

1 Эксперименты на infinite cluttered MNIST

1.1 Сравнение для Adam

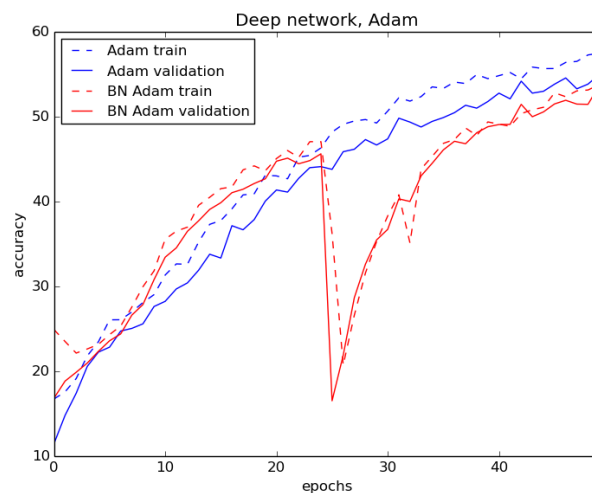
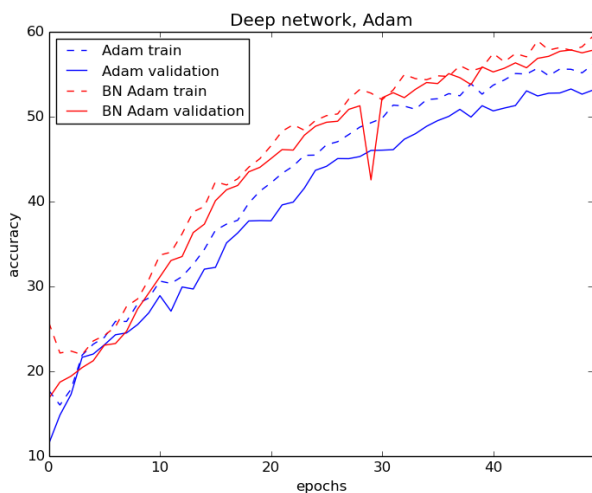


Рис. 1: Глубокая сеть ($10 \text{ HL} \times 100$)

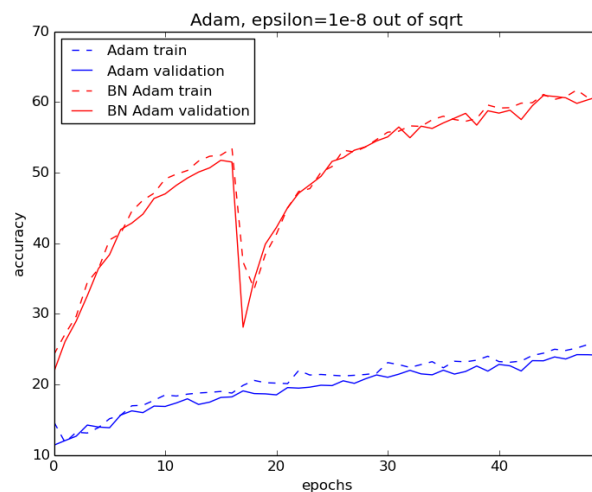
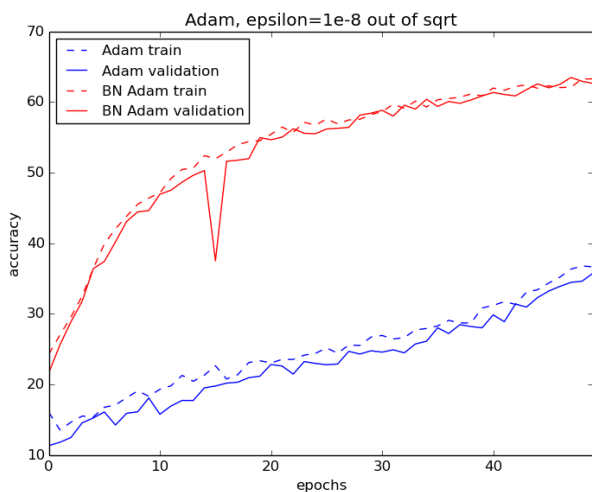


Рис. 2: $\epsilon = 10^{-8}$ вне корня ($3 \text{ HL} \times 100$)

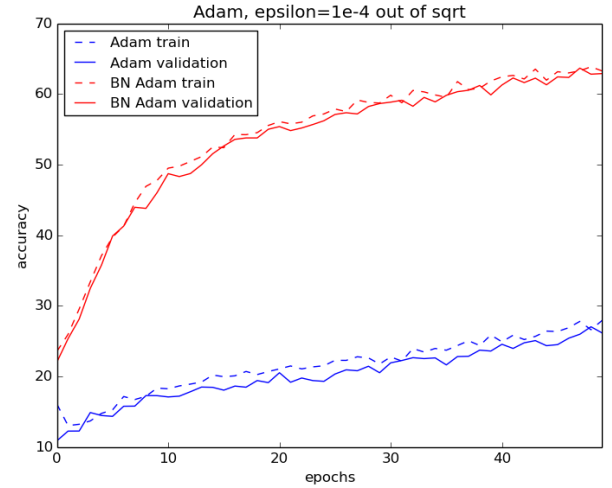
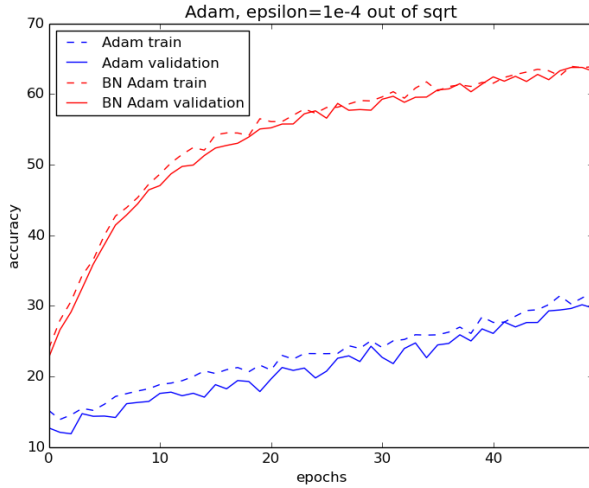


Рис. 3: $\epsilon = 10^{-4}$ вне корня ($3 \text{ HL} \times 100$)

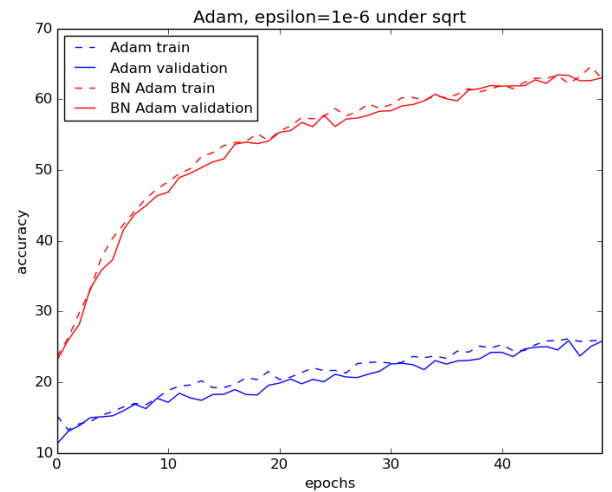
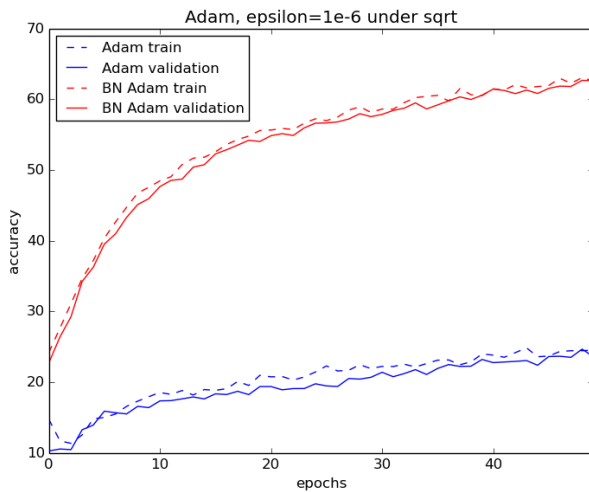


Рис. 4: $\epsilon = 10^{-6}$ под корнем, как сделано в других методах ($3 \text{ HL} \times 100$)

Делаем выводы, что

1. батч-нормализация может сильно прыгать при дефолтных параметрах метода;
2. батч-нормализация перестает прыгать, если увеличить эпсилон;
3. на неглубокой сети (3 скрытых слоя) Адам не успевает обучиться из-за сложной выборки и малого количества параметров;
4. на сети с 10 скрытыми слоями батч-нормализация слабо улучшает Адам.

1.2 Остальные методы

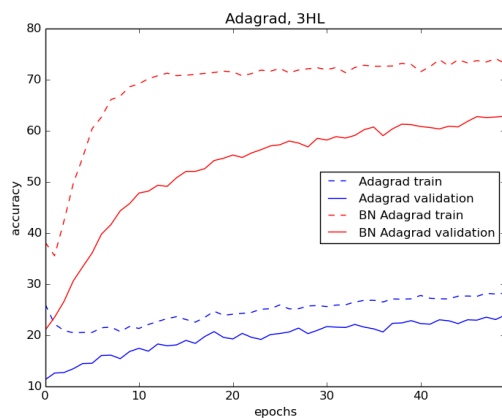


Рис. 5: Adagrad, 3 HL $\times 100$

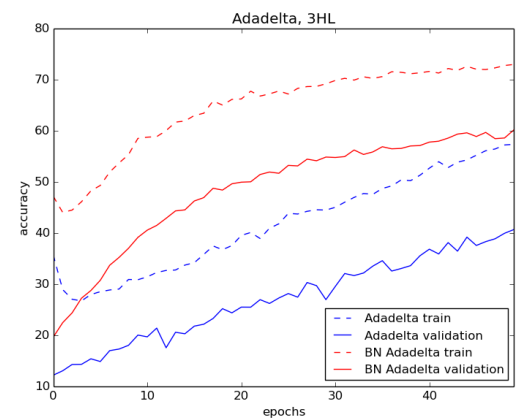


Рис. 6: Adadelata, 3 HL $\times 100$

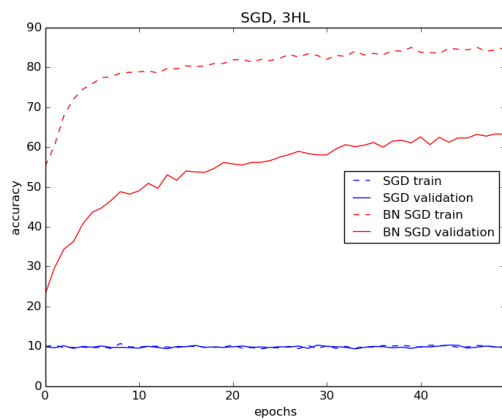


Рис. 7: SGD, 3 HL $\times 100$

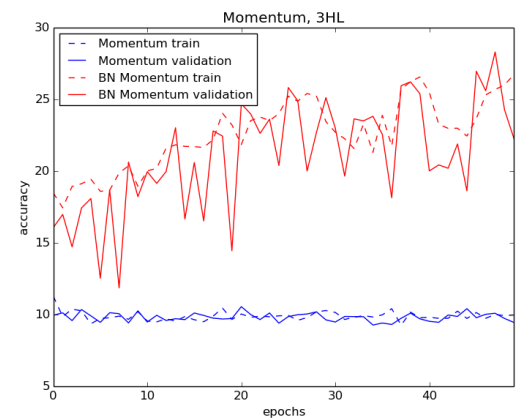


Рис. 8: Momentum, 3 HL $\times 100$

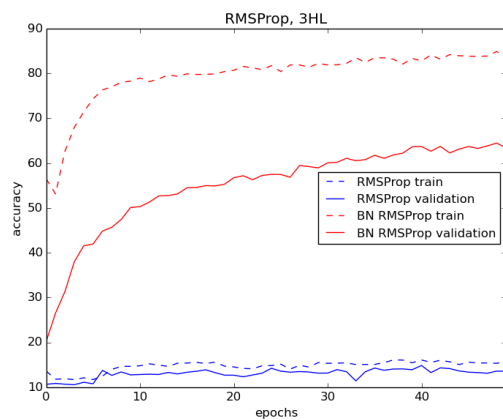


Рис. 9: RMSProp, 3 HL $\times 100$

Выводы:

1. рейтинг здесь особо не подбирался, поэтому нужно дополнительное исследование;
2. моментум сильно скачет, хотя рейтинг у них с SGD выставлен одинаковый;
3. моментум – единственный метод, для которого батч-нормализация на тренировочной и валидационной выборках почти не различается;

2 Эксперименты на MNIST

2.1 Сравнение для Adam

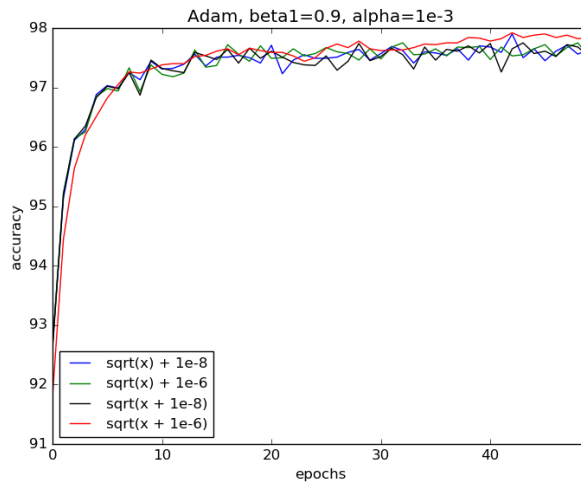


Рис. 10: Различные эpsilon-стратегии

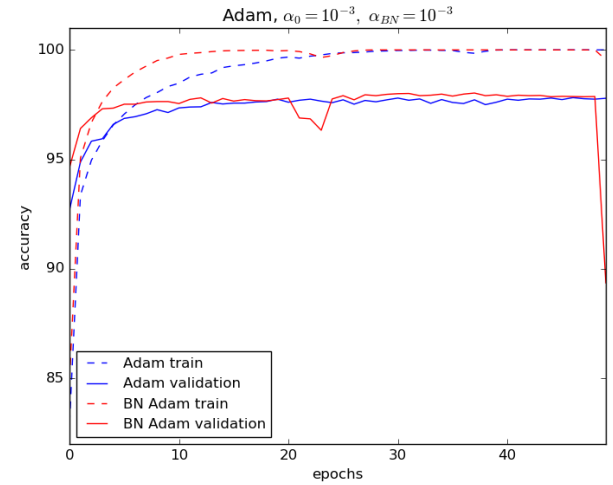


Рис. 11: Адам, 3 HL \times 100

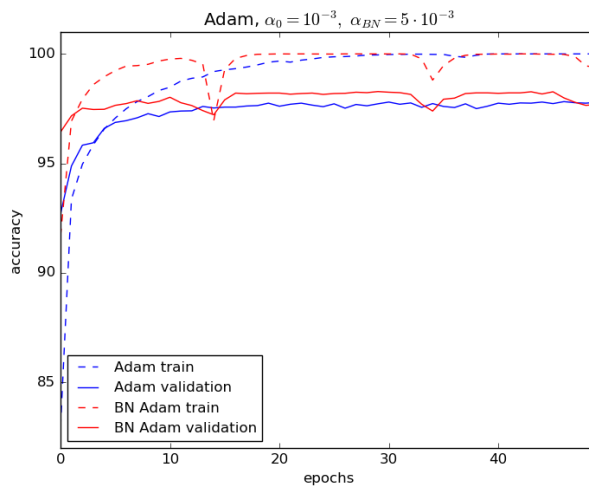


Рис. 12: Адам, 3 HL \times 100

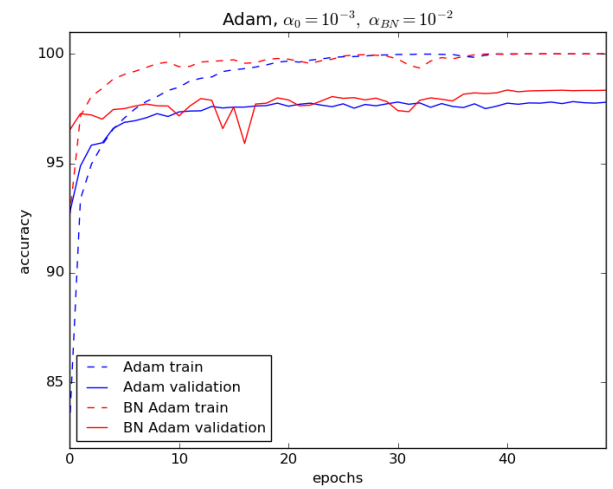


Рис. 13: Адам, 3 HL \times 100

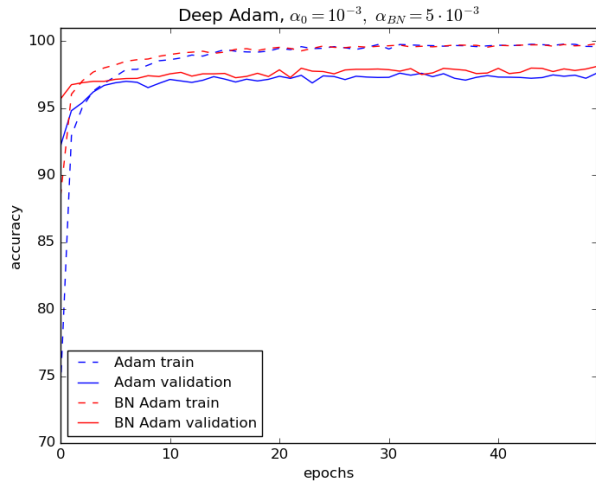


Рис. 14: Адам на глубокой сети, 10 HL $\times 100$

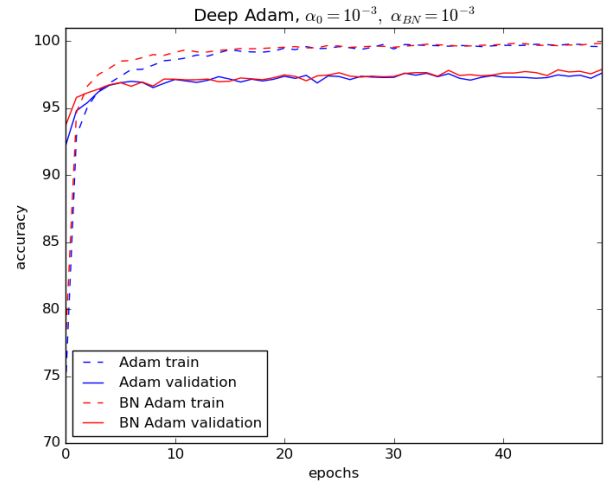


Рис. 15: Адам на глубокой сети, 10 HL $\times 100$

2.2 Остальные методы

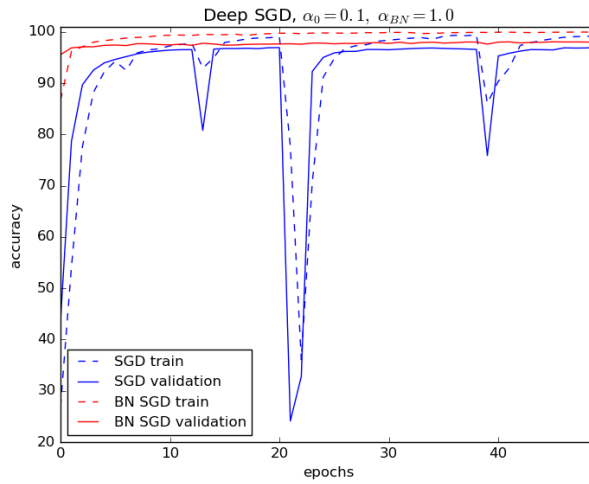


Рис. 16: SGD на глубокой сети, 10 HL $\times 100$

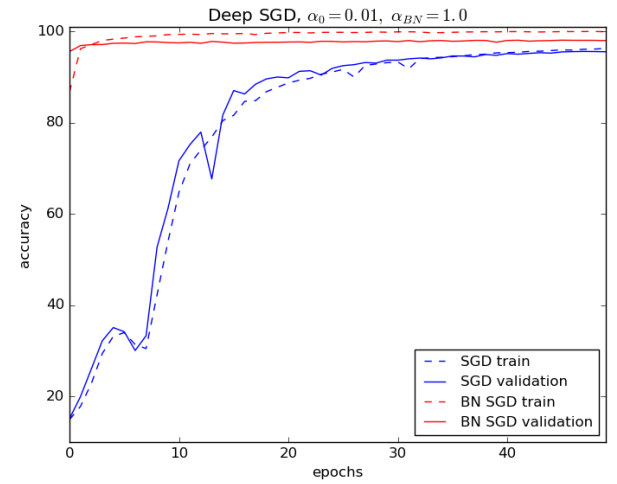


Рис. 17: SGD на глубокой сети, 10 HL $\times 100$

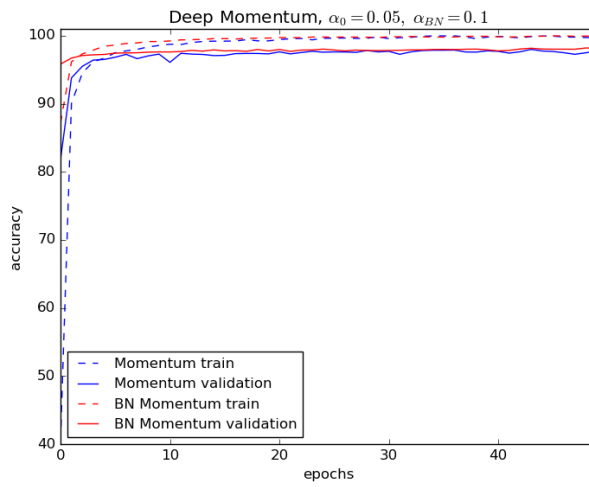


Рис. 18: Моментум на глубокой сети, 10 HL $\times 100$

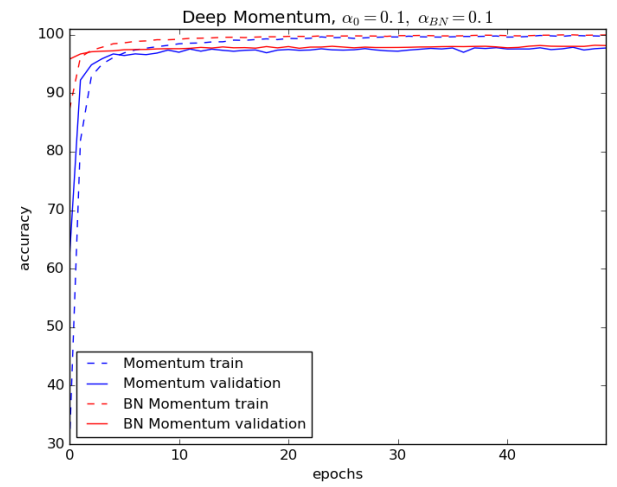


Рис. 19: Моментум на глубокой сети, 10 HL $\times 100$

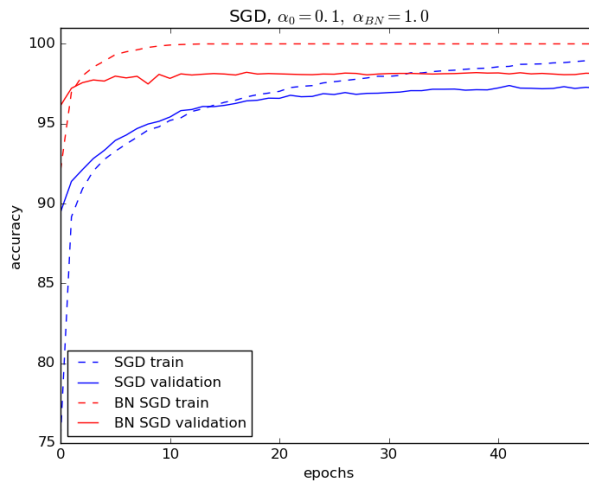


Рис. 20: SGD на неглубокой сети, 3 HL $\times 100$

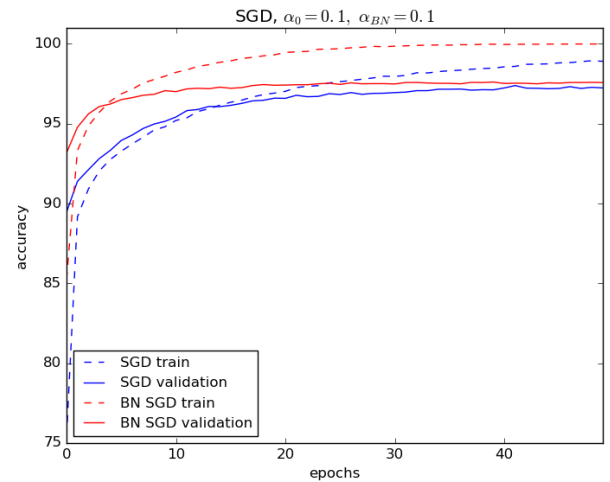


Рис. 21: SGD на неглубокой сети, 3 HL $\times 100$

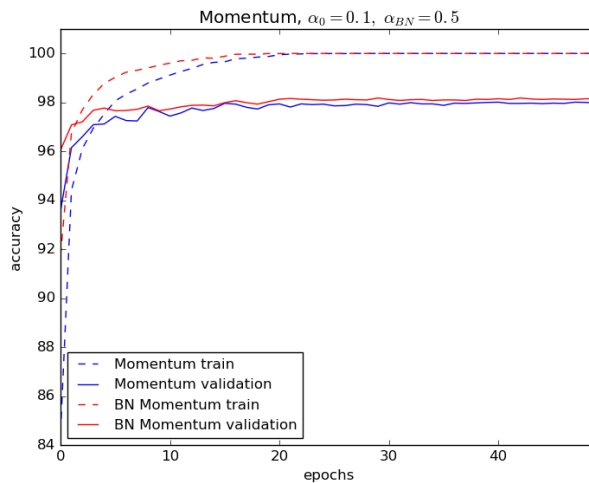


Рис. 22: Моментум на неглубокой сети, 3 HL $\times 100$

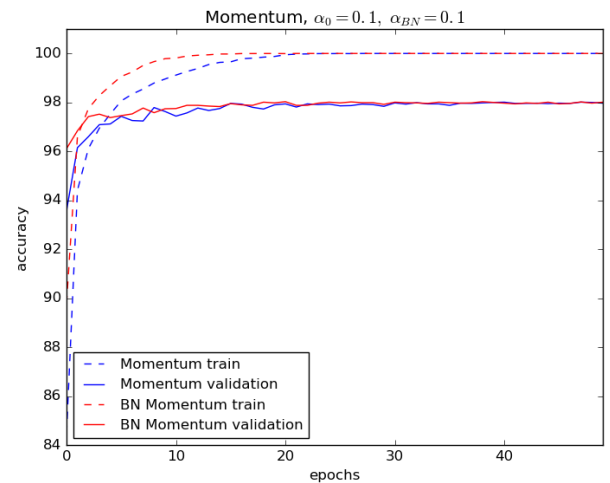


Рис. 23: Моментум на неглубокой сети, 3 HL $\times 100$

Выводы:

1. батч-нормализация для Адам ведет себя более плавно при увеличении эпсилон;
2. батч-нормализация слабо улучшает Адам для обеих архитектур сети;
3. на глубокой сети SGD плохо обучается либо из-за плохо выбранного рейта, либо так и должно быть;
4. батч-нормализация так же слабо улучшает Моментум для обеих архитектур;