

1. Постановка задачи

Целью работы было оценить практическую применимость статьи [1]. Были выбраны следующие критерии оценки:

- 1) Стабильность обучения
- 2) Качество обучения
- 3) Скорость сходимости сети
- 4) Обобщающая способность
- 5) Влияние гипер параметра (регуляризующий параметр λ) на пункты 1-4.

Для проведения тестов использовались следующие датасеты:

- 1) 2 концентрические, 100-мерные сферы различного радиуса.
- 2) 2 100-мерных набора семплов из нормальных распределений с одинаковыми дисперсиями и разными средними.
- 3) Подмножество датасета CIFAR (классы 0 и 1). Размер датасета: 2000 - train, 2000 - test, 2000 - validation.
- 4) Подмножество датасета MNIST (цифры 1 и 7). Размер датасета: 13000 - train, 2000 - test, 2000 - validation.

2. Стабильность

В ходе обучения сети с экспоненциальным нейроном проявлялся эффект нестабильности обучения - в любой момент ошибка на обучающей выборке может "взорваться" и, затем, как следствие, расходятся градиенты и веса сети. Причина этого в чрезвычайной чувствительности экспоненциального нейрона к малым изменениям весов в нём. Это подтверждает такой эксперимент: к весам нейрона добавляется нормальный шум из $N(0, \sigma)$, где σ в 10 раз меньше по величине чем среднее значение модулей весов в нейроне. Такое изменение экспоненциального нейрона сразу приводит к расхождению сети, хотя в случае нейрона с не экспоненциальной функцией активации (ReLU, Sigmoid, ...) такого не происходит. Частично эту проблему удаётся решить нормализацией входных данных, за счёт чего на ранних итерациях обучения не происходит "взрыва" градиентов. Так же, уменьшение learning rate в процессе обучения уменьшает частоту таких ситуаций. Однако полностью проблему решить не удалось.

3. Качество

Основным результатом статьи [1] было доказательство отсутствия локальных минимумов у получившейся сети. Было выдвинуто предположение, что существуют задачи, в которых обучение сети без экспоненциального приводит к "застреванию" в локальном минимуме, в то время как добавление экспоненциального нейрона приведёт к попаданию в глобальный минимум и, как следствие лучшему качеству на обучающей выборке. Однако, проведённые эксперименты показывают неоднозначную картину:

Датасет	Train Size	Без эксп. нейрона	С эксп. нейроном
2 сферы, Радиусы 1.0 и 3.0	10000	0.000\0	0.000\0
2 гауссианы, Средние 0.0 и 0.1	1000	0.000\0	0.000\0

Список литературы

- [1] Shiyu Liang, Ruoyu Sun, Jason D. Lee, R. Srikant, Adding One Neuron Can Eliminate All Bad Local Minima, arXiv:1805.08671