

Проект по методам оптимизации

On the Convergence of Adam and Beyond

Колесников Иван, Клюкин Ярослав Б05-004

1 Абстракт

Проект основан на статье "On the Convergence of Adam and Beyond" вышедшей в 2018 году. Авторы публикации обратили внимание на проблемы сходимости популярного метода градиентной оптимизации Adam и предложили модифицированный алгоритм AMSGrad, решающий эту проблему. В статье так же утверждается, что часто AMSGrad сходится лучше, чем Adam. В проекте мы реализовали оба алгоритма и проверили, что на задаче выпуклой оптимизации, представленной в статье, AMSGrad сходится, а Adam – нет. Также сравнили качество нейронных сетей, обученных при помощи данных алгоритмов на задачах MNIST и CIFAR.

2 Цели проекта

- Реализовать метод градиентной оптимизации Adam и его модификацию AMSGrad
- Проверить сходимость обоих методов на синтетическом примере выпуклой оптимизации из статьи
- Обучить нейронные сети на задаче MNIST. Для обучения первой сети использовать в качестве оптимайзера Adam, а для второй – AMSGrad. Сравнить графики обучения сетей и сделать выводы об относительной скорости сходимости алгоритмов.
- Выполнить предыдущий пункт для задачи CIFAR.

3 Итоги проекта

- Реализовали Adam и AMSGrad, проверили корректность работы алгоритмов на простой задаче выпуклой оптимизации.
- Результаты совпали с предоставленными в статье: на синтетической задаче онлайн-оптимизации (или стохастическом аналоге) Adam не сходится, AMSGrad сходится. Получены графики
- Для задачи MNIST функция потерь и на обучающей выборке, и на тестовой выборке немного быстрее убывает при использовании AMSGrad.
- На задаче CIFAR AMSGrad быстрее оптимизировал функцию потерь (из графиков видно, что она быстрее вышла на плато, а потом сеть начала переобучаться).

4 Промежуточные результаты

- Были реализованы классы Adam и AMSGrad с использованием библиотеки numpy. Далее был произведен простой тест методов для минимизации одномерной квадратичной функции.
- Был воспроизведен синтетический пример, на котором Adam не сходил к оптимуму, а AMSGrad справлялся с задачей. Был реализован онлайн-вариант задачи, очень долго пытались сделать стохастический но так и не вышло, точно не знаю почему, видимо что-то не так делали.
- Для решения задачи MNIST использовалась библиотека PyTorch, в которой была написана полносвязная нейронная сеть с одним скрытым слоем и сигмоидой в качестве функции активации. Стоит отметить, что задача классификации на 10 классов решается с помощью данной сети с хорошей точностью – порядка 96%, без переобучения. Из графиков видно, что до 50-ой итерации функции потерь на обучающей выборке убывают одинаково, но после функция потерь, оптимизируемая AMSGrad, убывает чуть быстрее.

- Задача CIFAR решалась похожим образом, только с использованием архитектуры свёрточной нейронной сети. Из графиков видно, что до 10-ой итерации функция потерь на обучающей выборке убывает сильнее при использовании AMSGrad, а после лучше себя показывает Adam. Однако обратившись к графику функции потерь на тесте мы поймём, что после 10-15 итерации сеть уже начала переобучаться.

5 Список литературы

- Оригинальная статья: <https://arxiv.org/abs/1904.09237v1>
- Оригинальная статья создателей Adam: <https://arxiv.org/pdf/1412.6980.pdf>
- Прекрасное видео человека из Индии чуть подробнее объясняющего синтетический пример:
<https://www.youtube.com/watch?v=j-uLz9TWov4>
- Курс по нейросетям, на котором основаны эксперименты: <https://stepik.org/course/50352/promo>