

1. Если двумерные сигналы имеют форму  $(m, k)$ , то датасет можно представить в виде тензора третьего ранга с формой  $(n, m, k)$ , т. е. в каждом слое записан двумерный сигнал.

2. Скорость обучения — это шаг в методах градиентного спуска, которые используются в обучении модели. Выбор шага может сильно повлиять на скорость сходимости и результат. Если брать большой шаг, то скорость сходимости будет большой, однако есть вероятность пропустить глобальный минимум. Если взять небольшой шаг, то скорость сходимости понизится, но пропустить минимум в такой случае сложнее, однако можно попасть в локальный минимум.

3. RMSProp учитывает историю обновлений весов. Часто обновляемые веса обновляются меньше. Для этого используется так называемое «бегущее среднее», которое вычисляется следующим образом:

$$E[g^2]_t = \beta E[g^2]_t + (1 - \beta) g_t^2,$$

где  $\beta$  - коэффициент сохранения,  $g_t$  – градиент функции потерь.

Обновление весов происходит по следующей формуле:

$$w_t = w_{t-1} - \frac{\eta}{\sqrt{E[g^2]_t}} g_t$$

4. Как видно из графиков, переобучение возникает очень быстро: точность на обучающих данных растет, когда как на проверочных не меняется в лучшую сторону. Это означает, что нейронной сети недостаточно этих обучающих данных для определения всех закономерностей. С точки зрения архитектуры сети(числа слоев, нейронов на них и функций активации), она достаточна для решения задачи, поскольку на обучающих данных была достигнута высокая точность. При недостатках архитектуры сети возникло бы недообучение. При разных количествах блоков при перекрестной проверке размер обучающего датасета также меняется: чем больше число блоков, тем больше тренировочных данных. Этим можно объяснить улучшение точности при большем числе блоков. При 6-7 блоках достигается наилучшая точность.

5. В строках 11-17 происходит нормализация входных признаков. После нормализации значения признаков датасета имеют выборочное среднее 0 и среднее квадратичное отклонение равное 1. При отличающихся порядках признаков также сильно будут отличаться веса, что усложнит обучение сети.

6. Для проверки работы сети датасет разбивается на обучающий и проверочный. От того, какие данные отправятся в обучающий, а какие — в проверочный, может зависеть результат проверки. Поэтому для получения более или менее адекватной оценки применяется перекрестная проверка. При перекрестной проверке обучение и проверка происходит по K блокам, после чего точность при проверках усредняется. Это значение и считается точностью перекрестной проверки.

7. MSE(mean square error) это среднеквадратичная ошибка, вычисляются как сумма квадратов разности предсказанных значений и истинных, деленная на число предсказанных значений:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

MAE(mean absolute error) это средняя абсолютная ошибка, вычисляются как абсолютная разность предсказанных значений и истинных, деленная на число предсказанных значений:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$