

1. Как можно представить датасет из изображений в формате CMYK?

Изображение: ширина, высота, цвет (CMYK – имеет 4 канала) $w \times h \times 4$

Датасет изображений : изображение, высота, ширина, цвет $(n, w, h, 4)$, то есть партия изображений может быть сохранена в таком тензоре $(n, w, h, 4)$

2. Что такое коэффициент регуляризации?

Регуляризация – метод борьбы с переобучением (один из способов уменьшение сложности сети). регуляризация выполняется посредством наложения штрафов на веса с наибольшими значениями с использованием параметра λ — коэффициент регуляризации, выражающий предпочтение минимизации нормы относительно минимизации потерь на обучающем множестве.

часто регуляризация представляет собой добавку $R(w)$ к функции потерь $L(f(x, w), y)$

$$\min_w \sum L(x_i, w, y_i) + \lambda R(w)$$

Если коэффициент слишком мал, то эффект от регуляризации будет ничтожен, если же слишком велик — модель обнулит все веса.

3. Для чего необходимо проводить стандартизацию данных?

это может быть небезопасно для ИНС, например, данные, которые имеют большое значение (которые намного больше изначальных весов в сети), или разнородные данные (например, одни находятся в промежутке от 200 до 300, а другие от 0 до 1 и т.д.). Это может привести к значительному изменению градиента, которые могут препятствовать сходимости сети. Поэтому данные стандартизируются так, чтобы среднее было 0, а стандартное отклонение 1.

4. на графиках ошибок в отчете, для контрольной выборки ошибка изначально меньше чем для обучающей выборки. Чем это можно обосновать?

Возможно сеть чему-то научилась. Это можно обосновать тем, что ошибка на тестовых данных считалась после итерации обучения и подсчета ошибки на тренировочных данных.

5. На основании какой метрики идет коррекция весов в данной работе?

Mean absolute error (MAE) или средняя абсолютная ошибка.

понятие точности неприменимо для регрессии, поэтому для оценки качества часто применяется средняя абсолютная ошибка

Абсолютная разность между предсказанными значениями и истинными.

$$\text{MAE} = 1/N \sum |y_i - y_i'|$$

6. Почему была выбрана проверка на 6 блоках?

Для того, чтобы проверить, как ведет себя сеть на бОльших количествах блоков. Разбив множество всех данных на более мелкие части, можно увидеть как будет меняться ошибка.

В общем случае смещение метода кросс-валидации зависит от размера валидационной выборки. Если размер валидационной выборки составляет 50% исходных данных (2-блочная кросс-валидация), итоговая оценка СКО будет более смещенной, чем в случае, когда этот размер составляет 10% исходных данных. Определенных рекомендаций по количеству блоков нет.

7. В 10 строчке "model.add(Dense(64, activation='relu', input_shape=(train_data.shape[1],)))" - почему input_shape имеет подобное значение (после запятой ничего не стоит)?

После запятой нет ничего, потому что может быть любое количество строк (образцов), то есть если input_shape=(16,), то модель может принимать массивы (*, 16)