**МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ, СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ ИМ. ПРОФ. М.А. БОНЧ-БРУЕВИЧА»**

**(СПбГУТ)**

Факультет инфокоммуникационных сетей и систем (ИКСС)

Курсовая работа по теме «Разработка нейронной сети для определения десятичных цифр по изображению»

по курсу

«Операционные системы и сети»

Группа: ИКПИ-14

Выполнил студент: Хохлов Т. В.

Принял преподаватель: Дагаев А.В.

Место для подписи \_\_\_\_\_\_\_\_\_\_\_\_\_

Санкт-Петербург

2023 г.

1. Что такое нейронная сеть.
   1. Тип используемой нейронной сети.
   2. Параметры
      1. Нейроны
      2. Слой
      3. Веса и связи
      4. Функции активации
      5. Оптимизатор
2. Что такое машинное зрение

**Нейронная сеть (Neural Network)**

Нейронная сеть — это вычислительная модель, вдохновленная структурой и функцией нервной системы живых организмов. Она применяется в области машинного обучения для анализа данных, распознавания образов, и решения различных задач, от классификации до генерации контента. Для полного понимания структуры и работы нейронных сетей, нужно подробно рассмотреть на её составляющие элементы и параметры.

**Связь с машинным зрением**

Машинное зрение — область машинного обучения, посвященная разработке методов и алгоритмов, позволяющих компьютерам анализировать и интерпретировать визуальную информацию, аналогично человеческому зрению.

Нейронные сети в машинном зрении применяются для распознавания образов на изображениях. Входные данные (изображения) подаются на входной слой сети, где они проходят через сверточные и пулинг слои, извлекая признаки. Далее, полносвязные слои и функции активации позволяют сети классифицировать или регрессировать данные.

В контексте машинного зрения, нейронные сети обучаются извлекать визуальные признаки, что позволяет им распознавать объекты, лица, и выполнять другие задачи визуального анализа. Эта способность нейронных сетей делает их мощным инструментом в сфере обработки изображений и расширяет возможности автоматизации и анализа визуальных данных.

**Что такое нейрон?**



Рис. 1 – Схема составного персептрона

Нейрон — это вычислительная единица, которая получает информацию, производит над ней простые вычисления и передает ее дальше. Они делятся на три основных типа (рис. 1): входной (синий), скрытый (красный) и выходной (зеленый). В том случае, когда нейросеть состоит из большого количества нейронов, вводят термин слоя. Соответственно, есть входной слой, который получает информацию, n скрытых слоев (обычно их не больше 3), которые ее обрабатывают и выходной слой, который выводит результат. У каждого из нейронов есть 2 основных параметра: входные данные **(input data)** и выходные данные **(output data)**. В случае входного нейрона: input=output (рис. 2). В остальных, в поле input попадает суммарная информация всех нейронов с предыдущего слоя, после чего, она нормализуется, с помощью функции активации и попадает в поле output.

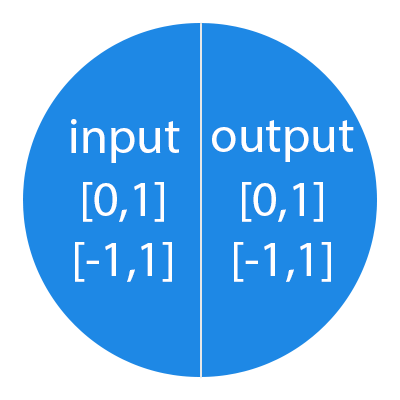


Рис. 2 – Структура входного нейрона

Нейроны оперируют числами в диапазоне [0,1] или [-1,1]. Для обработки чисел вне данного диапазона необходимо провести процесс нормализации (поделить 1 на необходимое число).

**Что такое синапс?**

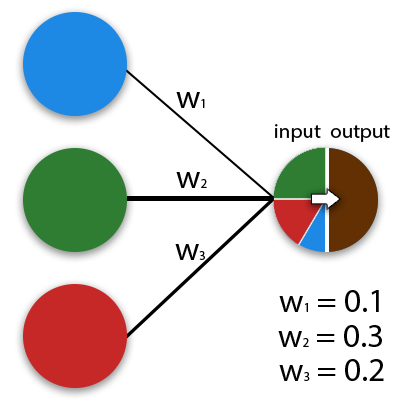


Рис. 3 – Схема примитивного персептрона

**Синапс** — это связь между двумя нейронами. У синапсов есть 1 параметр — **вес**. Благодаря ему, входная информация изменяется, когда передается от одного нейрона к другому. Например, имеется 3 нейрона, которые передают информацию следующему. Тогда имеются 3 веса, соответствующие каждому из этих нейронов. Нейроны, связанные синапсом с наибольшим весом будут наиболее актуальными, то есть информация предыдущего нейрона будет содержать доминирую в следующем нейроне. На самом деле, совокупность весов нейронной сети или матрица весов — это своеобразный мозг всей системы. Именно благодаря этим весам, входная информация обрабатывается и превращается в результат.

Важно помнить, что во время инициализации нейронной сети, веса расставляются в случайном порядке.

**Как работает нейронная сеть?**

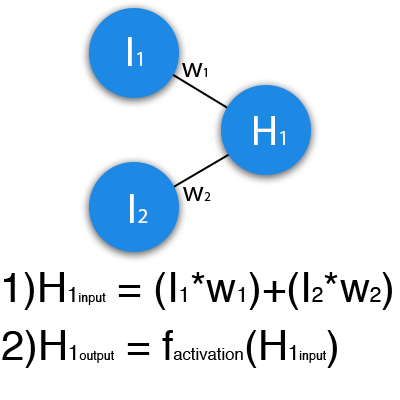


Рис. 3 – Схема примитивного персептрона, I-входной нейрон, H-скрытый нейрон.

Для простоты и полного понимая работы нейронных сетей, возьмём пример изображенный на рисунке 3. На нём изображён примитивный персептрон, состоящий из двух входных нейронов (I1 и I2) и одного скрытого (H1) соединённых весами w1 и w2. Из формулы 1 видно, что входная информация, переходящая на следующий слой — это сумма всех входных данных, умноженных на соответствующие им веса. В представленной формуле входные нейроны, представляют собой наличие определённой характеристики (или же степень привязанности) для рассматриваемого объекта, а веса – степень актуальности данной характеристики. Как было описано ранее, веса в только проинициализированной нейронной сети расставляются в случайном порядке, т.е. можно предположить, что у нас имеются какие-то входные данные (теоретические характеристики и проинициализированные веса), но как понять какие веса для определенной задачи будут актуальными? И ведь действительно, когда у нас есть входные данные, мы передаем по слоям «вперёд», выполняя вычисления степени их актуальности и так, определённое значение, пока не дойдем до выходного нейрона. Запустив такую сеть в первый раз, мы увидим, что ответ далек от правильно, потому что сеть не натренирована. Из этого можно сделать логический вывод, что тренировка нейронных сетей, заключается в определении конкретных диапазонов значений весов для связующих входных и выходных нейронов. Этим занимается промежуточный этап – нормализация данных с помощью **функции активации** и их сохранение в модели нейронной сети.

**Функция активации или слой активации**

Слои активации — это один из основных типов слоев, которые используются в нейронных сетях. Они представляют собой функцию, которая добавляет нелинейность к выходу предыдущего слоя. Это позволяет нейронной сети лучше моделировать сложные функции и более точно предсказывать результаты.

Слои активации принимают на вход результаты предыдущего слоя, называемые входом, и преобразуют их в выходное значение, которое передается следующему слою. Для этого они используют функцию активации, которая определяет, каким образом данные будут преобразованы.

В данной работе используются два вида функций активации, это **ReLU** и **SoftMax**.

**Функция активации ReLU**

ReLU (Rectified Linear Unit) — это нелинейная функция активации, которая широко используется в глубоком обучении. Она преобразует входное значение в значение от 0 до положительной бесконечности. Если входное значение меньше или равно нулю, то ReLU выдает ноль, в противном случае - входное значение.

Математически ReLU определяется следующим образом:



где max - функция, возвращающая максимальное значение из двух.

Графически ReLU выглядит как линейная функция с нулевым отсечением на оси абсцисс в точке 0. Это значит, что функция имеет постоянный наклон во всех точках, кроме точки 0, где происходит отсечение.

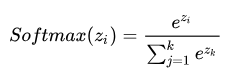


ReLU имеет несколько преимуществ по сравнению со сигмоидной функцией активации. Во-первых, ReLU более вычислительно эффективна, поскольку она является простой и быстрой операцией, которая не требует вычисления экспоненты. Во-вторых, ReLU решает проблему затухания градиента, так как она не вызывает затухания градиента при обратном распространении ошибки, как это происходит в случае с сигмоидной функцией активации.

**Функция активации Softmax**

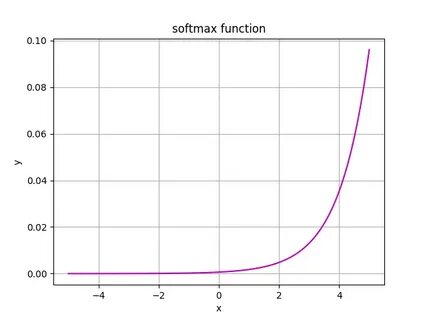
Функция Softmax используется для преобразования вектора значений в вероятностное распределение, которое суммируется до 1. Она особенно полезна в многоклассовой классификации, где необходимо определить вероятности для каждого класса.

Формула функции Softmax выглядит следующим образом:



где z\_i - это элемент входного вектора, а k - это общее число элементов в векторе.

График функции Softmax представляет собой гладкую кривую, начинающуюся от 0 и заканчивающуюся на 1, что соответствует сумме вероятностей всех элементов вектора. Кривая функции Softmax имеет свойство, что вероятность любого элемента вектора увеличивается, если значения других элементов уменьшаются, что позволяет использовать эту функцию для многоклассовой классификации.



Хотя функция Softmax имеет множество применений в машинном обучении, она также может иметь недостатки, такие как чувствительность к выбросам и несбалансированным данным, что может приводить к неверным вероятностным оценкам.