

Определение параметров нейронной сети, подлежащих оптимизации

К.Ф.Муравьев, О.Ю.Бахтеев, В.В.Стрижов

kirill.mouraviev@yandex.ru; bakhteev@phystech.edu; strijov@phystech.edu

Московский физико-технический институт

В работе рассматривается задача ускорения оптимизации параметров нейронной сети. Предлагается рассматривать градиентную оптимизацию параметров как стохастический процесс, позволяющий получить оценки на апостериорное распределение параметров и отклонение оптимизируемых параметров от точки экстремума. Экспериментальный анализ качества алгоритма проводится на выборке рукописных цифр MNIST.

Ключевые слова: нейронные сети, методы оптимизации параметров, стохастический градиентный спуск, стохастическая динамика, precondition-матрица.

Введение

Оптимизация глубоких нейронных сетей является задачей высокой вычислительной сложности и требует больших временных затрат и вычислительных мощностей [6]. При этом оптимизация сходится по большинству параметров сети уже после небольшого числа итераций. Своевременное определение сходимости параметров позволит существенно снизить вычислительные затраты на обучение нейросетей.

Основные современные методы обучения нейронных сетей [3] работают довольно быстро и достигают хорошего качества по сравнению с обычным стохастическим градиентным спуском, но все они проводят одинаковое число шагов оптимизации для всех параметров. В работе [5] описывается байесовский подход к обучению нейронных сетей с использованием стохастической динамики Ланжевена (SGLD). В работе [4] описывается использование метода SGLD с precondition-матрицей вместо константного шага, которое ускоряет время оптимизации глубоких нейросетей до четырех раз.

В данной работе предлагается, анализируя precondition-матрицу градиентного спуска, определять параметры, для которых метод оптимизации сошелся, и удалять их из множества параметров, подлежащих дальнейшей оптимизации.

Эксперимент проводится на датасете MNIST для нескольких типов полносвязных и сверточных нейронных сетей. Проводится сравнение метода с другими методами оптимизации параметров.

Список литературы

- [1] | Stephan Mandt, Matthew D. Hoffman, David M. Blei: Stochastic Gradient Descent as Approximate Bayesian Inference. 2017

- [2] Alex Graves: Practical Variational Inference for Neural Networks
- [3] Diederik P. Kingma, Jimmy Lei Ba: Adam: a Method for Stochastic Optimization. 2015
- [4] Chunyan Li, Changyou Chen, David Carlson, Lawrence Carin: Preconditioned Stochastic Gradient Langevin Dynamics for Deep Neural Networks. 2015
- [5] Max Welling, Yee Whye Teh: Bayesian Learning via Stochastic Gradient Langevin Dynamics
- [6] Barret Zoph, Quoc V.Le: Neural Architecture Search with Reinforcement Learning. 2017