

Современные методы оптимизации.

Лекция №2. Стохастическая оптимизация.

Камзолов Дмитрий

kamzolov.dmitry@phystech.edu

Московский физико-технический институт
Физтех-школа прикладной математики и информатики

25 сентября 2018

- 1 Задача стохастической оптимизации
- 2 SAA подход
- 3 SA подход
- 4 Robust SA
- 5 AC-SA

$$\min_{x \in X} \{f(x) = \mathbb{E}[F(x, \xi)]\}$$

- $X \subseteq \mathbb{R}^n$ – непустое ограниченное замкнутое выпуклое множество
- ξ – случайный вектор с вероятностным распределением P на множестве $\Xi \subseteq \mathbb{R}^d$
- $f : X \times \Xi \rightarrow \mathbb{R}$ – непрерывная и выпуклая функция на X

$$f(x) = \mathbb{E}[F(x, \xi)] = \int_{\Xi} F(x, \xi) dP(\xi)$$

Sample average approximation(SAA)

Сгенерируем набор из N случайных величин (данных)
 ξ_1, \dots, ξ_N и приблизим оригинальную задачу усредненной по
данным задачей

$$\min_{x \in X} \left\{ \bar{f}_N(x) = N^{-1} \sum_{j=1}^N F(x, \xi_j) \right\}$$

Теперь эта задача выпуклой оптимизации, которую можно
решать разными методами оптимизации

Stochastic approximation(SA)

Предположим, что существует механизм (оракул), который для данной точки (x, ξ) возвращает **стохастический субградиент** – вектор $G(x, \xi)$ такой, что функция $g(x) = \mathbb{E}[G(x, \xi)]$ определена и является субградиентом $f(x)$.

Отметим, что стох.субградиент можно выбрать, как $G(x, \xi) = \partial_x F(x, \xi)$.

SGD

$$x_{m+1} = x_m - \gamma_m G(x_m, \xi_m)$$

Шаг: $\gamma_m = \frac{\theta}{m}$.

Скорость сходимости: $\mathbb{E}[f(x_N) - f(x_*)] \leq O(1/N)$ для гладкого сильновыпуклого случая

Главный минус: неустойчив к неточному выбору шага \Rightarrow плохо работает на практике

SGD with Averaging (Б.Т.Поляк)

$$x_{m+1} = x_m - \gamma_m G(x_m, \xi_m)$$

Точка решения: $\bar{x}_N = \sum_{j=1}^N \nu_j x_j$

Шаг: $\gamma_m = \frac{\theta}{\sqrt{m}}$.

Скорость сходимости: $\mathbb{E}[f(\bar{x}_N) - f(x_*)] \leq O(1/N)$ для гладкого сильновыпуклого случая

Устойчив к неточному выбору шага без потери в скорости \Rightarrow лучше работает на практике

Пусть

$$A_j = \frac{1}{2} \|x_j - x_*\|_2^2, \quad a_j = \mathbb{E}[A_j] = \mathbb{E}[\|x_j - x_*\|_2^2]$$

Тогда

$$\begin{aligned} A_{j+1} &= \frac{1}{2} \|x_j - \gamma_j G(x_j, \xi_j) - x_*\|_2^2 \\ &\leq A_j + \frac{1}{2} \gamma_j^2 \|G(x_j, \xi_j)\|_2^2 - \gamma_j (x_j - x_*)^T G(x_j, \xi_j) \end{aligned}$$

Возьмем от обеих частей мат.ожидание и дополнительно предположим, что существует положительное число M такое, что $\mathbb{E}[\|G(x, \xi)\|_2^2] \leq M^2, \quad \forall x \in X$

Тогда

$$\begin{aligned}a_{j+1} &\leq a_j + \frac{1}{2}\gamma_j^2 M^2 - \gamma_j \mathbb{E}[(x_j - x_*)^T G(x_j, \xi_j)] \\&= a_j + \frac{1}{2}\gamma_j^2 M^2 - \gamma_j \mathbb{E} \left[\mathbb{E} \left\{ (x_j - x_*)^T G(x_j, \xi_j) | \xi_{j-1} \right\} \right] \\&= a_j + \frac{1}{2}\gamma_j^2 M^2 - \gamma_j \mathbb{E} \left[(x_j - x_*)^T \mathbb{E} \{ G(x_j, \xi_j) | \xi_{j-1} \} \right] \\&= a_j + \frac{1}{2}\gamma_j^2 M^2 - \gamma_j \mathbb{E} \left[(x_j - x_*)^T g(x_j) \right] \\&\leq a_j + \frac{1}{2}\gamma_j^2 M^2 - \gamma_j \mathbb{E} [f(x_j) - f(x_*)]\end{aligned}$$

Просуммировав это неравенство получаем

$$\begin{aligned} \sum_{j=1}^N \gamma_j \mathbb{E} [f(x_j) - f(x_*)] &\leq \sum_{j=1}^N (a_j - a_{j+1}) + \frac{1}{2} \sum_{j=1}^N \gamma_j^2 M^2 \\ &\leq a_1 + \frac{1}{2} M^2 \sum_{j=1}^N \gamma_j^2 \end{aligned}$$

Пусть $\nu_j = \frac{\gamma_j}{\sum_{i=1}^N \gamma_i}$, $\bar{x}_N = \sum_{j=1}^N \nu_j x_j$, $D_X = \max_{x \in X} \|x - x_1\|_2$ тогда

$$\mathbb{E} [f(\bar{x}_N) - f(x_*)] \leq \mathbb{E} \left[\sum_{j=1}^N \nu_j f(x_j) - f(x_*) \right] \leq \frac{D_X + \frac{1}{2} M^2 \sum_{j=1}^N \gamma_j^2}{\sum_{i=1}^N \gamma_i}$$

Для получения постоянного шага минимизируем правую часть неравенства по γ получим:

$$\gamma_j = \frac{D_X}{M\sqrt{N}}$$

И соответственно скорость сходимости:

$$\mathbb{E}[f(\bar{x}_N) - f(x_*)] \leq \frac{D_X M}{\sqrt{N}}$$

SA versus SAA on the stochastic max-flow problem.

- (m, n)		F1 (50,500)		F2 (100, 1,000)		F3 (100, 2,000)		F4 (250, 5,000)	
ALG.	N	Obj	CPU	Obj	CPU	Obj	CPU	Obj	CPU
N-SA	100	0.1140	0	0.0637	0	0.1296	1	0.1278	3
	1,000	0.1254	1	0.0686	3	0.1305	6	0.1329	15
	2,000	0.1249	3	0.0697	6	0.1318	11	0.1338	29
	4,000	0.1246	5	0.0698	11	0.1331	21	0.1334	56
E-SA	100	0.0840	0	0.0618	1	0.1277	2	0.1153	7
	1,000	0.1253	3	0.0670	6	0.1281	16	0.1312	39
	2,000	0.1246	5	0.0695	13	0.1287	28	0.1312	72
	4,000	0.1247	9	0.0696	24	0.1303	53	0.1310	127
SAA	100	0.1212	5	0.0653	12	0.1310	20	0.1253	60
	1,000	0.1223	35	0.0694	84	0.1294	157	0.1291	466
	2,000	0.1223	70	0.0693	170	0.1304	311	0.1284	986
	4,000	0.1221	140	0.0693	323	0.1301	636	0.1293	1,885

SA versus SAA on the stochastic utility problem.

-		L1: $n = 500$		L2: $n = 1,000$		L3: $n = 2,000$		L4: $n = 5,000$	
ALG.	N	Obj	CPU	Obj	CPU	Obj	CPU	Obj	CPU
N-SA	100	-7.7599	0	-5.8340	0	-7.1419	1	-5.4688	3
	1,000	-7.8781	2	-5.9152	2	-7.2312	6	-5.5716	13
	2,000	-7.8987	2	-5.9243	5	-7.2513	10	-5.5847	25
	4,000	-7.9075	5	-5.9365	12	-7.2595	20	-5.5935	49
E-SA	100	-7.6895	0	-5.7988	1	-7.0165	1	-4.9364	4
	1,000	-7.8559	2	-5.8919	4	-7.2029	7	-5.3895	20
	2,000	-7.8737	3	-5.9067	7	-7.2306	15	-5.4870	39
	4,000	-7.8948	7	-5.9193	13	-7.2441	29	-5.5354	77
SAA	100	-7.6571	7	-5.6346	8	-6.9748	19	-5.3360	44
	1,000	-7.8821	31	-5.9221	68	-7.2393	134	-5.5656	337
	2,000	-7.9100	72	-5.9313	128	-7.2583	261	-5.5878	656
	4,000	-7.9087	113	-5.9384	253	-7.2664	515	-5.5967	1,283

Accelerated SA(AC-SA)

- $x_k^{md} = \alpha_k x_k + (1 - \alpha_k) x_k^{ag}$
- Вычисляем $G(x_k^{md}, \xi_k)$
- $x_{k+1} = x_k - \gamma_k G(x_k^{md}, \xi_k)$
- $x_{t+1}^{ag} = \alpha_k x_{k+1} + (1 - \alpha_k) x_k^{ag}$

И соответственно скорость сходимости:

$$\mathbb{E}[f(\bar{x}_N) - f(x_*)] \leq O(1) \left\{ \frac{L}{N^2} + \frac{M}{\sqrt{N}} \right\}$$

Спасибо за внимание!