

“On the Periodic Behavior of Neural Network Training with Batch Normalization and Weight Decay”(Ekaterina Lobacheva, Maxim Kodryan, Nadezhda Chirkova, Andrey Malinin, Dmitry Vetrov).

Автор обзора-рецензии: Стриженок Сергей, БПМИ-191.

В статье описывается интересный факт, который получается при одновременном использовании weight decay и batch normalization: при обучении наблюдается дестабилизация, которая не приводит к расхождению, но инициирует собой новый период обучения. Авторы в статье показывают как и эксперименты, доказывающие это, так и теоретическое обоснование.

Процесс работы над статьей происходил в течение 2021 года, первая версия была опубликована летом 29 июня 2021 года, крайняя на текущий момент 15 января 2022 года. Статья была опубликована на 35-й конференции NeurIPS(Neural Information Processing Systems) в 2021 году. Это одна из крупнейших в мире конференций по машинному обучению и нейронным сетям, которая проводится с 1989 года. Авторы статьи – исследователи в центре глубинного обучения и байесовских методов НИУ ВШЭ. Часть из авторов является нашими преподавателями по разным предметам, связанным с машинным обучением. На похожую тему у авторов обнаружено не было, но все предыдущие статьи авторов были написаны на тему глубинного обучения.

На мой взгляд повлиявших статей на рецензируемую нет, но есть как минимум 2 важных статьи, на которые ссылались авторы: “Batch normalization: Accelerating deep network training by reducing internal covariate shift” – первая статья про batch normalization и “Understanding batch normalization” и множество аналогичных, в которых описывается как “работает” и как помогает batch normalization при обучении, но не до конца детально.

По google scholar найдено 10 цитирований и ссылок, прямых продолжений не найдено. Также статья упомянута в блоге на [towardsdatascience](https://towardsdatascience.com/).

Сильная сторона работы состоит в том, что с помощью несложной интересной идеи и простым математическим выкладкам, которые поймет неподготовленный слушатель, авторы доносят сложный и неожиданный результат.

Слабые стороны – из статьи не до конца понятно, как можно избежать такого поведения при обучении нейронной сети. Для всей теории, представленной в статье, нужно соблюдения многих условий, которые, как в том числе говорят авторы, часто в индустрии не выполняются.

В качестве улучшения я бы предложил авторам подробно рассказать, как можно избежать поведения при обучении нейронной сети и попытаться обобщить теорию на более широкие срезы методов и данных, которые чаще встречаются в реальном мире.