**Основы вычислительных нейронных сетей**

**Архитектура нейронной сети**

Для начала введем определение вычислительных нейронных сетей. Вычислительные нейронные сети — это модель машинного обучения, вдохновленная структурой и функциями человеческого мозга. Они широко используются в различных областях, таких как компьютерное зрение, обработка естественного языка и распознавание речи, для выполнения сложных задач, таких как классификация изображений, языковой перевод и распознавание речи. В этом параграфе показаны основные принципы вычислительных нейронных сетей, включая их архитектуру, процесс обучения и различные типы нейронных сетей, обычно используемые на практике. Понимание этих основ необходимо для создания эффективных моделей нейронных сетей для широкого спектра приложений.

Рассмотрим подробнее архитектуру нейронных сетей. Перед этим вкратце будет описано устройство биологической нейронной сети. Биологическая нейронная сеть состоит из огромного количества взаимосвязанных клеток, называемых нейронами, которые взаимодействуют друг с другом через специализированные структуры, называемые синапсами. Каждый нейрон обычно имеет тело клетки, дендриты и аксон. Дендриты получают электрические импульсы от других нейронов, а аксон передает этот импульс дальше другим нейронам. Эта связь между нейронами позволяет обрабатывать и передавать информацию по всему мозгу и нервной системе. Сложность и разнообразие нейронных связей и сетей позволяют мозгу выполнять широкий спектр функций, включая восприятие, память и поведение.

Аналогично биологической нейронной сети, искусственная нейронная сеть состоит из большого количества взаимосвязанных узлов, называемых искусственными нейронами, которые обрабатывают входные данные и вырабатывают выходные сигналы.

Базовая архитектура нейронной сети состоит из трех типов слоев: входного слоя, скрытого слоя и выходного слоя. Входной уровень является первым уровнем в нейронной сети и отвечает за прием входных данных, которые необходимо обработать. Количество нейронов во входном слое соответствует количеству входных объектов в данных. Например, в задаче классификации изображений входной слой будет иметь нейрон для каждого пикселя изображения. Пример с классификацией изображений будет раскрываться позднее.

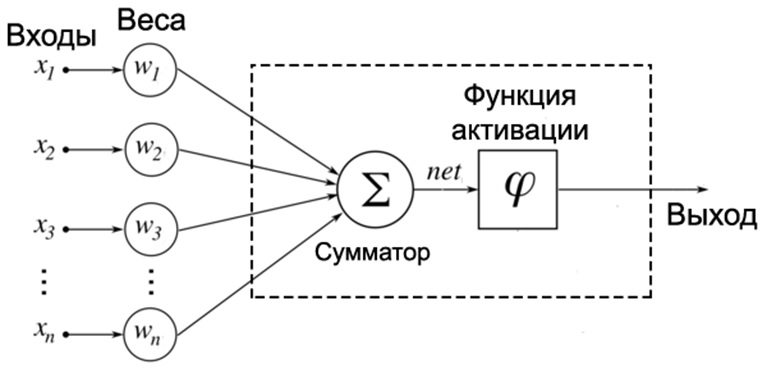
Скрытые слои нейронной сети - это то место, где происходит большая часть вычислений. Каждый скрытый слой состоит из набора нейронов, которые получают входные данные от предыдущего слоя и применяют ряд математических операций для преобразования входных данных. Эти преобразования определяются весами и смещениями, связанными с каждым нейроном и соединением. Количество скрытых слоев и нейронов в каждом слое зависит от сложности задачи и объема данных, доступных для обучения.

Скрытые уровни называются "скрытыми", потому что их выходные данные непосредственно не наблюдаемы, в отличие от входов и выходов сети. Выходные данные скрытых слоев передаются следующему слою до тех пор, пока не будет создан последний слой, который называется выходным. Выходной слой создает конечный результат работы сети, который может быть меткой класса, распределением вероятностей или числовым значением, в зависимости от задачи.

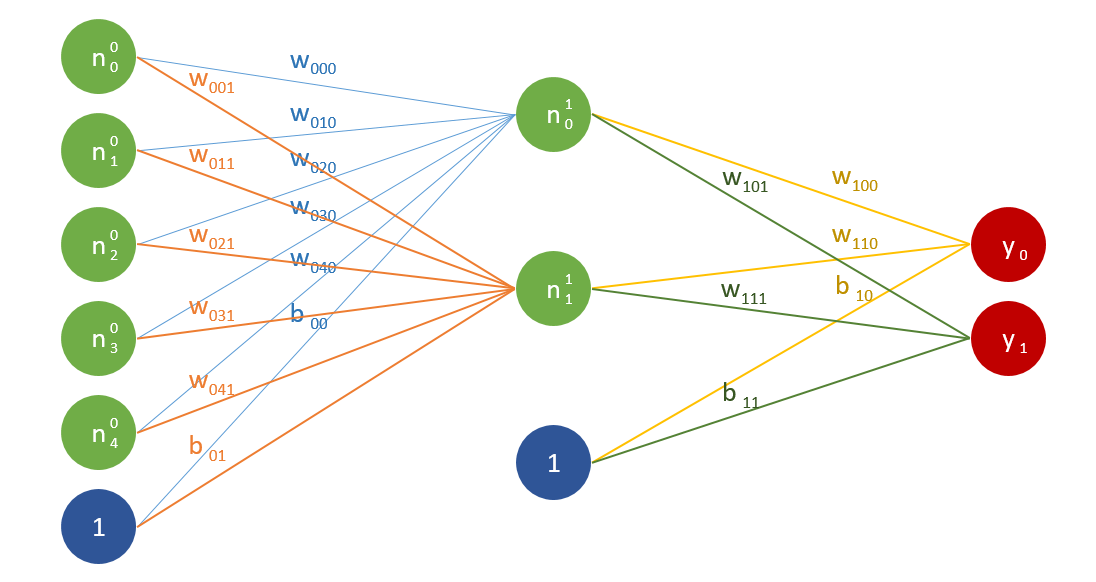
Количество нейронов в выходном слое зависит от типа выходных данных, необходимых для выполнения задачи. Например, в задаче двоичной классификации выходной уровень будет иметь один нейрон, который выдает значение от 0 до 1, указывающее на вероятность принадлежности входных данных к классу. В задаче многоклассовой классификации выходной слой будет содержать несколько нейронов, каждый из которых соответствует отдельному классу. Например, в задаче распознавания цифр десятичной системы счисления, на выходном слое будет 10 нейронов: по одному на каждую цифру.

Преобразования, которые происходят в скрытых слоях, имеют решающее значение для того, чтобы сеть научилась распознавать закономерности во входных данных и производить точные прогнозы. Связи между нейронами в нейронной сети моделируются по образцу синапсов в биологической нейронной сети, где электрические импульсы передаются между нейронами. В нейронной сети каждое соединение между нейронами имеет связанный с ним вес, который определяет прочность соединения (в некоторых источниках насчет этого проводят параллели с биологической нейронной сетью: нейрон передает импульс связанным с ними нейронам с разной силой. Эту силу как раз и характеризует вес). Веса корректируются в процессе обучения, чтобы повысить производительность сети. Процесс корректировки весов называется обучением нейронной сети.

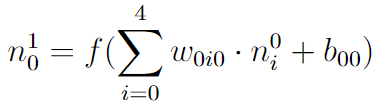
Теперь опишем понятие функции активации. Функция активации каждого нейрона определяет, будет ли он "срабатывать" и выдавать выходной сигнал, основываясь на получаемых им входных данных. Функции активации применяются к выходному сигналу каждого нейрона после того, как он был умножен на соответствующий ему вес и суммирован с членом смещения. Результирующее значение затем передается через функцию активации, которая выдает конечный выходной сигнал нейрона.

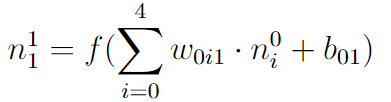
Схематично, описываемая математическая модель нейрона выглядит так: 

Приведем пример простой нейронной сети. Она будет состоять из одного входного слоя, одного выходного слоя и одного скрытого слоя. Раньше такая модель называлась бы трехслойной, однако сейчас принято считать только число скрытых слоев. Итак, имеем однослойную нейронную сеть. Пусть на входном слое 6 нейронов: 5 “классических” и 1 нейрон смещения (константа, как правило, равная 1). На скрытом слое, положим, 3 нейрона: 2 обычных и 1 нейрон смещения. На выходном слое 2 нейрона. Ниже схематически показана подобная сеть, где у каждого нейрона есть пара индексов: нижний показывает условный порядковый номер нейрона, а верхний – номер слоя, которому данный нейрон принадлежит. Каждый нейрон следующего слоя связан со всеми нейронами предыдущего слоя. Каждая такая связь имеет свой вес, который условно обозначим где i отвечает за номер исходящего слоя, j – за номер исходящего нейрона, k – за номер входящего нейрона.

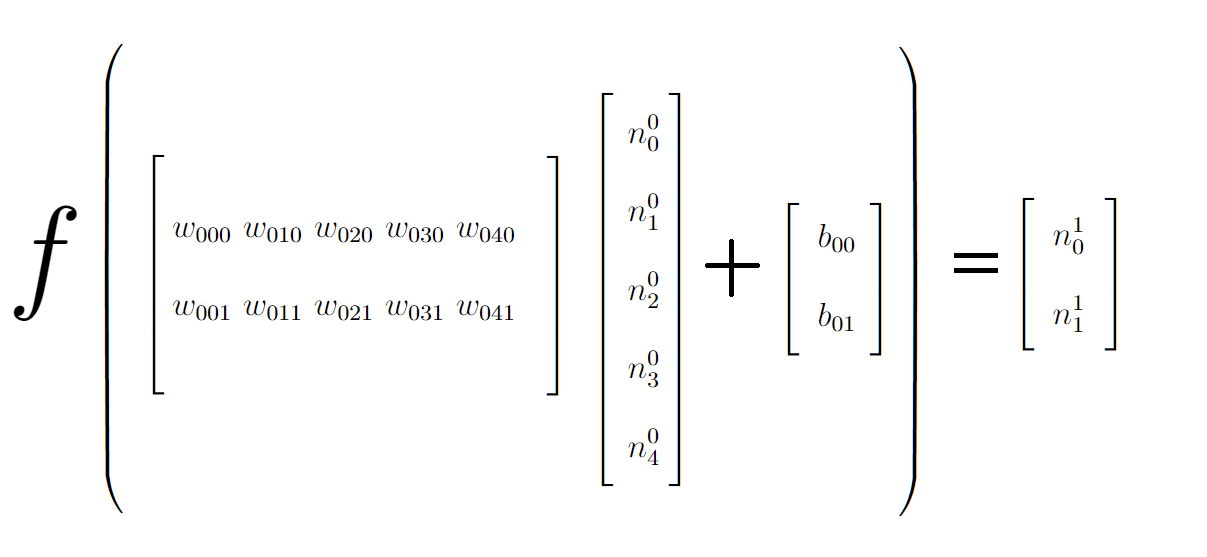


Математическая запись нейронов второго уровня будет иметь следующий вид:

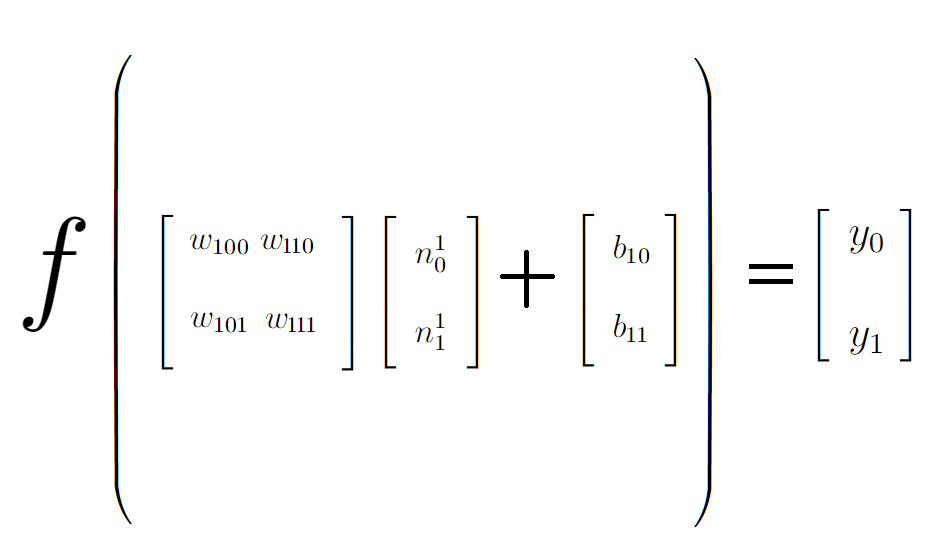




В терминах линейной алгебры запись выглядит так:



Аналогично для нейронов выходного слоя:



Вышеописанные математические выкладки называются прямой трансляцией или feedforward или алгоритмом прямого распространения.

Теперь посмотрим, что будет, если взять в качестве функции f активации линейную функцию. В таком случае, очевидно, получим суперпозицию линейных преобразований между входным слоем и выходным, что, в свою очередь, лишает смысла наличие скрытого слоя, ведь можно обойтись только упомянутыми двумя. Поэтому в качестве функции активации принято использовать нелинейные функции, такие как следующие:

**Сигмоид**: Сигмоидальная функция выдает выходные данные в диапазоне от 0 до 1, что полезно для задач двоичной классификации. Функция определяется как

f(x) = 1 / (1 + e^-x).

**ReLU**: Функция ReLU выдает значение 0 для отрицательных входных значений и само входное значение для положительных входных значений. Функция определяется как

f(x) = max(0, x).

**Softmax**: Функция определяется как

f(x) = e ^ x\_i / sum(e ^ x\_j),

где i - индекс текущего нейрона, а j - индекс всех нейронов в выходном слое.

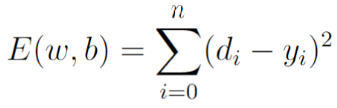
**Обучение нейронной сети**

Обучение нейронной сети заключается в минимизации отклонения выходных результатов от «правильных» результатов. Если нейросеть обучается на наборе из входных и заведомо известных «правильных» выходных данных, такой процесс обучения называют обучение с учителем или контролируемый процесс обучения. Помимо этого, существуют другие подходы в обучении нейросетей. Например:

* Неконтролируемое обучение: при неконтролируемом обучении нейронная сеть обучается на данных без каких-либо явных меток или указаний. Вместо этого сеть пытается самостоятельно найти закономерности и структуру во входных данных.
* Обучение с подкреплением: при обучении с подкреплением нейронная сеть обучается, получая обратную связь в виде вознаграждений или наказаний за определенные действия, предпринятые в окружающей среде. Сеть пытается выработать стратегию, которая максимизирует ожидаемое вознаграждение с течением времени.
* Трансфертное обучение: это метод, при котором предварительно обученная сеть настраивается под новую задачу с меньшим объемом данных, что может ускорить обучение.

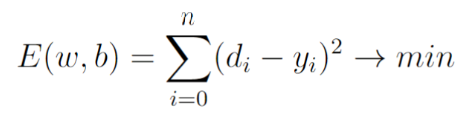
Отклонение полученных результатов от идеальных характеризует функция ошибки.

Она имеет следующий вид:



Где n – это число выходных нейронов – 1.

Задача обучения нейронной сети сводится к задаче поиска минимума функции ошибки, то есть к задаче вида:



Эта задача может быть решена с помощью метода градиентного спуска

(описание метода градиентного спуска, геометрический смысл)

(Пример кода нейросети)