УДК 004

ТЕХНОЛОГИИ РАБОТЫ НЕЙРОННЫХ СЕТЕЙ

М.А. ИВАНОВ
(Представлено: канд. техн. наук, доц. А.Ф. ОСЬКИН)

Рассматривается принцип построения и работы нейронных сетей. Представлены поняиия «нейронная сеть», «нейрон», «синапс».

Многие материалы по нейронным сетям сразу начинаются с демонстрации довольно сложных архитектур. При этом самые базовые вещи, касающиеся функций активаций, инициализации весов, выбора количества слоев в сети и т.д., если и рассматриваются, то вскользь. Получается, начинающему практику нейронных сетей приходится брать типовые конфигурации и работать с ними фактически вслепую.

В данной статье рассмотрим механизм работы нейронных сетей, начиная с основ, с самой простой конфигурации. Это позволит наработать практическую интуицию в построении архитектур нейросетей, которая на практике оказывается очень ценным активом [1].

Нейронная сеть – это последовательность нейронов, соединенных между собой синапсами. Схематическое представление нейронов показано на рисунке 1. 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. Другими словами, нейросеть – это машинная интерпретация мозга человека, в котором находятся миллионы нейронов, передающих информацию в виде электрических импульсов.

Базовый тип нейронных сетей – это сеть прямого распространения (далее СПР). Это сеть с последовательным соединением нейронных слоев, в ней информация всегда идет только в одном направлении.

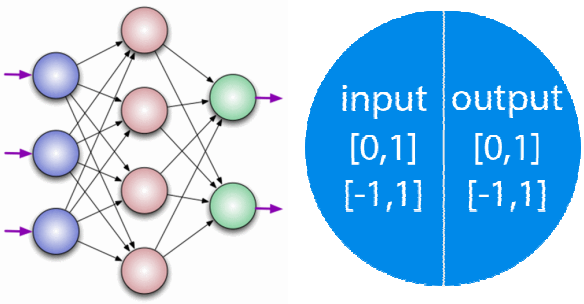

Рисунок 1. – Схематическое представление нейронов

Нейрон – это вычислительная единица, которая получает информацию, производит над ней простые вычисления и передает ее дальше. Они делятся на три основных типа: входной, скрытый и выходной (рис. 2). 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 (inputdata) и выходные данные (outputdata). 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 и попадает в поле output.

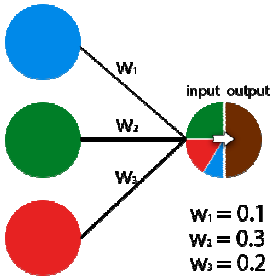

Рисунок 2. – Схематическое представление синапсиса

Синапс – это связь между двумя нейронами. У синапсов есть 1 параметр – вес. Благодаря чему входная информация изменяется, когда передается от одного нейрона к другому. К примеру, есть 3 нейрона, которые передают информацию следующему. В таком случае, существуют 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 [1].

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

В примере на рисунке 3 изображена часть нейронной сети, где буквами I обозначены входные нейроны, буквой H – скрытый нейрон, а буквой w – веса.

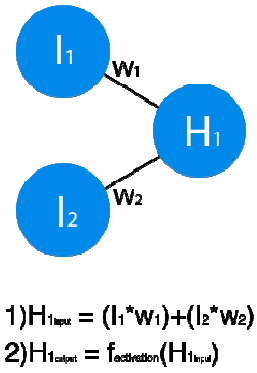

Рисунок 3. – Пример части нейронной сети

Из формулы видно, что входная информация – это сумма всех входных данных, умноженных на соответствующие им веса. Тогда дадим на вход 1 и 0. Пусть $w_1 = 0{,}4$ и $w_2 = 0{,}7$

Входные данные нейрона H_1 будут следующими: 1x0,4 + 0x0,7 = 0,4. Теперь, когда существуют входные данные, есть возможность получить выходные данные, подставив входное значение в функцию активации. Выходные данные передаются дальше. Так повторяется для всех слоев до выходного нейрона. Запустив такую сеть в первый раз можно увидеть, что ответ далек от правильного, потому что сеть не натренирована. Для улучшения результатов сеть необходимо тренировать.

Функция активации – это способ нормализации входных данных. Если на входе будет большое число, пропустив его через функцию активации, на выходе число попадает в нужный диапазон. Функций активации достаточно много, имеет смысл рассмотреть самые основные: Сигмоид (Логистическая) и Гиперболический тангенс. Главные их отличия – это диапазон значений.

Логистическая функция (рис. 4) – самая распространенная функция активации, ее диапазон значений [0, 1]. Именно на ней показано большинство примеров в сети. Соответственно, если в конкретном случае присутствуют отрицательные значения (например, акции могут идти не только вверх, но и вниз), понадобится функция, которая захватывает и отрицательные значения.

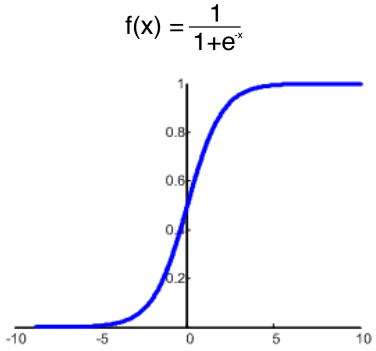

Рисунок 4. – Логистическая функция активации

Имеет смысл использовать гиперболический тангенс (рис. 5), только тогда, когда значения могут быть и отрицательными, и положительными, так как диапазон функции [–1; 1]. Использовать эту функ-

цию только с положительными значениями нецелесообразно, так как это значительно ухудшит результаты нейросети [3].

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

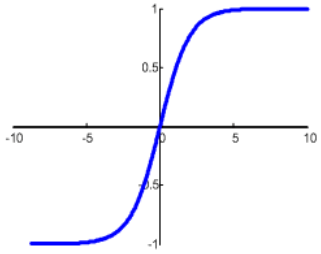

Рисунок 5. – Гиперболический тангенс

Тренировочный сет – это последовательность данных, которыми оперирует нейронная сеть.

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 – это величина, которая устанавливается в 0 при инициализации нейронной сети и имеет потолок, задаваемый вручную. Чем больше эпоха, тем лучше натренирована сеть и соответственно, ее результат. Эпоха увеличивается каждый раз, когда проходит весь набор тренировочных сетов.

Важно не путать итерацию с эпохой и понимать последовательность их инкремента. Сначала n раз увеличивается итерация, а потом эпоха и никак не наоборот. Нельзя сначала тренировать нейросеть только на одном сете, потом на другом и т.д. Нужно тренировать каждый сет один раз за эпоху. Данным способомможно избежать ошибок в вычислениях.

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, нейросеть спроектирована не лучшим образом.

Принцип подсчета ошибки во всех случаях одинаков. За каждый сет от идеального ответа отнимается полученный. Далее, либо возводится в квадрат, либо вычисляется квадратный тангенс из этой разности, после чего полученное число делится на количество сетов [2].

Заключение

Нейронные сети – это мощный, но при этом нетривиальный прикладной инструмент. Лучший способ научиться строить рабочие нейросетевые конфигурации – начинать с более простых моделей и экспериментировать, нарабатывая опыт и интуицию практики нейронных сетей.

ЛИТЕРАТУРА

1. Искусственная_нейронная_сеть // Википедия – свободная общедоступная мультиязычная универсальная интернет-энциклопедия [Электронный ресурс]. – Режим доступа: https://ru.wikipedia.org/wiki/Искусственная_нейронная_сеть/. – Дата доступа: 24.09.2017.
2. Нейронные сети в картинках: от одного нейрона до глубоких архитектур [Электронный ресурс] // Хабрахабр. – Режим доступа: https://habrahabr.ru/post/322438/. – Дата доступа: 25.09.2017.
3. Самое главное о нейронных сетях. Лекция в Яндексе [Электронный ресурс] // Хабрахабр. – Режим доступа: https://habrahabr.ru/company/yandex/blog/307260/. – Дата доступа: 25.09.2017.