

## Stochastic Training is Not Necessary for Generalization

Автор исследования: Рахматуллин Рамазан

Опубликована на arXiv 29 сентября 2021 года, отправлена на ICLR 2022. Вторая версия загружена на OpenReview 22 ноября после комментариев рецензентов. Изначально рецензии были в большинстве негативные, после комментариев авторов и обновления статьи мнение поменялось в лучшую сторону, так что теперь средний рейтинг у статьи немного выше чем acceptable [openreview](#).

Главный автор статьи переехал из University of Siegen (Зиген, Германия) в University of Maryland (Вашингтон, США). Его соавторами стали руководитель из Siegen, руководитель и коллеги из Maryland:

(главный автор) [Jonas Geiping](#) University of Siegen(phd), postdoc at UMD, в списке публикаций статья только у него

(коллега) [Micah Goldblum](#) University of Maryland(bs/phd), postdoc at NYU

(другой коллега) [Phillip E. Pope](#) University of Maryland(phd)

(руководитель из Siegen) [Michael Moeller](#) University of Siegen(prof)

(главный руководитель из Maryland) [Tom Goldstein](#) University of Maryland(prof)

В целом группа из Maryland (и автор в частности) занимаются вопросами безопасности в машинном обучении (например статья о том, как в распределенном обучении зная градиент можно вычислить элементы датасета, который ты как одна из машин не видела).

Предыдущей статьей можно считать [эту](#). Авторы изучают разные общеизвестные предположения о нейронных сетях. Полагаю, что автор заинтересовался свойствами поверхности графиков, можно упомянуть [статью](#), [SAM](#)(улучшают максимум вокруг точки) и прочие такие же.

7 цитирований, в основном по поводу примененных регуляризаций. Есть интересная [статья](#), доказывающая неэффективность GD если использовать только градиент множества. Они упоминают, что в нашей статье используется статистика по батчам. Наша статья не имеет каких-то сильных выводов(пока что) и проверена только на одном датасете, поэтому ссылок особо нет.

В рецензиях предлагают много разных непроведенных экспериментов, увеличить число итераций, применить регуляризации к SGD, посмотреть другой датасет. Однако скорость обучения невелика, поэтому авторы часть из этого решили не делать, другую часть они сделали и добавили в обновленную версию.