Минимаксная оценка сложности распределенной стохастической выпуклой оптимизации в случае перепараметризации

Вероника Озернова
Научный руководитель: Гасников А.В.

Московский физико-технический институт

26 марта 2024

Содержание

Постановка задачи

- Рассматриваем классическую гладкую выпуклую задачу:

$$\min_{w \in \mathbb{R}^d} F(w) := \mathbb{E}_{z \sim \mathcal{D}}[\ell(w; z)]$$

где $F-$ неотрицательная выпуклая функция, $\|w^*\| \leq B$ и F - H-гладкая, т.е.

$$F(w)+\langle \nabla F(w), y-w \rangle \leq F(y) \leq F(w)+\langle \nabla F(w), y-w \rangle + \frac{H}{2}\|y-w\|^2$$

$$g(w_t) = \frac{1}{b}\sum_{i=1}^{b} \nabla \ell(w_t; z^i) \text{ for } i.i.d. z^1, \ldots, z^b \sim \mathcal{D}$$

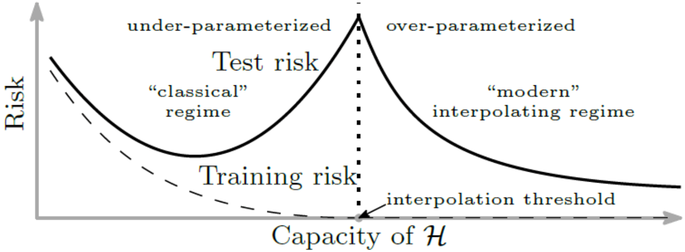
$$\mathbb{E}_z[g(x;z)] = \nabla F(x) \quad \text{and} \quad \mathbb{E}_z\|g(x;z) - \nabla F(x)\|^2 \leq \sigma^2$$

Постановка задачи

- Мы рассматриваем федеративные алгоритмы с прерывистой коммуникацией, которые пытаются оптимизировать F, используя M параллельных машин, каждой из которых разрешается K запросов к g в каждом из R раундов коммуникации. Наконец, используем IID-данные, где каждая из машин имеет доступ к стохастическим градиентам из одного и того же распределения, в отличие от более сложной "гетерогенной" среды.

Перепараметризация

- Задача рассматривается в контексте режима перепараметризации. С этим связано то, что на тренировочных данных мы можем получить почти нулевое значени функции потерь. Основой этого режима является bias-variance trade-off



Мотивация

- Федеративные алгоритмы
- DeepLearning

Существующие результаты

- Доказаны оценки сложности и сходимости и приведены соответсвующие алгоритмы для классической постановки задачи федеративного обучения в [3]. Так же соответсвенно доказаны оценки и проработана тематика перепараметризации для стохастических методов в [1] и [2].
- Перепараметризация: $E_k\|\nabla \ell(w^*; z) - \nabla L(w^*)\|_2^2 \leq 2HL^*$

Существующие результаты

Theorem 1 *For any $H, B, \sigma, K, R > 0$ and $M \geq 2$, and any intermittent communication algorithm, there exists a convex, H-smooth objective which has a minimizer with norm at most B in any dimension*

$$d \geq 2KR + \left(10^9\left(1 + KR + \left(\frac{HB}{\sigma}\right)^{3/2} M(KR)^{5/4}\right) + \frac{6144H^2B^2MKR}{\sigma^2}\right)\log(64MK^2R^2)$$

and a stochastic gradient oracle, g, with $\mathbb{E}_z\|g(x;z) - \nabla F(x)\|^2 \leq \sigma^2$ such that the algorithm's output will have error at least

$$\mathbb{E}F(\hat{x}) - F^* \geq c \cdot \left(\frac{HB^2}{K^2R^2} + \min\left\{\frac{\sigma B}{\sqrt{MKR}}, HB^2\right\} + \min\left\{\frac{HB^2}{R^2(1+\log^2 M)}, \frac{\sigma B}{\sqrt{KR}}\right\}\right)$$

for a numerical constant c.

Верхняя оценка: $E[F(x_{KR}) - F_*] \leq c \cdot \left(\frac{HB^2}{R^2} + \frac{\sigma B}{\sqrt{K}}\right)$

План работы

- Получить оптимальные оценки для федеративных алгоритмов в случае перепараметризации, объединив результаты предыдущих исследований
- Провести эксперименты и замерить соответсвующие показатели

Список литературы

- Sasila Ilandarideva, Anatoli Juditsky, Guanghui Lan2, Tianjiao Li, Accelerated stochastic approximation with state-dependent noise, 2023.
- Blake Woodworth, Nathan Srebro, An Even More Optimal Stochastic Optimization Algorithm: Minibatching and Interpolation Learning, 2021.
- Blake Woodworth, Brian Bullins, Ohad Shamir, Nathan Srebro, The Min-Max Complexity of Distributed Stochastic Convex Optimization with Intermittent Communication, 2021.