

1. Что такое коэффициент момента?
Если я правильно понял вопрос, то момент является синонимом импульса в методе оптимизации с использованием градиентного спуска. Он необходим для того, чтобы перемещение точки по графику описывалось не только шагом но и кривизной кривой, это позволяет снизить риск застрять в локальных минимумах функции.
2. Что такое ядро свертки?
Ядро свёртки можно назвать фильтром, он относится к оператору, применяемому ко всему изображению таким образом, что преобразует информацию, закодированную в пикселях. На практике ядро это матрица меньшего размера по сравнению с входными размерами изображения, состоящая из вещественных записей.
3. К чему ведет застревание в локальном минимуме в ходе обучения?
При попытках выбраться из локального минимума будут увеличиваться потери, также если шаг будет обучения слишком мал, мы можем застрять в локальном минимуме, не найдя глобальный.
4. Чем можно объяснить поведение графиков на рисунке 4?
С некоторого момента при увеличении количества эпох начинается переобучение, в связи с чем увеличивается ошибка.
5. Всегда ли необходима нормировка данных?
Нет, не всегда. Сеть сможет адаптироваться и под ненормализованные данные, но это усложнит и замедлит обучение, т.к. данные имеют значения из различных диапазонов.
6. Что это за параметры "best_mae = [1, 0, 0] best_val_mae = [1, 0, 0]"
Первый параметр значение mae, второй и третий количество эпох и блоков, соответственно.
Эти параметры я хотел использовать для упрощения нахождения модели с наименьшей абсолютной ошибкой, но так как я некорректно указал начальные параметры, пришлось опираться исключительно на графики.
7. В чем ошибка у Вас при разбиении данных на обучающую и контрольную выборку?
Возможно модель следовало обучать на всём объёме обучающих данных, без разбиения их на части.