

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. Н.Э. Баумана

Кафедра «Систем обработки информации и управления»

ОТЧЕТ

Домашняя работа №1, 2
по курсу «Проектирование интеллектуальных систем»

ИСПОЛНИТЕЛЬ:

группа ИУ5-22

Колпаков М.О.

ФИО

подпись

"__" _____ 2020 г.

ПРЕПОДАВАТЕЛЬ:

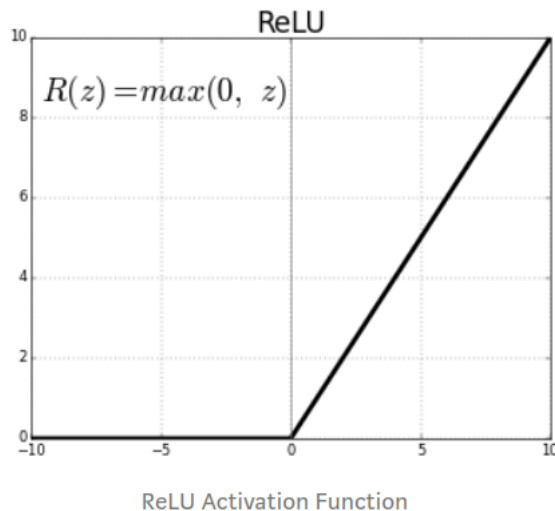
ФИО

подпись

"__" _____ 2020 г.

Москва - 2020

1. Расчет градиента. Выведите градиент для relu функции.



The ReLU activation function $g(z) = \max\{0, z\}$ is not differentiable at $z = 0$. A function is differentiable at a particular point if there exist left derivatives and right derivatives and both the derivatives are equal at that point. ReLU is differentiable at all the point except 0. the left derivative at $z = 0$ is 0 and the right derivative is 1.

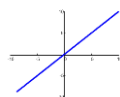
2. Для чего используется функция активации? Перечислите виды активационной функции.

Функция активации решает, должен ли нейрон быть активирован или нет, путем вычисления взвешенной суммы и дальнейшего добавления смещения с ней. Цель функции активации состоит в том, чтобы ввести нелинейность в выход нейрона.

Функция активации — это способ нормализации входных данных (мы уже говорили об этом ранее). То есть, если на входе у вас будет большое число, пропустив его через функцию активации, вы получите выход в нужном вам диапазоне.

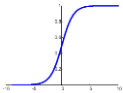
Линейная функция

$$f(x) = x$$



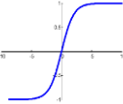
Эта функция почти никогда не используется, за исключением случаев, когда нужно протестировать нейронную сеть или передать значение без преобразований.

Сигмоид

$$f(x) = \frac{1}{1+e^{-x}}$$


Это самая распространенная функция активации, ее диапазон значений $[0,1]$. Именно на ней показано большинство примеров в сети, также ее иногда называют логистической функцией. Соответственно, если в вашем случае присутствуют отрицательные значения (например, акции могут идти не только вверх, но и вниз), то вам понадобится функция которая захватывает и отрицательные значения.

Гиперболический тангенс

$$f(x) = \frac{e^{2x} - 1}{e^{2x} + 1}$$


Имеет смысл использовать гиперболический тангенс, только тогда, когда ваши значения могут быть и отрицательными, и положительными, так как диапазон функции $[-1,1]$. Использовать эту функцию только с положительными значениями нецелесообразно так как это значительно ухудшит результаты вашей нейросети.

Функция ReLU

$$f(z) = \max(0, z)$$



Rectified Linear Unit — это наиболее часто используемая функция активации при глубоком обучении. Данная функция возвращает 0, если принимает отрицательный аргумент, в случае же положительного аргумента, функция возвращает само число. То есть она может быть записана как $f(z) = \max(0, z)$. На первый взгляд может показаться, что она линейна и имеет те же проблемы что и линейная функция, но это не так и ее можно использовать в нейронных сетях с множеством слоев. Функция ReLU обладает несколькими преимуществами перед сигмойдой и гиперболическим тангенсом:

Очень быстро и просто считается производная. Для отрицательных значений — 0, для положительных — 1.

Разреженность активации. В сетях с очень большим количеством нейронов использование сигмоидной функции или гиперболического тангенса в качестве активационной функции влечет активацию почти всех нейронов, что может сказаться на производительности обучения модели. Если же использовать ReLU, то количество включаемых нейронов станет меньше, в силу характеристик функции, и сама сеть станет легче.

У данной функции есть один недостаток, называемый проблемой умирающего ReLU^[2]. Так как часть производной функции равна нулю, то и градиент для нее будет нулевым, а это значит, что веса не будут изменяться во время спуска и нейронная сеть перестанет обучаться.

Функцию активации ReLU следует использовать, если нет особых требований для выходного значения нейрона, вроде неограниченной области определения. Но если после обучения модели результаты получились не оптимальные, то стоит перейти к другим функциям, которые могут дать лучший результат.

3. Свертка. а) Чему равна размерность карт признаков? б) Напишите значения карт признаков, если шаг ядра (stride) равен 2.

Дано:

$$h=w=5$$

$$k_w=k_h=3$$

Решение:

h - высота изображения

w - ширина изобра

s - шаг

k_w - ширина ядра

k_h - высота ядра

p -отступ

$$\left\lfloor \frac{w + 2 * p - k_w}{s} + 1 \right\rfloor \times \left\lfloor \frac{h + 2 * p - k_h}{s} + 1 \right\rfloor$$

$$p = 0$$

$$s=2$$

$$[(5+2*0 - 3)/2 + 1] \times [(5+2*0 - 3)/2 + 1]=2 \times 2 - \text{размерность}$$

Значения карт вычисляем по принципу:

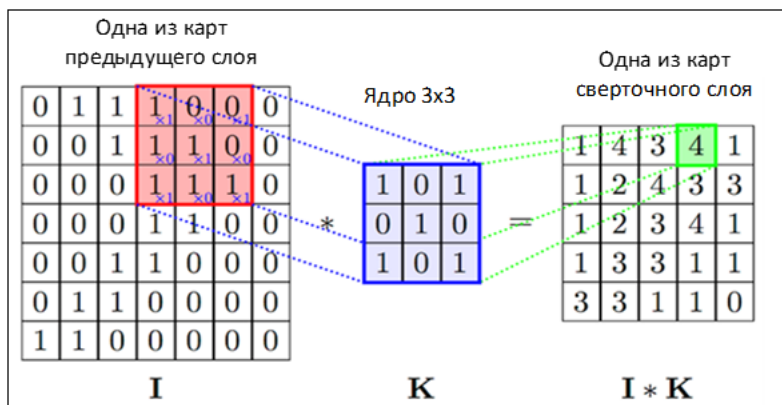


Рисунок 4 — Операция свертки и получение значений сверточной карты (valid)

Значения карт:

189	167
129	197