

Приближение функций с помощью нейронных сетей

В последние годы появился новый алгоритмический аппарат приближения функций многих переменных с помощью линейных операций и суперпозиций функций одного переменного. Такое приближение осуществляется специальными формальными устройствами – *нейронными сетями*, состоящими из формальных *нейронов*.

Нейрон получает на входе вектор сигналов x , вычисляет его скалярное произведение на вектор весов α и некоторую функцию одного переменного $\varphi(z)$, где z — скалярное произведение x на α . Результат рассылается на входы других нейронов или передается на выход. Таким образом, нейронные сети вычисляют суперпозиции простых функций одного переменного и их линейных комбинаций.

Для описания алгоритмов и устройств в нейро-информатике выработана специальная схмотехника, в которой элементарные устройства — сумматоры, синапсы, нейроны и т. п. — объединяются в сети, предназначенные для решения задач. Наиболее важные элементы нейросистем — *адаптивный сумматор* и *нелинейный преобразователь*. Адаптивный сумматор вычисляет скалярное произведение входного сигнала x на вектор параметров α (рис. 2.3).

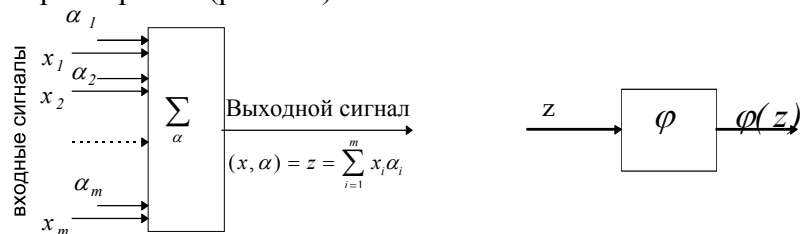


Рис. 2.3 – Адаптивный сумматор
Рис. 2.4 – Нелинейный преобразователь сигнала

Адаптивным его называют из-за наличия вектора настраиваемых параметров α . Нелинейный преобразователь получает скалярный входной сигнал z и переводит его в $\varphi(z)$ (рис. 2.4).

Стандартный формальный нейрон составлен из входного сумматора, нелинейного преобразователя и точки ветвления (рис. 2.5).

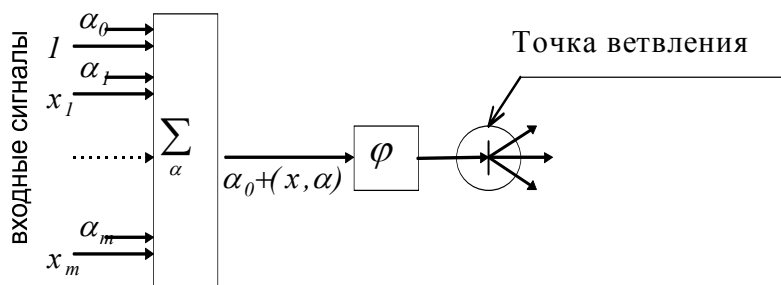


Рис. 2.5 – Формальный нейрон

Точка ветвления служит для рассылки одного сигнала по нескольким адресам. Она получает скалярный входной сигнал z и передает его выходам. Среди нейронных сетей можно выделить две базовые архитектуры: *слоистые* и *полносвязные* сети.

Слоистые сети. Нейроны расположены в несколько слоев (рис. 2.6). Нейроны первого слоя получают входные сигналы, преобразуют их и через точки ветвления передают нейронам второго слоя. Далее срабатывает второй слой и т. д. до k -го слоя, который выдает выходные сигналы для пользователя. Если не оговорено противное, то каждый

выходной сигнал i -го слоя подается на вход всех нейронов $(i+1)$ -го слоя. Число нейронов в каждом слое может быть любым и никак заранее не связано с количеством нейронов в других слоях. Стандартный способ подачи входных сигналов: все нейроны первого слоя получают каждый входной сигнал. Особое распространение получили трехслойные сети, в которых каждый слой имеет свое наименование: первый — входной, второй — скрытый, третий — выходной.

Полносвязные сети. Каждый нейрон передает свой выходной сигнал остальным нейронам, включая самого себя. Выходными сигналами сети могут быть все или некоторые выходные сигналы нейронов после нескольких тактов функционирования сети. Все выходные сигналы подаются всем нейронам.

Таким образом, нейронные сети вычисляют линейные функции, нелинейные функции одного переменного, а также все возможные суперпозиции — функции от функций, получаемые при каскадном соединении сетей.

Рассмотрим более подробно слоистую сеть (рис. 2.6). Ее структура характеризуется числом K и количеством нейронов m в каждом слое. Заметим, что в слоистой сети связи между нейронами в слое отсутствуют.

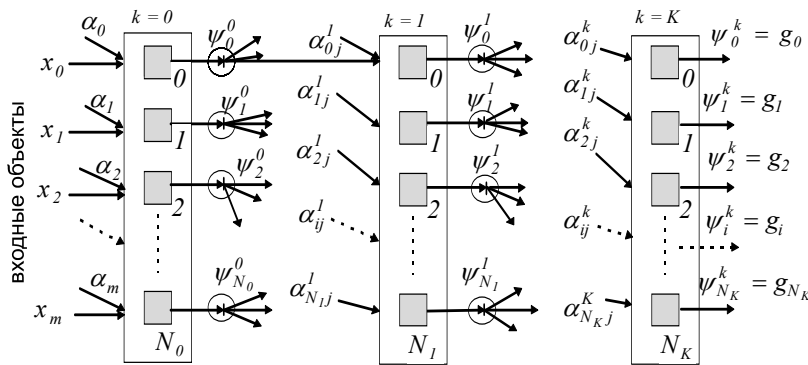


Рис. 2.6 – Слоистая сеть

Введем новые обозначения: вход i -го нейрона k -го слоя — z_i^k , выход i -го нейрона — ψ_i^k , количество нейронов в k -ом слое — $N_k, k = 1, 2, \dots, K$. Тогда суперпозиция входных сигналов i -го нейрона имеет вид:

$$z_i^k = \sum_{j=0}^{N_{k-1}} \alpha_{ij}^k \cdot \psi_j^{k-1}, i = \overline{1, N_k}, k = \overline{1, K}.$$

Здесь α_{ij}^k — весовые коэффициенты, являющиеся настраиваемыми параметрами и характеризующими связь j -го нейрона $(k-1)$ -го слоя с i -ым нейроном k -го слоя.

Для нулевого слоя имеем $\psi_j^0 = x_j, j = \overline{1, m}$. С учетом принятых обозначений аппроксимирующая функция $g_i, i = \overline{1, N_k}$, представляет собой персептрон и может быть записана в виде

$$g_i = \psi_i^k, i = \overline{1, N_k},$$

$$\psi_i^k = \varphi(z_i^k), i = \overline{1, N_k}, k = \overline{1, K},$$

$$\psi_0^k = 1, k = \overline{0, K-1}.$$

В качестве функций активации нейронов (нелинейного преобразователя нейронов φ) часто используют гладкие функции вида:

$$\varphi(z) = z; \varphi(z) = \frac{1}{1 - \exp(-z)}; \varphi(z) = \frac{\exp(z) - \exp(-z)}{\exp(z) + \exp(-z)}.$$

Приближение функций с помощью нейронных сетей сводится к их обучению. При этом входные сигналы x подаются обучаемой сети на обработку, задаются значения весовых коэффициентов α , а получаемые выходные сигналы g сравниваются с

экспериментальными данными y . Затем строится оценка работы сети, например, как критерий максимального правдоподобия

$$E(\alpha) = \frac{1}{2} \sum_{\lambda=1}^P \sum_{i=1}^{N_K} (y_{i\lambda} - g_i(x^{(\lambda)}, \alpha))^2,$$

где $g_i(x^{(\lambda)}, \alpha)$ — i -ый выход сети, соответствующий векторам входных сигналов $x^{(\lambda)}$ и весовых коэффициентов α , P — объем обучающей выборки $(x^{(\lambda)}, y_\lambda)$.

Поиск оптимальных значений весовых коэффициентов α , при которых критерий $E(\alpha)$ минимален, производится с помощью известных методов решения экстремальных задач.

При обучении нейронных сетей целесообразно использовать метод регуляризации, позволяющий получить сглаженные функции $g_i(x^{(\lambda)}, \alpha)$. При этом оценка работы сети выбирается в виде

$$\hat{E}(\beta, \alpha) = E(\alpha) + \beta \Omega(\alpha),$$

где β — параметр регуляризации, $\Omega(\alpha)$ — равномерно выпуклая функция, например, $\Omega(\alpha) = \frac{1}{2} \alpha^T \cdot \alpha$.

Оптимальное значение параметра регуляризации β подбирается итерационным методом.