

# 1 Постановка экспериментов

1. Хотим подтвердить экспериментально, что скачки Адама с батч-нормализацией уменьшаются при увеличении эпсилон, а также сравнить сглаженную версию Адама с батч-нормализацией со стандартной версией Адама. Для этого мы запустим следующие эксперименты на датасете MNIST на трехслойной полносвязной сети:
  - BN Adam со стандартным  $\epsilon = 10^{-8}$  и сохраним параметры  $v_t$  (3 раза)
  - BN Adam с увеличенным  $\epsilon = 10^{-4}$  и сохраним параметры  $v_t$  (3 раза)
  - Adam со стандартным  $\epsilon = 10^{-8}$
2. Хотим провести эксперименты с дисперсией градиента и может быть подтвердить, что BN уменьшает дисперсию. Для этого мы запустим  $K$  раз SGD и BN SGD на датасете MNIST для 3х- и 10ти-слойной полносвязных сетей из одного начального приближения. На каждой итерации будем записывать угол (или косинус угла) отклонения текущего стохастического шага от полного градиента в текущей точке. Затем усредним все отклонения и проверим, доминирует ли "дисперсия"SGD над "дисперсией"BN SGD.
3. Хотим убедиться, что на трехслойной сети на данных cluttered MNIST метод Адам не успевает обучиться за 50 эпох, для чего мы запустим уже существующий эксперимент на большее количество эпох.
4. Хотим проверить гипотезы об *эпсилон* и *улучшении методов при добавлении BN* для другой архитектуры сети, а именно сверточной (2 сверточных слоя с макс-пулингом, а затем 1 скрытый полносвязный слой). Для этого мы запустим следующие эксперименты:
  - Все методы с их BN версией на датасете MNIST
  - Все методы с их BN версией на датасете CIFAR-10
  - Метод BN Adam с  $\epsilon_2 = 10^{-4}$  на обоих датасетах