bar{x}^N = \frac{1}{N}\sum_{k=0}^{N-1} x^k$$

$$\mathbb{E}[f(\bar{x}^N)] - f(x_*) \le \underbrace{\frac{MR}{\sqrt{N}}}_{\varepsilon}$$

$$N \sim \frac{M^2 R^2}{\varepsilon^2} \quad \Big\{ \quad \mathbb{E}\|\nabla f(x,\xi)\|_2^2 \le M^2$$

$$R = \|x^0 - x_*\|_2$$

Рестарты

$$f(x) - f(x_*) \ge \frac{\mu_\gamma}{2}\|x - x_*\|_2^\gamma, \ \gamma \ge 1$$

$$\frac{\mu_\gamma}{2}\mathbb{E}[\|\bar{x}^N - x_*\|_2^\gamma] \le$$

$$\le \mathbb{E}[f(\bar{x}^N)] - f(x_*) \le$$

$$\le \frac{M\sqrt{\mathbb{E}\|x^0 - x_*\|_2^2}}{\sqrt{N}}$$

$\gamma = 2$

$$\boxed{\mathbb{E}\|\bar{x}^N - x_*\|_2^2 \le \frac{1}{2}\mathbb{E}\|x^0 - x_*\|_2^2}$$

$$x^0 := \bar{x}^N \ ; \ N \sim \frac{M^2}{\mu\varepsilon}, \ \mu_2 = \mu.$$

офлайн

ОМП

О-ценка М-аксим. П-равдоп.

$$\frac{1}{n}\sum_{k=1}^{n} f(x, \xi^k) \to \min_x$$

$\downarrow$

реш. задачи было бы ε-реш. исход.

$$n \sim \frac{M^2}{\mu\varepsilon}$$ (если есть μ-сильн. вып.)

Если $\mu = 0$

$$n \sim d\frac{M^2 R^2}{\varepsilon^2}, \ d = \dim x$$

Регуляризация

$$\mu \sim \varepsilon / R^2$$

$$\frac{1}{n}\sum_{k=1}^{n} f(x, \xi^k) + \frac{\mu}{2}\|x\|_2^2 \to \min_x$$

Батч-параллелизация

Допущ. $\|\nabla f(y)-\nabla f(x)\|_2 \le L\|y-x\|_2$

$$f(x) := \mathbb{E}_\xi f(x,\xi) \to \min_x$$

$$x^{k+1} = x^k - h\nabla_\delta f(x^k)$$

выпуклость

$$(*) \quad -\delta_1 \le f(y) - f(x) - \langle \nabla_\delta f(x), y-x\rangle \le \frac{L}{2}\|y-x\|_2^2 + \delta_2$$

Предполож. $\mathbb{E}\delta_1 \equiv 0, \quad \mathbb{E}\delta_2 \le \delta.$

$$(0) \quad \|x^{k+1}-x_*\|_2^2 = \|x^k - x_*\|_2^2 - 2h\langle \nabla_\delta f(x^k), x^k - x_*\rangle + h^2\|\nabla_\delta f(x^k)\|_2^2$$

Из $(*)$ 1) $\|\nabla_\delta f(x^k)\|_2^2 \le 2L(f(x^k) - f(x^{k+1}) + \delta_2)$

2) $\langle \nabla_\delta f(x^k), x^k - x_*\rangle \ge f(x^k) - f(x_*) - \delta_1$

Подставляем в (0)

$$2h\left(f(x^k) - f(x_*) - \delta_1\right) \le \|x^k - x_*\|_2^2 - \|x^{k+1} - x_*\|_2^2 + h^2 2L\left(f(x^k) - f(x^{k+1}) + \delta_2\right)$$

$h = \frac{1}{L}$

$$\frac{2}{L}\left(f(x^{k+1}) - f(x_*) - \delta_1\right) \le \|x^k - x_*\|_2^2 - \|x^{k+1}-x_*\|_2^2 + \frac{2\delta_2}{L}$$

$\frac{L}{2N}\sum_{k=0}^{N-1}$

$$\boxed{f(\bar{x}^N) - f(x_*) - \delta_1 \le \frac{LR^2}{2N} + \delta_2} \quad (*)$$

$$\frac{1}{N}\sum_{k=0}^{N-1} f(x^{k+1}) \ge f\left(\frac{1}{N}\sum_{k=0}^{N-1} x^{k+1}\right) = f\Big(\underbrace{\frac{1}{N}\sum_{k=1}^{N} x^k}_{\bar{x}^N}\Big)$$

$\mathbb{E}\ (*)$ $\qquad \mathbb{E}\delta_1 = 0$ $\qquad \mathbb{E}\delta_2 \le \delta$

$$\mathbb{E} f(\bar{x}^N) - f(x_*) \le \frac{LR^2}{2N} + \delta \quad (\Diamond)$$

Батчирование $\qquad R^2 = \|x^0 - x_*\|_2^2$

$$\nabla_\delta f(x) = \frac{1}{r}\sum_{i=1}^{r} \nabla_x f(x, \xi^i) = \underbrace{\nabla^r f(x, \{\xi\})}_{\nabla^r f(x)}$$

Предположения: $\begin{cases} \mathbb{E}_\xi \nabla_x f(x,\xi) = \nabla f(x) \\ \mathbb{E}\left[\|\nabla_x f(x,\xi) - \nabla f(x)\|_2^2\right] \le \sigma^2 \end{cases}$

$\Downarrow$

$$\mathbb{E}\left[\|\nabla^r f(x) - \nabla f(x)\|_2^2\right] \le \frac{\sigma^2}{r} \quad (\square)$$

$f(x)$ - выпуклая

$0 \le f(y) - f(x) - \langle \nabla f(x), y - x\rangle$ $\}$ выпуклость

$0 \le f(y) - f(x) - \langle \nabla^r f(x), y - x\rangle + \langle \nabla^r f(x) - \nabla f(x), y - x\rangle$

1) $\underbrace{\langle \nabla f(x) - \nabla^r f(x), y - x\rangle}_{\delta_1} \le f(y) - f(x) - \langle \nabla^r f(x), y - x\rangle$

из $\|\nabla f(y) - \nabla f(x)\|_2 \le L\|y - x\|_2$

$f(y) \le f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|y - x\|_2^2$

$f(y) \le f(x) + \langle \nabla^r f(x), y - x\rangle + \frac{L}{2}\|y - x\|_2^2 + \langle \nabla f(x) - \nabla^r f(x), y - x\rangle$

$$\langle a, b\rangle \le \frac{\|a\|_2^2}{2L} + \frac{L\|b\|_2^2}{2} \qquad \forall L>0$$

$$\langle \nabla f(x) - \nabla^r f(x), y-x\rangle \le \frac{\|\nabla f(x) - \nabla^r f(x)\|_2^2}{2L} + \frac{L}{2}\|y-x\|_2^2$$

2) $$f(y) \le f(x) + \langle \nabla^r f(x), y-x\rangle + \frac{2L}{2}\|y-x\|_2^2 + \underbrace{\frac{\|\nabla f(x) - \nabla^r f(x)\|_2^2}{2L}}_{\delta_2}$$

$$\underset{new}{L} := 2\underset{old}{L}$$

Из (□) $\Rightarrow$ $$\mathbb{E}\,\delta_2 = \frac{1}{2L}\mathbb{E}\,\|\nabla f(x) - \nabla^r f(x)\|_2^2 \le \frac{\sigma^2}{2Lr}$$

$$\delta = \frac{\sigma^2}{2Lr}$$

Из (◊) следует, что

$$x^{k+1} = x^k - \frac{1}{2L}\,\frac{1}{r}\sum_{i=1}^{r}\nabla f(x^k, \xi^{k,i})$$

пробатч. град. спуск.

$$\bar{x}^N = \frac{1}{N}\sum_{k=1}^{N} x^k$$

$$\mathbb{E}\, f(\bar{x}^N) - f(x_*) \le \underbrace{\frac{LR^2}{N}}_{\varepsilon/2} + \underbrace{\frac{\sigma^2}{2Lr}}_{\varepsilon/2} = \varepsilon$$

$$N = \frac{2LR^2}{\varepsilon}, \qquad r = \frac{\sigma^2}{L\varepsilon} \ge 1$$

$\# \nabla f(x, \xi)$ число вызовов оракула

$$T = N\cdot r = \frac{2LR^2}{\varepsilon}\cdot\frac{\sigma^2}{L\varepsilon} = \frac{2\sigma^2 R^2}{\varepsilon^2}$$

$$N \sim \frac{LR^2}{\varepsilon}, \quad T \sim \max\left\{\frac{\sigma^2 R^2}{\varepsilon^2}, \frac{LR^2}{\varepsilon}\right\}$$

GM: $\mathbb{E} f(\bar{x}^N) - f(x_*) \simeq \frac{LR^2}{N} + \delta$

FGM: $\mathbb{E} f(\bar{x}^N) - f(x_*) \simeq \underbrace{\frac{LR^2}{N^2}}_{\varepsilon/2} + \underbrace{N \cdot \delta}_{\varepsilon/2}$

Devolder et al 2013

$$\boxed{N \sim \sqrt{\frac{LR^2}{\varepsilon}}}$$

$$N \cdot \frac{\sigma^2}{2Lr} = \frac{\varepsilon}{2}$$

$$r = \frac{\sigma^2}{L\varepsilon} N \sim \frac{1}{\varepsilon^{3/2}}$$

$$T = N \cdot r = \frac{\sigma^2}{L\varepsilon} N^2 = \frac{\sigma^2}{L\varepsilon} \cdot \frac{LR^2}{\varepsilon}$$

оптимальна → $\boxed{T \sim \frac{\sigma^2 R^2}{\varepsilon^2}}$ // $T \sim \max\left\{\frac{\sigma^2 R^2}{\varepsilon^2}, \sqrt{\frac{LR^2}{\varepsilon}}\right\}$

Оценка снизу

какое-то число → $\tilde{\varepsilon} x \to \min\limits_{x \in [-1,1]}$

знаем $|\tilde{\varepsilon}| = \varepsilon$

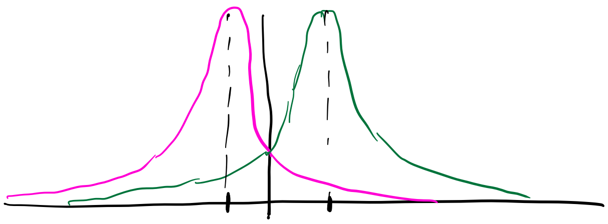
Но не знаем знак $\tilde{\varepsilon}$

$$\nabla f(x, \xi) = f'(x, \xi) = \tilde{\varepsilon} + \xi, \quad \xi \in N(0,1)$$

$\tilde{\varepsilon} + \xi^k$ — выборка

неизв. пар.

выборка из $\to N(\tilde{\varepsilon}, 1)$



$-\varepsilon$ ε

$x_* = 1$

или

$x_* = -1$

если не угадали

но $\Delta f = 2\varepsilon$

и не реш. задачу.

из Стобилена $\Rightarrow$ $n \geq \frac{1}{\varepsilon^2}$

объем выборки