

МЕТОДЫ ОБУЧЕНИЯ СТОХАСТИЧЕСКИХ БИНАРНЫХ НЕЙРОННЫХ СЕТЕЙ

Януш Виктор Янович

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: yanushvictor@gmail.com

Научный руководитель — Кропотов Дмитрий Александрович

Нейронные сети с бинарными весами и активациями являются крайне эффективными с вычислительной точки зрения моделями. В [1] говорят об ускорении операции свертки в 58 раз по сравнению с обычными вычислениями с плавающей запятой, а также об уменьшении размера модели в 32 раза благодаря переходу от чисел одинарной точности к однобитному представлению.

Было показано, что сети с квантизацией весов и активаций, в частности, с бинаризацией способны достигать того же качества работы, что и базовые методы, работающие с действительными числами [4]. Тем не менее, при обучении бинарных нейронных сетей зачастую используются эвристики, не обоснованные математически, например, такие как Straight-Through Estimator [5].

Основная проблема, возникающая при обучении бинарных сетей это то, что состояния нейронов изменяются дискретно при изменении параметров, что приводит к необходимости использования методов комбинаторики или сглаживания для поиска направлений оптимизации. Стохастические бинарные нейронные сети [2] внедряют шум перед дискретными функциями активации. Это приводит к тому, что выход нейросети становится гладкой функцией после матожидания по всем шумам, а следовательно производная функции потерь корректно определена. Более того, внедренный во все слои шум превращает сеть в глубокую модель с латентными переменными с очень гибким прогнозным распределением. Такие модели могут оценивать неопределенность в данных. Работая с весами как со случайными переменными можно также оценить неопределенность модели используя байесовские методы (например, см. [3]).

В работе рассматривается метод обучения бинарной нейронной сети на основе последовательного сэмпинга. Вычисление градиента функции потерь по выходу определенного слоя происходит следующим образом: аналитически подсчитывается и дифференцируется матожидание по сэмплам отличающимся не более чем в одном бите. Такой подсчет происходит в каждом последующем слое и на выхо-

де получается аппроксимация градиента матожидания функции потерь. Полученный метод похож по форме на обычный метод обратного распространения ошибки и может быть реализован через него, однако позволяет работать с дискретными переменными и имеет ту же вычислительную сложность.

Литература

1. Rastegari, Mohammad, et al. "Xnor-net: Imagenet classification using binary convolutional neural networks." European conference on computer vision. Springer, Cham, 2016.
2. Raiko, Tapani, et al. "Techniques for learning binary stochastic feedforward neural networks." arXiv preprint arXiv:1406.2989 (2014).
3. Neal, R. M. Bayesian learning for neural networks, volume 118. Springer Science Business Media, 2012.
4. Bethge, Joseph, et al. "Training competitive binary neural networks from scratch." arXiv preprint arXiv:1812.01965 (2018).
5. Bengio, Yoshua, Nicholas Léonard, and Aaron Courville. "Estimating or propagating gradients through stochastic neurons for conditional computation." arXiv preprint arXiv:1308.3432 (2013).