## Нейронные сети

## Что такое нейронная сеть

Нейро́нная сеть[[1]](https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D1%81%D0%B5%D1%82%D1%8C#cite_note-1) (также иску́сственная нейро́нная сеть, ИНС, или просто нейросе́ть) — [математическая модель](https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C), а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей — сетей [нервных клеток](https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D0%B9%D1%80%D0%BE%D0%BD) живого организма. Это понятие возникло при изучении процессов, протекающих в [мозге](https://ru.wikipedia.org/wiki/%D0%9C%D0%BE%D0%B7%D0%B3), и при попытке [смоделировать](https://ru.wikipedia.org/wiki/%D0%9C%D0%BE%D0%B4%D0%B5%D0%BB%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5) эти процессы. Первой такой [попыткой](https://ru.wikipedia.org/wiki/%D0%9C%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C_%D0%BC%D0%BE%D0%B7%D0%B3%D0%B0) были нейронные сети [У. Маккалока](https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D0%BA%D0%BA%D0%B0%D0%BB%D0%BE%D0%BA,_%D0%A3%D0%BE%D1%80%D1%80%D0%B5%D0%BD) и [У. Питтса](https://ru.wikipedia.org/wiki/%D0%9F%D0%B8%D1%82%D1%82%D1%81,_%D0%A3%D0%BE%D0%BB%D1%82%D0%B5%D1%80)[[2]](https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D1%81%D0%B5%D1%82%D1%8C#cite_note-Мак-Каллок-2).

Нейронная сеть — это последовательность нейронов, соединенных между собой синапсами. Структура нейронной сети пришла в мир программирования прямиком из биологии. Благодаря такой структуре, машина обретает способность анализировать и даже запоминать различную информацию. Нейронные сети также способны не только анализировать входящую информацию, но и воспроизводить ее из своей памяти.

Нейронные сети используются для решения сложных задач, которые требуют аналитических вычислений подобных тем, что делает человеческий мозг. Самыми распространенными применениями нейронных сетей является:  
  
**Классификация** — распределение данных по параметрам. Например, на вход дается набор людей и нужно решить, кому из них давать кредит, а кому нет. Эту работу может сделать нейронная сеть, анализируя такую информацию как: возраст, платежеспособность, кредитная история и тд.  
  
**Предсказание** — возможность предсказывать следующий шаг. Например, рост или падение акций, основываясь на ситуации на фондовом рынке.  
  
**Распознавание** — в настоящее время, самое широкое применение нейронных сетей. Используется в Google, когда вы ищете фото или в камерах телефонов, когда оно определяет положение вашего лица и выделяет его и многое другое.

## Нейрон

Нейрон — это вычислительная единица, которая получает информацию, производит над ней простые вычисления и передает ее дальше. Они делятся на три основных типа: 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. В остальных, в поле input попадает суммарная информация всех нейронов с предыдущего слоя, после чего, она нормализуется, с помощью функции активации (пока что просто представим ее f(x)) и попадает в поле output.

**Важно помнить**, что нейроны оперируют числами в диапазоне [0,1] или [-1,1]. А как же, вы спросите, тогда обрабатывать числа, которые выходят из данного диапазона? На данном этапе, самый простой ответ — это разделить 1 на это число. Этот процесс называется нормализацией, и он очень часто используется в нейронных сетях.

## Синапс

Синапс это связь между двумя нейронами. У синапсов есть 1 параметр — вес. Благодаря ему, входная информация изменяется, когда передается от одного нейрона к другому. Допустим, есть 3 нейрона, которые передают информацию следующему. Тогда у нас есть 3 веса, соответствующие каждому из этих нейронов. У того нейрона, у которого вес будет больше, та информация и будет доминирующей в следующем нейроне (пример — смешение цветов). 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.

## Функции активации

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. Функций активации достаточно много поэтому мы рассмотрим самые основные: Линейная, Сигмоид (Логистическая), ReLU и Гиперболический тангенс. Главные их отличия — это диапазон значений.

### Тренировочный сет

Тренировочный сет — это последовательность данных, которыми оперирует нейронная сеть. В нашем случае исключающего или (xor) у нас всего 4 разных исхода то есть у нас будет 4 тренировочных сета: 0xor0=0, 0xor1=1, 1xor0=1,1xor1=0.

### Итерация

Это своеобразный счетчик, который увеличивается каждый раз, когда нейронная сеть проходит один тренировочный сет. Другими словами, это общее количество тренировочных сетов пройденных нейронной сетью.

## Эпоха

При инициализации нейронной сети эта величина устанавливается в 0 и имеет потолок, задаваемый вручную. Чем больше эпоха, тем лучше натренирована сеть и соответственно, ее результат. Эпоха увеличивается каждый раз, когда мы проходим весь набор тренировочных сетов, в нашем случае, 4 сетов или 4 итераций.

Важно не путать итерацию с эпохой и понимать последовательность их инкремента. Сначала n  
раз увеличивается итерация, а потом уже эпоха и никак не наоборот. Другими словами, нельзя сначала тренировать нейросеть только на одном сете, потом на другом и тд. Нужно тренировать каждый сет один раз за эпоху. Так, вы сможете избежать ошибок в вычислениях.

## Ошибка

Ошибка — это процентная величина, отражающая расхождение между ожидаемым и полученным ответами. Ошибка формируется каждую эпоху и должна идти на спад. Если этого не происходит, значит, вы что-то делаете не так.

## Алгоритм обратного распространения ошибки

Принцип работы обратного распространения ошибки основан на математическом методе поиска градиента. Когда нейронная сеть получает входные данные и проходит через слои, она вычисляет свое текущее предсказание и сравнивает его с ожидаемым выходом. Затем алгоритм обратного распространения ошибки вычисляет ошибку для каждого нейрона и корректирует веса, чтобы минимизировать эту ошибку. Процесс повторяется множество раз, пока сеть не достигнет приемлемой точности.

**Обратное распространение ошибки** – это алгоритм, используемый в обучении нейронных сетей. Он позволяет оценивать и корректировать веса связей между нейронами на основе разницы между предсказанным и ожидаемым результатом.

Когда нейронная сеть производит предсказание, оно может быть несовершенным или даже неправильным. Обратное распространение ошибки помогает нейронной сети понять, какие именно связи между нейронами нуждаются в корректировке, чтобы улучшить качество предсказаний.

Алгоритм обратного распространения ошибки основан на принципе градиентного спуска. Суть его заключается в том, что для каждой связи между нейронами вычисляется величина, насколько изменение этой связи влияет на ошибку предсказания. Затем веса связей корректируются в направлении, противоположном градиенту, чтобы минимизировать ошибку предсказания.

Обратное распространение ошибки проходит через все слои нейронной сети, начиная с выходного слоя и заканчивая входным слоем. Каждый слой вычисляет свою ошибку и передает ее назад по связям к предыдущему слою. Таким образом, веса связей нейронной сети постепенно корректируются, чтобы обеспечить более точные предсказания.

Обратное распространение ошибки является одним из ключевых методов обучения нейронных сетей. Он позволяет нейронной сети самостоятельно «обучаться» на основе предоставленных данных и корректировать свои параметры, чтобы достичь наилучшего качества предсказаний.

Применение обратного распространения ошибки включает несколько этапов:

1. Прямое распространение сигнала. Входные сигналы передаются по сети от нейрона к нейрону. Каждый нейрон вычисляет свою активационную функцию от входных данных и передает результат следующему нейрону.
2. Вычисление ошибки. После прямого распространения сигнала вычисляется разница между полученными и ожидаемыми выходными значениями сети. Эта разница является ошибкой, которую необходимо минимизировать.
3. Обратное распространение ошибки. Ошибка сети распространяется в обратном направлении, от выходных нейронов к входным. Каждый нейрон вычисляет градиент ошибки по отношению к его входным данным и весам связей.
4. Корректировка весов. Используя градиент ошибки, веса связей между нейронами корректируются с помощью заданного обновления весового коэффициента. Целью является минимизация ошибки и улучшение результатов сети.