

### 1. Как можно представить датасет двумерных сигналов?

Датасет из двумерных сигналов, которые имеют форму  $(s, k)$ , можно представить в виде тензора третьего ранга с формой  $(n, s, k)$ , то есть в каждом слое записан двумерный сигнал.

### 2. Что отличает методы обучения первого порядка от методов второго порядка?

Методы обучения первого порядка используют методы оптимизации первого порядка, использующие градиент. А в методах второго порядка используются методы оптимизации второго порядка, использующие гессиан (матрицу вторых производных).

### 3. Что такое L2-регуляризация?

Регуляризация – метод, который добавляет некоторые ограничения к условию, с целью предотвратить переобучение. Так как “симптомом” переобучением являются большие веса модели, то имеет смысл их ограничить с помощью L2-регуляризации.

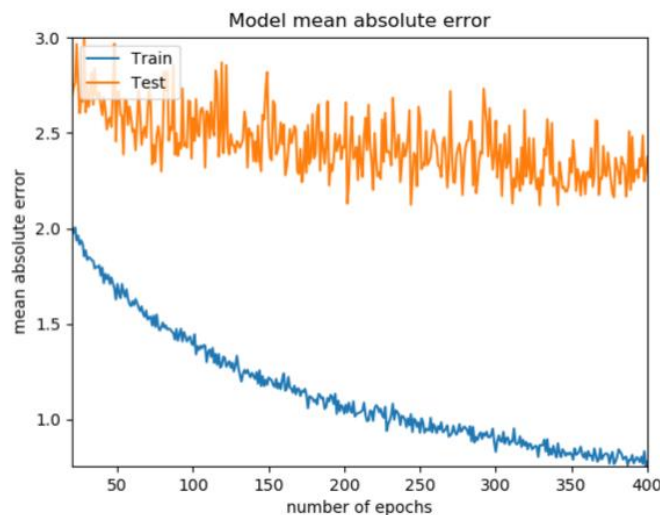
L2-регуляризации – это когда к функционалу ошибки (доля неправильных ответов)  $Q(w, X)$  прибавляется  $\lambda \sum_j w_j^2$ , где  $\lambda$  это коэффициент регуляризации,  $w_j$  веса модели. Чем больше  $\lambda$

тем ниже сложность модели, но чем она меньше, тем выше риск переобучения. Таким образом, нужно искать некоторый баланс - выбирать  $\lambda$  такой, что, с одной стороны, она не допускает переобучения, с другой - позволяет делать модель достаточно сложной, чтобы уловить все закономерности в данных.

В итоге получается, что задача  $Q(w, X) + \lambda \sum_j w_j^2 \rightarrow \min_w$  заменяется на 
$$\begin{cases} Q(w, X) \rightarrow \min_w \\ \sum_j w_j^2 \leq C \end{cases}$$

То есть сумма квадратов весов модели ограничивается некоторой константой  $C$

### 4. Как была найдена точка переобучения?



По графику видно, что после 50 эпохи ошибка на обучающих данных продолжает уменьшаться, а на тренировочных данных растет.

## 5. Для чего необходима стандартизация данных?

Иногда бывает, что признаки могут находиться в разных диапазонах. Передать такие признаки в модель будет проблематично, так как это усложнит ее обучение. Для того чтобы избежать этого усложнения принято делать стандартизацию входных признаков: для каждого признака во входных данных (столбца в матрице входных данных) из каждого значения вычитается среднее по этому признаку, и разность делится на стандартное отклонение, в результате признак центрируется по нулевому значению и имеет стандартное отклонение, равное единице.

## 6. Почему анализируется $mae$ , а не $mse$ ?

$mae$  дает возможность понять, насколько сильно число, которое нам выдала нейронная сеть, отличается от правильного ответа. Также можно использовать  $mse$ , но  $mae$  проще интерпретировать.

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

Несомненным преимуществом  $mae$  является то, что модули не увеличивают в разы отклонения, считающиеся выбросами. Поэтому эта оценка является лучше, чем  $MSE$  и фактически соответствует медиане.

## 7. Почему анализируется $mae$ , а не $mse$ ?

Перед применением перекрёстной проверки перемешать данные