

1. Что такое слой субдискретизации в сверточных нейронных сетях?

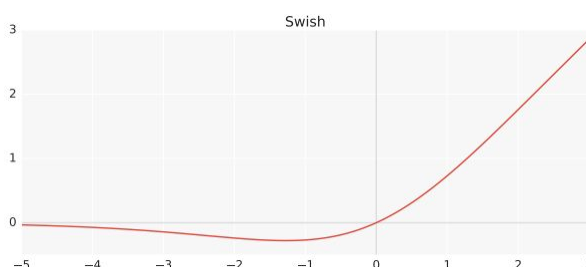
В сверточных нейронных сетях используется 3 вида слоёв: сверточные, субдискретизирующие и полносвязные (так сказать “обычные” слои НС типа персептрон). Задачей конкретно субдискретизирующего слоя является уменьшение размерности карты признаков предыдущего слоя. Это делается для сокращения вычислительных затрат и расхода памяти, избежания переобучения, а также для выделения наиболее значимых признаков (т.к. после выявления каких-либо признаков на предыдущем шаге нам больше не нужно настолько подробное изображение). Сам процесс описывается так: мы разделяем карту признаков, полученную на сверточном слое, на фиксированное количество частей n (обычно размер каждой части 2×2) и на каждой части выбираем максимальное по значению. В итоге мы уменьшаем размерность нашей карты признаков в 2 раза.

2. Что такое L1 регуляризация?

L1 регуляризация - способ разрежения. Используется при большом наборе входных данных и малом количестве наблюдений. Мы должны оставить небольшое количество важных факторов и удалить те, что являются шумом. При L1 регуляризации дополнительный член является модулем, производная модуля является константой с абсолютной величиной, равной единице. Градиентный спуск при такой регуляризации стремится к 0 с постоянной скоростью. Тем самым получается, что в одномерном случае L1 регуляризация способствует обращению в 0 весов тех признаков, которые являются шумом.

3. Всегда ли функции активации должны быть возрастающими или убывающими?

Нет, например функция активации Swish.



4. Найдите ошибки в подписях к графикам 5 и далее

Усредненные графики для средней абсолютной ошибки.

5. Как называется нормировка, используемая в лаб. работе?

Стандартизованная оценка (или Z-оценка).

6. Сколько слоев в модели в лаб. работе?

В данной лаб. работе для каждого блока строится модель из 3-ёх слоев.

7. Почему в качестве оптимизатора выбран rmsprop?

Метод rmsprop является модификацией алгоритма Adagrad. Идея алгоритма состоит в менее частом обновлении весов для типичных признаков. Также данный оптимизатор способен уменьшать скорость обучения, когда это нужно.

Формула обновления весов:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} g_t$$

Знаменатель есть корень из среднего квадратов градиентов, отсюда RMSProp — root mean square propagation:

$$RMS[g]_t = \sqrt{E[g^2]_t + \epsilon}$$

В данной работе в качестве оптимизатора выбран rmsprop скорее всего по той причине, что какие-то типичные признаки у нас будут обновляться часто, поэтому нам нужно уменьшить их обновление.