

Минимаксная оценка сложности распределенной стохастической выпуклой оптимизации в случае перепараметризации

В.С. Озернова, А.В. Гасников

Московский физико-технический институт (национальный исследовательский университет)

Масштабные размеры многих современных моделей машинного обучения и наборов данных порождают сложные цели обучения с высокой размерностью, оптимизация которых может быть очень вычислительно затратной. Для уменьшения вычислительной стоимости важно разрабатывать алгоритмы оптимизации, способные использовать параллелизм для сокращения времени обучения. Стохастические методы первого порядка, использующие стохастические оценки градиента целевой функции обучения, являются наиболее распространенным подходом. Более того, такие методы могут быть масштабированы на несколько устройств, что позволяет улучшить производительность.

Распределенные вычисления формируют такой подход в ML как федеративное обучение. Оно характеризуется ограниченностью коммуникаций клиентов, на которых обучаются локальные модели. Это позволяет разделить данные на несколько машин, которые в свою очередь уже проводят вычисления на меньших объемах, чем если бы все велось на одной машине. Однако такие алгоритмы также влекут ряд ограничений на задачи, которые могут быть реализованы. В том числе эти задачи подразумевают классификацию по типу распределения данных на клиентах - IID и non-IID. IID - базовое предположение о независимости и одинаковом равномерном распределении обучающих примеров между клиентами. Если же это предположение не выполняется, мы говорим о независимой и неодинаковой ситуации(non-IID).

В контексте задач машинного обучения также существует такое понятие как обучение в режиме перепараметризации. Этот режим говорит о том, что модель может иметь число параметров, сильно превышающее размерность данных, что позволяет свести функцию потерь на тренировочной выборке до близкого к нулю или же нулевого значения. Такой концепт характерен для задач DeepLearning.

В данной работе рассматривается классическая постановка задачи распределенной стохастической выпуклой оптимизации с целевой функцией потерь с федеративными алгоритмами с прерывистой коммуникацией. В качестве основы обучения используются IID данные. Данная задача рассматривается в контексте перепараметризации, поэтому можем также ввести ограничения на оптимальное значение целевой функции. Целью исследования является улучшение существующих верхних и нижних оценок для сложности распределенной оптимизации [1] с учетом условий, накладываемых режимом перепараметризации [2, 3].

Были получены улучшенные оценки и сформулированы теоремы и промежуточные леммы, отражающие значения ошибок, которых можно добиться с заданными условиями на задачу.

Литература

- [1] *Blake Woodworth, Brian Bullins, Ohad Shamir, Nathan Srebro*. The Min-Max Complexity of Distributed Stochastic Convex Optimization with Intermittent Communication. *arXiv:2102.01583*. 2021
- [2] *Sasila Ilandarideva, Anatoli Juditsky, Guanghui Lan, Tianjiao Li*. Accelerated stochastic approximation with state-dependent noise. *arXiv:2307.01497*. 2023
- [3] *Blake Woodworth, Nathan Srebro*. An Even More Optimal Stochastic Optimization Algorithm: Minibatching and Interpolation Learning. *arXiv:2106.02720v2*. 2021