Принципы построения нейронных сетей [M.163]

Функционирование нейронной сети определяется не только тем, какие модели нейронов в ней используются и видом их активационных функций, но и количеством нейронов, а также способом их соединения внутри сети. Параметры выбираются исходя из особенностей решаемой с помощью нейронной сети задачи, уровня ее сложности и трудоемкости, количества используемых переменных, характера исходных данных.

При реализации нейронной сети и определении ее параметров необходимо выбрать нейросетевую архитектуру и конфигурацию. При этом под *архитектурой* понимаются общие принципы построения нейронных сетей для определенного класса задач, а под *конфигурацией* — параметры конкретной нейросетевой модели. Если решаемая задача является типовой и хорошо известной, что характерно для задач Data Mining, то при создании нейронной сети можно воспользоваться одной из готовых конфигураций. В научных и технических приложениях, где сложность и разнообразие задач намного выше, в некоторых случаях приходиться синтезировать нестандартные нейросетевые конфигурации, что может оказаться весьма трудоемким и длительным процессом.

Построение и обучение нейронных сетей часто называют искусством, поскольку выбор конфигурации сети и параметров ее обучения не всегда однозначен, он во многом определяется опытом и даже интуицией аналитика. Конечно, в данной области существуют теория и общие рекомендации, но сильная зависимость работы нейронных сетей от особенностей входных данных и характера искомых закономерностей делает процесс построения классификационных и регресионных моделей на их основе неоднозначным: при поиске оптимального решения приходится использовать метод проб и ошибок.

Прежде чем приступить к рассмотрению основных нейросетевых архитектур и конфигураций, а также рекомендаций по выбору соответствующих параметров, введем некоторые базовые понятия.

Основные группы нейронов в сети

В зависимости от расположения в сети и выполняемых функций все нейроны могут быть разбиты на группы.

1. Входные нейроны. На них поступают значения переменных вектора входного воздействия $\mathbf{X} = (x_1, x_2, \ldots, x_n)$, где n — число компонентов вектора, равное числу входных переменных модели; соответственно, ему же будет равно число входных нейронов. Совокупность входных нейронов образует вход нейронной сети. Входные нейроны не выполняют обработку информации: их задача — принять компоненты вектора входного воздействия и распределить по другим нейронам сети.
2. Выходные нейроны. Служат для вывода результатов обработки нейронной сетью входного вектора. На их выходах формируются компоненты выходного вектора $\mathbf{Y} = (y_1, y_2, \ldots, y_m)$, где m — число выходных нейронов, которое соответствует числу выходных переменных модели. В общем случае $m \neq n$. Выходной вектор $\mathbf{Y}$ часто называют откликом или реакцией сети на заданное входное воздействие, а сеть в процессе функционирования реализует функцию многих переменных $\mathbf{Y} = f(\mathbf{X})$.
3. Скрытые нейроны. Расположены внутри сети. Если у входных нейронов с другими нейронами сети связаны только выходы, а у выходных нейронов — только входы, то скрытыми можно назвать все нейроны, не имеющие связи с внешним окружением сети. Именно скрытые нейроны выполняют преобразование данных в нейронной сети.

Архитектуры нейронных сетей

Функциональные возможности нейронной сети возрастают с увеличением числа нейронов и связей между ними. Чем больше связей, тем больше количество весов, настраиваемых в процессе обучения нейронной сети. При увеличении числа весовых коэффициентов возрастает число возможных состояний нейросетевой модели, а значит, и количество возможных функциональных преобразований. Однако усложнение модели неизбежно приводит к росту вычислительных затрат, связанных с ее обучением и работой. Кроме того, уровень сложности задач Data Mining не требуют применения нейронных сетей с большим числом связей, поэтому при выборе конфигурации нейронной сети всегда следует искать компромисс.

Единственное жесткое требование, предъявляемое к конфигурации сети, — это соответствие размерности вектора входного воздействия числу входов нейронной сети и размерности вектора отклика числу выходных нейронов.

Искусственную нейронную сеть прямого распространения, в которой присутствует хотя бы один скрытый слой, называют *многослойным персептроном* (*multilayer perceptron*, MLP). Из всех архитектур нейронных сетей именно он с сигмоидной функцией активации является базовым для решения задач классификации и регрессии. Существуют нейронные сети, в которых сигнал проходит не только в прямом направлении, но и в обратном, в структуре их связей присутствуют замкнутые циклы, однако такие сети реже применяются в Data Mining.

Принципы функционирования многослойного персептрона

Рассмотрим сначала несложную конфигурацию многослойного персептрона, содержащую один скрытый слой (рисунок 1).

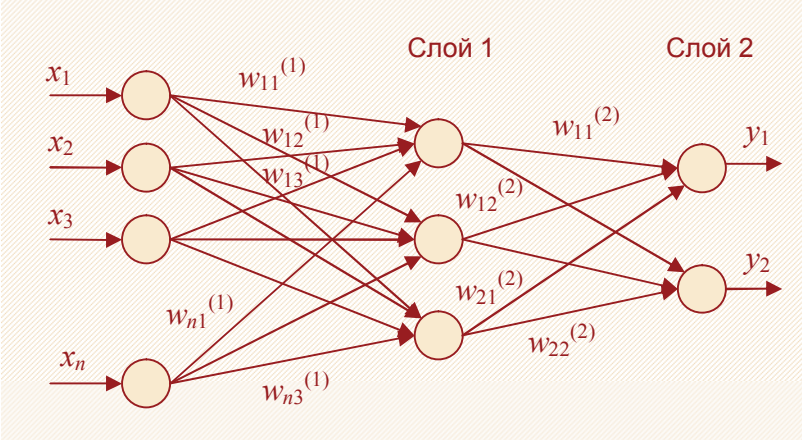

Рисунок 1 – Многослойный персептрон

Нейроны выходного слоя обрабатывают данные и выполняют вывод результата. Обозначим каждую связь внутри сети как w_{ij}^{k}, где i — номер нейрона входного слоя, с выхода которого выходит связь; j — номер нейрона выходного слоя, на вход которого поступает связь (полагаем, что нейроны в слоях нумеруются сверху вниз). Индекс k указывает на слой сети, к которому относится данная связь. Например, $w_{11}^{(1)}$ — это первая связь первого нейрона первого слоя сети.

При подаче на вход вектора $X = (x_1, x_2 \ldots x_n)$ сеть сформирует на выходе двухкомпонентный вектор $Y = (y_1, y_2)$, значения элементов которого вычисляются по формуле:

$$y_j = f\left[\sum_{i=1}^{n} x_i w_{ij}\right] \quad (1)$$

где f — активационная функция нейрона.

Рассмотрим пример, который иллюстрирует процесс формирования выхода в многослойном персептроне, содержащем три входных нейрона, три нейрона скрытого слоя и два нейрона выходного слоя (рисунок 2).

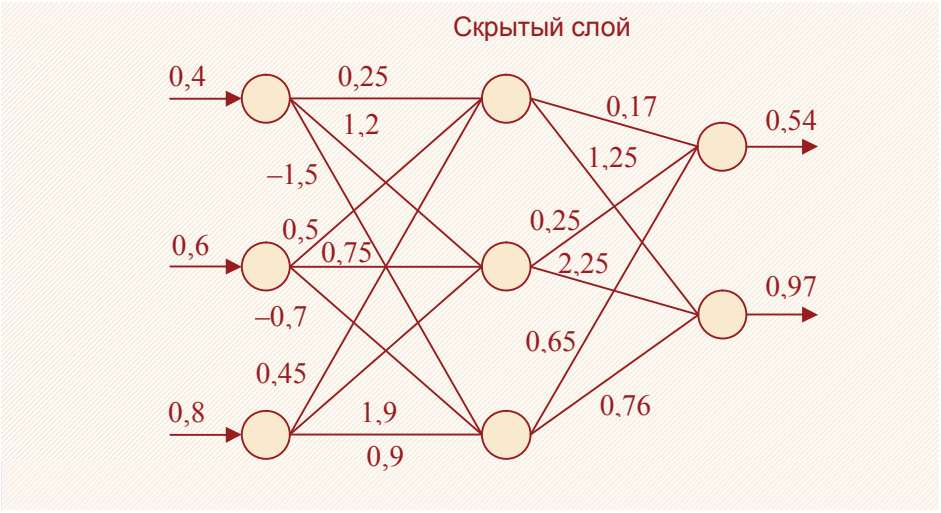

Рисунок 2 – Сеть 3 × 3 × 2

Пусть на вход нейронной сети поступает вектор $\mathbf{X} = (0{,}4;\ 0{,}6;\ 0{,}8)$, а веса связей установлены в соответствии с таблицей 1.

Таблица 1 – Начальные веса нейронов

1 слой	2 слой
$w_{11}^{(1)} = 0{,}25$	$w_{11}^{(2)} = 0{,}17$
$w_{12}^{(1)} = 1{,}2$	$w_{12}^{(2)} = 1{,}25$
$w_{13}^{(1)} = -1{,}5$	$w_{21}^{(2)} = -0{,}25$
$w_{21}^{(1)} = 0{,}5$	$w_{22}^{(2)} = 2{,}25$
$w_{22}^{(1)} = 0{,}75$	$w_{31}^{(2)} = 0{,}65$
$w_{23}^{(1)} = -0{,}7$	$w_{32}^{(2)} = 0{,}76$
$w_{31}^{(1)} = 0{,}45$	
$w_{32}^{(1)} = 1{,}9$	
$w_{33}^{(1)} = 0{,}9$	

На каждый нейрон первого скрытого слоя поступает один и тот же входной вектор, распределяемый нейронами входного слоя. Тогда на выходах сумматоров нейронов 1-го скрытого слоя будут получены следующие значения:

$$s_1^{(1)} = x_1 \cdot w_{11}^{(1)} + x_2 \cdot w_{21}^{(1)} + x_3 \cdot w_{31}^{(1)} = 0{,}4 \cdot 0{,}25 + 0{,}6 \cdot 1{,}2 + 0{,}8 \cdot 0{,}45 = 0{,}1 + 0{,}72 + 0{,}36 = 1{,}18$$

$$s_2^{(1)} = x_1 \cdot w_{12}^{(1)} + x_2 \cdot w_{22}^{(1)} + x_3 \cdot w_{32}^{(1)} = 0{,}4 \cdot 1{,}2 + 0{,}6 \cdot 0{,}75 + 0{,}8 \cdot 1{,}9 = 0{,}48 + 0{,}45 + 1{,}52 = 2{,}45$$

$$s_3^{(1)} = x_1 \cdot w_{13}^{(1)} + x_2 \cdot w_{23}^{(1)} + x_3 \cdot w_{33}^{(1)} = 0{,}4 \cdot (-1{,}5) + 0{,}6 \cdot (-0{,}7) + 0{,}8 \cdot 0{,}9 = -0{,}6 - 0{,}42 + 0{,}72 = -0{,}30$$

Если используется логистическая активационная функция с параметром крутизны, равным 1, то выходы нейронов 1-го слоя будут:

$$y_1^{(1)} = f\left(s_1^{(1)}\right) = 1/\left(1 + e^{1{,}18}\right) = 0{,}86$$

$$y_3^{(1)} = f\left(s_3^{(1)}\right) = 1/\left(1 + e^{0{,}3}\right) = 0{,}42$$

Таким образом, вектор входов для нейронов второго слоя составит $X_1 = (0{,}86;\ 0{,}92;\ 0{,}42)$. Тогда

$$s_1^{(2)} = x_1 \cdot w_{11}^{(2)} + x_2 \cdot w_{21}^{(2)} + x_3 \cdot w_{31}^{(2)} = 0{,}86 \cdot 0{,}17 + 0{,}92 \cdot (-0{,}25) + 0{,}42 \cdot 0{,}65 = 0{,}15 - 0{,}23 + 0{,}27 = 0{,}9$$

$$s_2^{(2)} = x_1 \cdot w_{12}^{(2)} + x_2 \cdot w_{22}^{(2)} + x_3 \cdot w_{32}^{(2)} = 0{,}86 \cdot 1{,}25 + 0{,}92 \cdot 2{,}25 + 0{,}42 \cdot 0{,}76 = 1{,}08 + 2{,}07 + 0{,}32 = 3{,}47$$

Выходы нейронной сети будут иметь следующие значения:

$$y_1^{(2)} = f\left(s_1^{(2)}\right) = 1/\left(1 + e^{-0{,}19}\right) = 0{,}54;\ y_1^{(1)} = f\left(s_2^{(2)}\right) = 1/\left(1 + e^{-3{,}47}\right) = 0{,}97.$$

Выбор числа нейронов в многослойном персептроне

При выборе конфигурации нейронной сети для решения конкретной задачи очень важно правильно определить оптимальное число нейронов. Строгих соотношений, позволяющих точно рассчитать его, не существует. Тем не менее можно привести ряд рекомендаций (отчасти эмпирических), которые помогут оценить приемлемое для решения той или иной задачи число нейронов. Вообще, интерес представляет не столько количество нейронов сети, сколько число связей между ними, так как именно настройка весов связей определяет выполняемое сетью функциональное преобразование.

Чтобы исключить переобучение нейронной сети, нужно придерживаться следующих рекомендаций.

- Число нейронов во входном и выходном слоях жестко определяется числом входных и выходных переменных модели соответственно.
- Число нейронов в скрытых слоях и число скрытых слоев выбираются таким образом, чтобы количество образованных ими связей было как минимум в два-три раза меньше числа обучающих примеров.

Рассмотрим конкретный случай. Пусть обучающее множество содержит 150 примеров. Полагая, что число связей в нейронной сети должно быть в три раза меньше числа примеров, получаем общее число связей в сети $C = 150/3 = 50$. Пусть обучающий набор данных содержит четыре входных переменных и одну выходную. Тогда нейронная сеть должна иметь четыре входных нейрона и один выходной. Рассмотрим две возможные конфигурации — с одним скрытым слоем и с двумя. В случае одного скрытого слоя можно составить соотношение $4 \cdot t + 1 \cdot t = 50$, где t — число нейронов в скрытом слое. Можно записать: $4 \cdot t + 1 \cdot t = t \cdot (4 + 1) = 5 \cdot t = 50$, откуда $t = 10$. Следовательно, для обеспечения 50 связей нейронная сеть с четырьмя входными нейронами, одним выходным и одним скрытым слоем должна содержать в нем 10 нейронов (рисунок 3*а*). Аналогично определяем, что в случае двух скрытых слоев можно использовать варианты по 5 нейронов в слое; в первом слое 7 нейронов, во втором — 3 и т. д. (рисунок 3*б*).

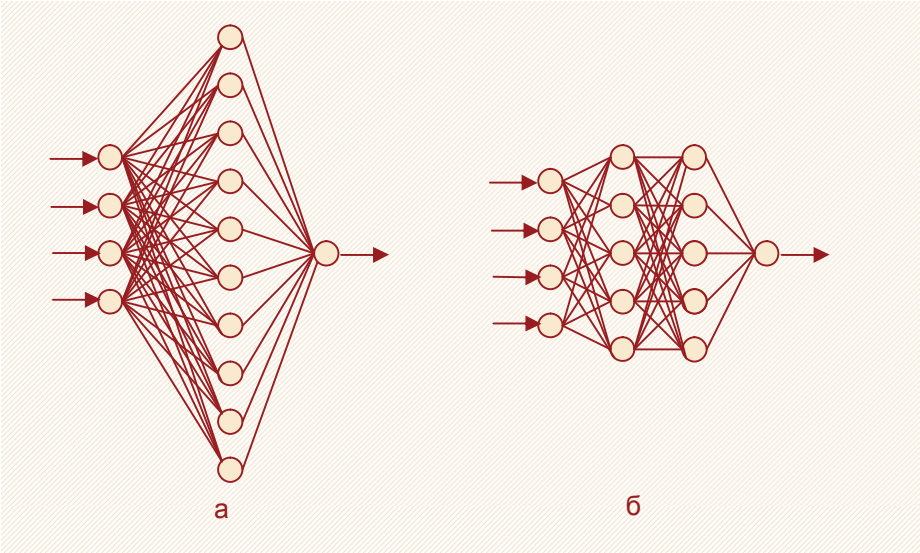

Рисунок 3 – Возможные конфигурации сетей с 50 связями