Лекция 2

Тема 1.1 Классификация нейронных сетей и их свойства

Содержание:

1. Функции активации нейронов

2. Процесс обучения нейронных сетей

Сигнал NET далее, как правило, преобразуется активационной функцией *f* и дает выходной нейронный сигнал OUT. Активационная функция может быть обычной линейной функцией:

$$\text{OUT} = K(\text{NET}), \qquad (1)$$

где *K* – постоянная, пороговой функции

$$\text{OUT} = 1, \text{ если NET} > \text{T}, \qquad (2)$$

$$\text{OUT} = 0 \text{ в остальных случаях, (или -1)} \qquad (3)$$

где *T* – некоторая постоянная пороговая величина, или же функцией, более точно моделирующей нелинейную передаточную характеристику биологического нейрона и представляющей нейронной сети большие возможности.

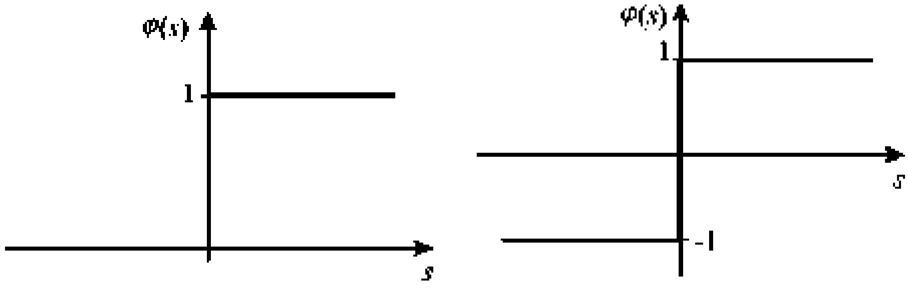

Рис.2. Пороговая функция активации

Функция активации f сужает диапазон изменения величины NET так, что при любых значениях NET значения OUT принадлежат некоторому конечному интервалу, то *F* называется *"сжимающей" функцией.* Наиболее распространенные функции активации приведены в табл. 1.

В качестве "сжимающей" функции часто используется логистическая или "сигмоидальная" (S-образная) функция. Эта функция математически выражается как $F(x) = 1/(1 + e^{-x})$, или описывается в виде:

$$\mathrm{OUT} = \frac{1}{1+e^{-\mathrm{NET}}}. \tag{4}$$

По аналогии с электронными системами активационную функцию можно считать нелинейной усилительной характеристикой искусственного нейрона. Коэффициент усиления вычисляется как отношение приращения величины OUT к вызвавшему его небольшому приращению величины NET. Он выражается наклоном кривой при определенном уровне возбуждения и изменяется от малых значений при больших отрицательных возбуждениях (кривая почти горизонтальна) до максимального значения при нулевом возбуждении и снова уменьшается, когда возбуждение становится большим положительным. Гроссберг (1973) обнаружил, что подобная нелинейная характеристика решает поставленную им дилемму шумового насыщения. Каким образом одна и та же сеть может обрабатывать как слабые, так и сильные сигналы? Слабые сигналы нуждаются в большом сетевом усилении, чтобы дать пригодный к использованию выходной сигнал. Однако усилительные каскады с большими коэффициентами усиления могут привести к насыщению выхода шумами усилителей (случайными флуктуациями), которые присутствуют в любой физически реализованной сети.

Таблица 1 – Функции активации нейронов

№	Название функции активации	Формула	Область значений

1	Пороговая бинарная	$f(s)=\begin{cases}0, s\le\theta\\1, s>\theta\end{cases}$	(0,1)
2	Сигнум (модифицированная пороговая)	$f(s)=\begin{cases}-1, s\le\theta\\1, s>\theta\\0, s=0\end{cases}$	(-1,1)
3	Линейная	$f(s)=k\cdot s$	(-∞, ∞)
4	Полулинейная		(0, ∞)
5	Логистическая (сигмоидальная)	$f(s)=\frac{1}{1+e^{-as}}$	(0,1)
6	Гиперболический тангенс	$f(s)=\frac{e^{as}-e^{-as}}{e^{as}+e^{-as}}$	(-1,1)
7	Экспоненциальная	$f(s)=e^{as}$	(0, ∞)
8	Синусоидальная	$f(s)=\sin(s)$	(-1,1)
9	Сигмоидальная (рациональная)	$f(s)=\frac{s}{a+\lvert s\rvert}$	(-1,1)
10	Шаговая (линейная с насыщением)	$f(s)=\begin{cases}-1, s\le 0\\s, -1<s<1\\s\ge 1\end{cases}$	(-1,1)
11	Модульная	$f(s)=\lvert s\rvert$	(0, ∞)
12	Знаковая (сигнатурная)	$f(s)=\begin{cases}-1, s>0\\1, s\le 0\end{cases}$	(-1,1)
13	Квадратическая	$f(s)=s^2$	(0, ∞)

Сильные входные сигналы в свою очередь также будут приводить к насыщению усилительных каскадов, исключая возможность полезного использования выхода.

Центральная область логистической функции, имеющая большой коэффициент усиления, решает проблему обработки слабых сигналов, в то время как области с падающим усилением на положительном и отрицательном концах подходят для больших возбуждений. Таким образом, нейрон функционирует с большим усилением в широком диапазоне уровня входного сигнала.

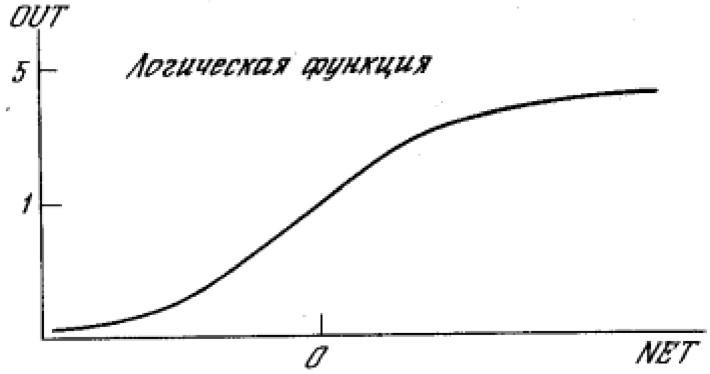

Рис. 3. Логистическая функция активации

Другой широко используемой активационной функцией является гиперболический тангенс. По форме она сходна с логистической функцией и часто используется биологами в качестве математической модели активации нервной клетки. В качестве активационной функции искусственной нейронной сети она записывается следующим образом:

$$OUT = th(x). \qquad (5)$$

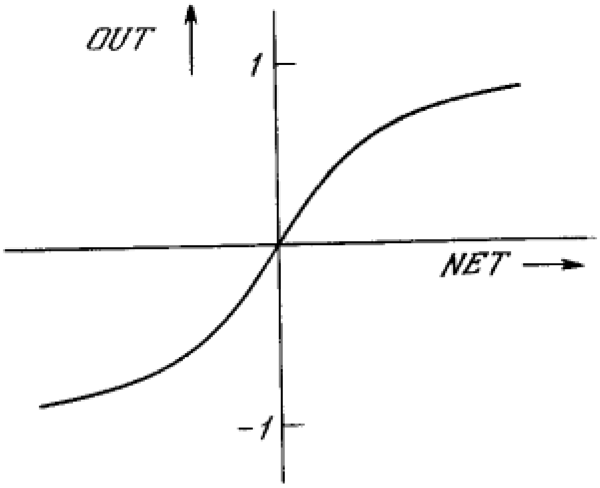

Рис.4. Гиперболический тангенс

Подобно логистической функции гиперболический тангенс является S-образной функцией, но он симметричен относительно начала координат, и в точке NET = 0 значение выходного сигнала OUT равно нулю. В отличие от логистической функции гиперболический тангенс принимает значения различных знаков, что оказывается выгодным для ряда сетей.

Рассмотренная простая модель искусственного нейрона игнорирует многие свойства своего биологического двойника. Например, она не принимает во внимание задержки во времени, которые воздействуют на динамику системы. Входные сигналы сразу же порождают выходной сигнал. И, что более важно, она не учитывает воздействий функции частотной модуляции или синхронизирующей функции биологического нейрона, которые ряд исследователей считают решающими.

Несмотря на эти ограничения, сети, построенные из этих нейронов, обнаруживают свойства, сильно напоминающие биологическую систему. Только время и исследования смогут ответить на вопрос, являются ли подобные совпадения случайными или следствием того, что в модели верно схвачены важнейшие черты биологического нейрона.

ИНС представляет собой совокупность нейронов, определенным образом соединенных друг с другом. Нейронные сети можно классифицировать по различным характеристикам: по типу связей между нейронами, топологии, типу обрабатываемых данных, направлению обработки информации и др.

По типу исходной обрабатываемой информации можно выделить сети аналогового типа и дискретные. Сети первого типа используют в качестве исходных данных действительные числа. Сети второго типа работают с данными, представленными в двоичной системе счисления.

Нейронные сети могут обучаться с учителем и без учителя. При обучении с учителем для каждого входного примера известно выходное пространство решений нейронной сети. Такое пространстве решений связано с ожидаемыми результатами работы сети. Для этого при обучении используют обучающие выборки, позволяющие сравнивать полученные результаты с ожидаемыми. На основе определенного алгоритма обучения производится корректировка весовых коэффициентов, для уменьшения ошибки обучения. При отсутствии таких выборок сети обучаются без учителя. В этом случае сеть формирует выходное пространство решений только на основе векторов входа.

Процесс обучения без учителя производится только на основании распределения входных данных, без вмешательства учителя в процесс обучения. Входные данные распределяются на категории, или классы, и количество классов заранее не известно.

По признаку корректировки синаптических коэффициентов, выполняющих роль связей между нейронами, нейронные сети можно разделить на два класса: статические сети и динамические сети.

В статических сетях, связи между нейронами фиксированы и весовые коэффициенты сети выбираются сразу из условия задачи. В сетях с динамическими связями, для которых настройка связей осуществляется в процессе обучения. Это осуществляется посредством использования обратных связей между нейронами отдельных слоев. В том случае, если в сети существует хотя бы одна связь она называется рекуррентной.

По направлению информации нейронные сети делятся на сети:

- прямой передачи сигнала (feed-forward), в которых информация распространяется последовательно от слоя к слою в прямом направлении, т.е. от выходов к выходам;

- с обратным распространением информации (feed-back), характеризующиеся как прямым, так и обратным распространением данных между слоями сети.

По признаку структур нейронов в сетях ИНС различают на гомогенные и гетерогенные. В гомогенных сетях нейроны для преобразования информации используют одну функцию активации, в гетерогенной сети нейроны разных слоев используют различные функции активации.

Известно большое разнообразие архитектур нейронных сетей, в которой нейроны соединены между собой определенным образом. По характеру связей между нейронами в сети, т.е. топологии нейронные сети делятся на:

- полносвязные;

- слоистые (многослойные);

- слабосвязанные (нейронные сети с локальными связями).

Цель обучения ИНС заключается в следующем: для множества входных сигналов выдавать желаемое значение или близкое к нему с заданной погрешностью. Различают несколько десятков алгоритмов обучения нейронных сетей, которые можно разделить на два класса:

- обучение с учителем (supervised learning);
- обучение без учителя (unsupervised learning).

При обучении с учителем предполагается, что помимо входных сигналов x_i, известны также и ожидаемые выходные сигналы d_i. Для этого формируется база данных, в которой каждый входной образ описывается набором параметров $(x_1, \ldots. x_n)$. Для обучения из общей базы данных выбираются обучающие примеры (training samples), для которых заранее известен желаемый результат или отклик сети. Исследователь самостоятельно формирует *обучающую выборку или привлекает экспертов*, т.е. совокупность множества пар «вход-выход» $(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots. (x_n, y_n)$.

При обучении ИНС каким-либо алгоритмом, сеть изменяет синаптические веса таким образом, чтобы расхождения от полученного (y_i) выходного сигнала с ожидаемыми (d_i) были минимальными по заданному критерию. Обучение продолжается до тех пор, пока изменения синаптических весов не станут незначительными. Обученная нейронная сеть может быть использована для решения конкретных задач. Множество входных или выходных сигналов рассматривается как векторы и обучение производится путем последовательного предъявления входных векторов с одновременной подстройкой весов. Таким образом, в алгоритмах с учителем необходима обучающая выборка с метками классов.

Поэтому процесс обучения нейронных сетей можно рассматривать как совокупность пяти компонентов:

$$< x, y, r, G, E >, \tag{6}$$

где x – множество входных сигналов;

y – множество выходных значений сети;

R – некоторая функция, для которой *y=r(x)*, задаваемая парами входов-выходов в виде $(x_1, y_1)\ldots\ldots(x_p, y_p)$;

G – множество всех возможных значений функции g при выбранной архитектуре сети, т.е. вид функции g определяется значениями синаптических весов и смещений сети;

E – функция ошибки или функционал качества обучения сети, отражает для каждой из функций g степень близости к *r*.

В процессе обучения нейронная сеть формирует выходной сигнал *Y* в соответствии с заданным входным сигналом *X*, формируя некоторую функцию *G(x)=y*. Процесс обучения заключается в поиске функции *G* близкой к *r* с заданной величиной ошибки.

Например, в задачах распознавания образов исследователь имеет некоторое множество входных данных, или образов (*x*). Под образом в общем случае понимают цифровое изображение (аэрофотоснимок, видеоснимок, космический снимок, номеров автомобилей, отпечатков пальцев, текста и т.д.). При обучении нейронных сетей исходный образ должен быть отнесен к определенному классу объектов (например, класс земель, буква алфавита и т.д.).

При использовании множества обучающих примеров процесс обучения превращается в задачу многомерной оптимизации. В общем случае обучение сети заключается в поиске такой функции g, оптимальной по E. Это требует длительных вычислений, число итераций при обучении сетей может составлять от 10^3 до 10^8 числа итераций.

Обучение без учителя. Основная черта алгоритмов без учителя – это отсутствие обучающих данных о принадлежности входных данных конкретному классу (группе) объектов. Обучение ИНС производится на основании конкуренции нейронов согласно стратегии: победитель получает все (Winner Takes All, WTA); победитель получает большее (Winner Takes Most, WTM).

Все алгоритмы обучения ИНС делят:

- локальной оптимизации с вычислением частных производных первого порядка;

- локальной оптимизации с вычислением частных производных первого и второго порядков;

- стохастической оптимизации;

- глобальной оптимизации