Искусственные нейронные сети:
Другие виды активационных функций:

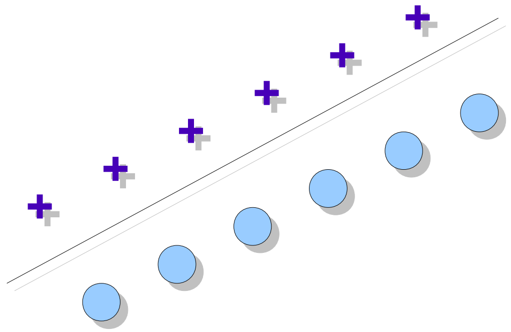
- Радиально-базисные ИНС
- Стохастическая модель нейрона

- Осовский С. **Нейронные сети для обработки информации** / Пер. с польского И.Д. Рудинского - М.: Финансы и статистика, 2002

Радиально-базисные ИНС

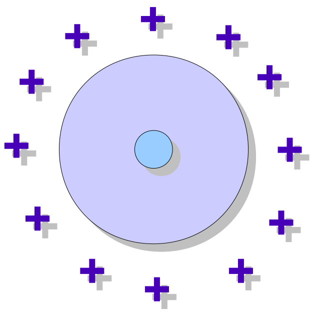
ИНС, использующие в качестве активационной функции радиального типа. Такие сети также называют RBF-сетями.

Сигмоидальный нейрон реализует линейное разделение пространства входных значений:



либо $\sum_j w_{ij} x_j > 0$, либо $\sum_j w_{ij} x_j < 0$

Радиальный нейрон представляет собой гиперсферу, которая осуществляет шаровое разделение пространства вокруг центральной точки.



$$\phi(x) = \phi(\|\vec{x} - \vec{c}\|)$$

Нейроны такого типа реализуют функции, радиально изменяющиеся вокруг выбранного центра *c*.

Общий вид радиальной функции:

$$\phi(x)=\phi\left(\frac{x^2}{\sigma^2}\right), \qquad \phi(x) \text{ - убывающая функция}$$

например,

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

где σ - ширина «окна» функции.

Нетрудно видеть, что приведённая функция изменяется радиально относительно нуля

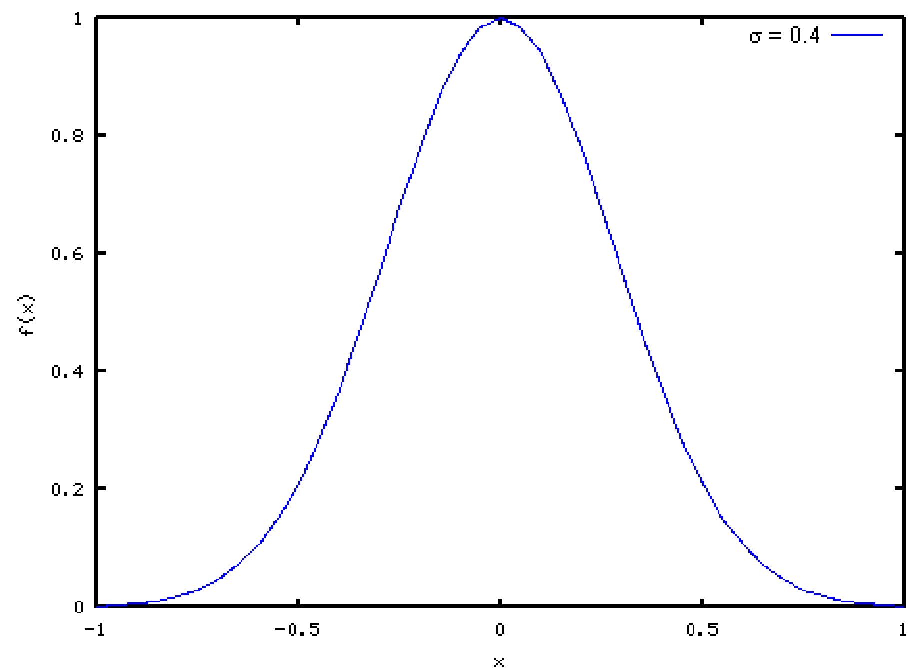


Рис. 1. График радиальной функции

Существуют и другие виды радиально изменяющихся функций, однако функция Гаусса является наиболее распространённой.

Особенности RBF-сетей

- Нет необходимости в большом количестве скрытых слоёв. Достаточно только одного скрытого слоя.
- Нейроны скрытого слоя имеют радиальный тип активационной функции.
- Выходной слой состоит из одного или нескольких линейных нейронов

RBF-сети используются для решения задач классификации, либо аппроксимации функций многих переменных и в прогнозировании — во многих приложениях, где работают сети с линейными нейронами, однако RBF реализуют иные методы обработки данных, что обеспечивает упрощение и ускорение обучения.

Рассмотрим пример использования радиально-базисной сети для решения задачи аппроксимации функции одной переменной на интервале [-1, 1].

Пусть известны 9* точек кривой, описываемой функцией:

$$f(x) = (x-3)^4 - 4x^3$$

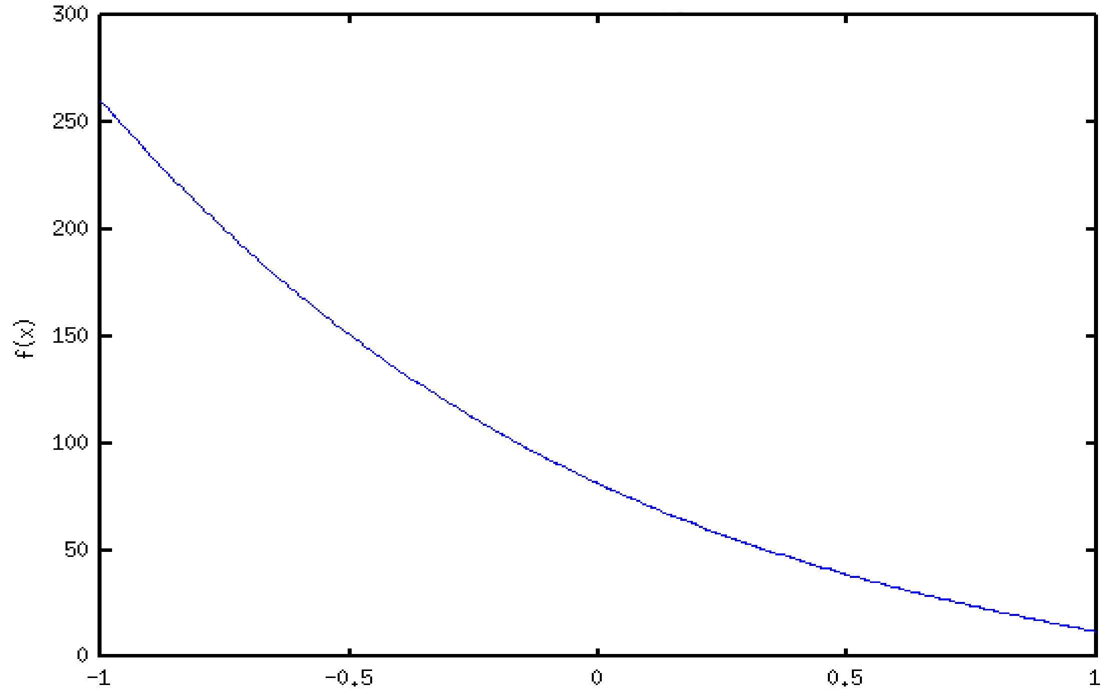


Рис. 2. График аппроксимируемой функции

* Пояснение относительно количества точек будет дано далее

Эта кривая будет аппроксимироваться с помощью взвешенных функций Гаусса:

$$v_j = e^{-\frac{(x-c_j)^2}{2\sigma}}, \; j=1,2,\ldots,9$$

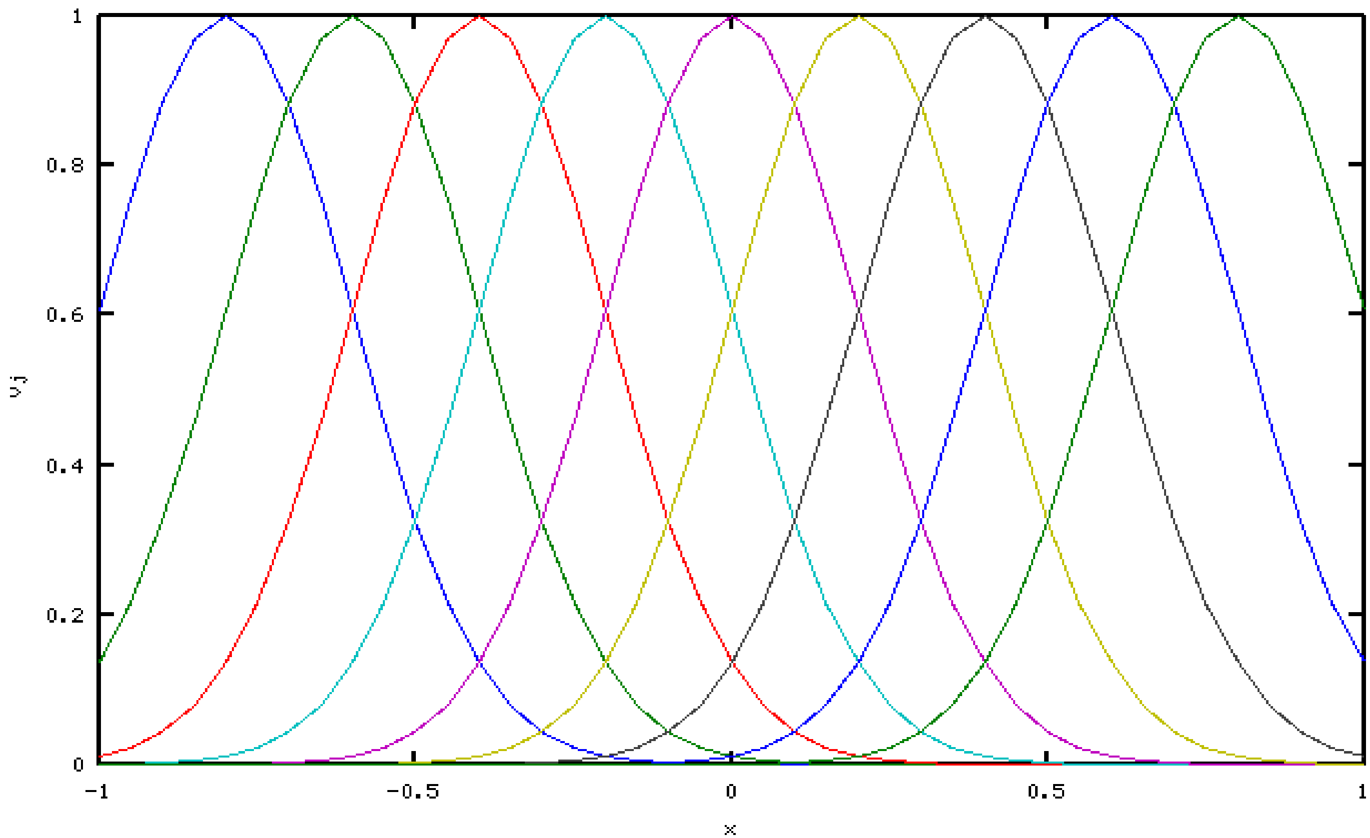


Рис. 3. Множество радиальных функций на интервале

Топология сети имеет вид, представленный ниже. Сеть имеет скалярный вход, значение которого передаётся на 9 радиальных базисных нейронов, взвешенные выходы которых формируют скалярный выход сети.

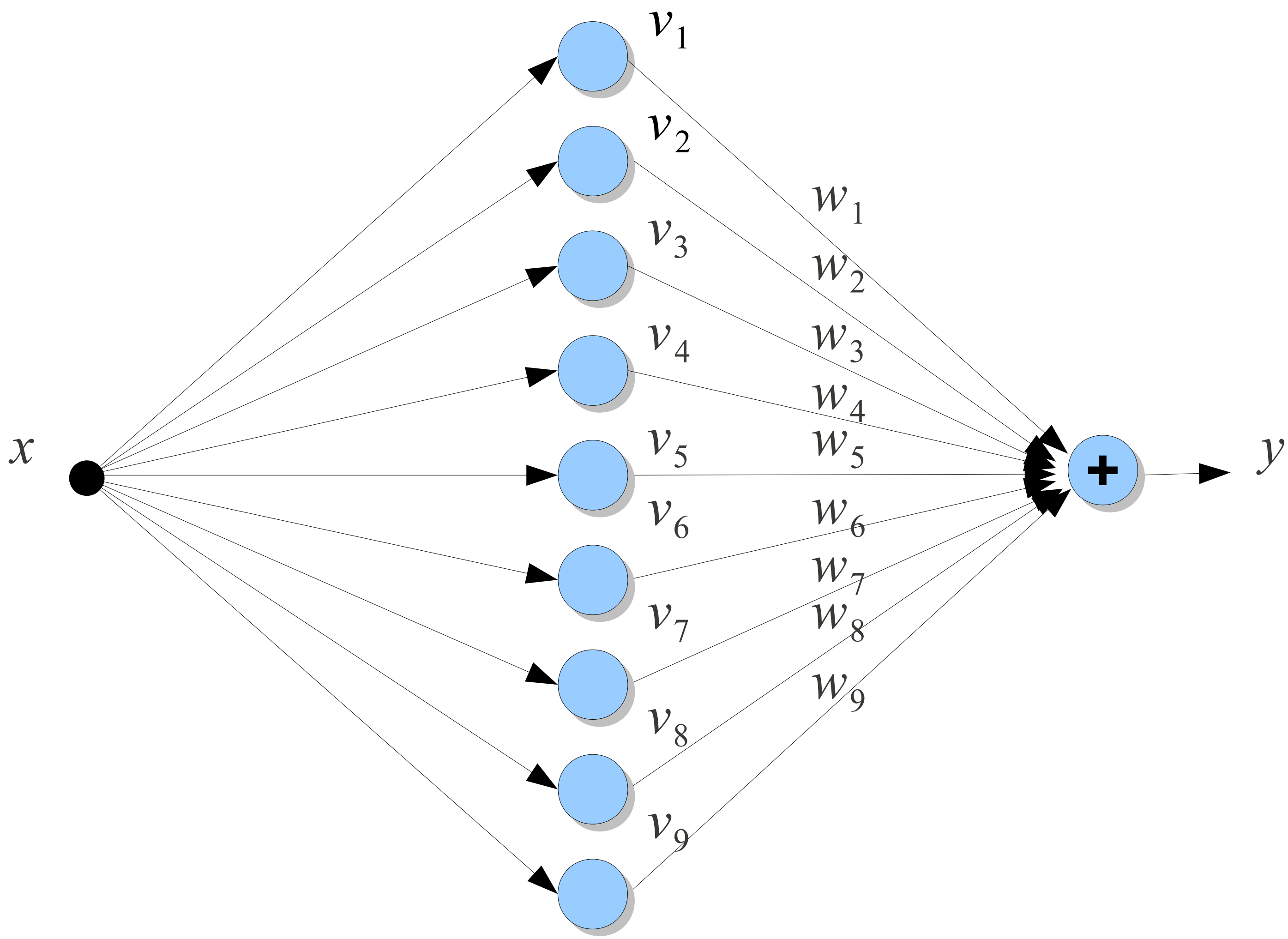


Рис. 4. Топология радиально-базисной нейронной сети

Количество нейронов скрытого слоя, расположение их центров, а также ширина окна радиальной функции подбираются экспертом под конкретные условия. От этих параметров зависит точность приближения.

Не имея определённых соображений относительно расположения центрв радиальных функций, разместим их равномерно на оси внутри выбранного интервала аппроксимации.

Центры функций выберем в точках: [-0.8; -0.6; -0.4; -0.2; 0; 0.2; 0.4; 0.6; 0.8]

Ширина «окна» функции $\sigma = 0.04$

Кривая аппроксимируется взвешенной суммой (линейной комбинацией) базисных функций, реализуемых нейронами скрытого слоя. Таким образом, эти функции представляют собой базис для построения аппроксимирующей кривой (рис. 3).

Необходимо найти весовые (масштабирующие) коэффициенты для этих функций.

Весовые коэффициенты $(w_{1,} w_{2,} \ldots, w_9)$ радиальных функций должны обеспечивать получение значения аппроксимируемой функции в точках $(x_{1,} x_{2,} \ldots, x_p)$ обучающей выборки, т.е.:

$$\begin{gathered} v_1(x_1) \cdot w_1 + v_2(x_1) \cdot w_2 + \ldots + v_9(x_1) \cdot w_9 = f(x_1) \\ v_1(x_2) \cdot w_1 + v_2(x_2) \cdot w_2 + \ldots + v_9(x_2) \cdot w_9 = f(x_2) \\ \ldots \\ v_1(x_p) \cdot w_1 + v_2(x_p) \cdot w_2 + \ldots + v_9(x_p) \cdot w_9 = f(x_p) \end{gathered}$$

что в векторном виде запишется как:

$$V \boldsymbol{w} = \boldsymbol{f}$$

где $\boldsymbol{w} = (w_{1,} w_{2,} \ldots, w_9)^T, \quad \boldsymbol{f} = (f(x_1), f(x_2), \ldots, f(x_p))^T$

$$V = \begin{pmatrix} v_1(x_1) & v_2(x_1) & \ldots & v_9(x_1) \\ v_1(x_2) & v_2(x_2) & \ldots & v_9(x_1) \\ & & \ldots & \\ v_1(x_p) & v_2(x_p) & \ldots & v_9(x_p) \end{pmatrix}$$

Отсюда видно, что количество *p* точек обучающей выборки должно быть не меньше числа нейронов скрытого слоя $p \geq 9$

Взяв количество точек обучающей выборки *p* = 9, получим квадратную матрицу *V*, тогда найти весовые коэффициенты скрытого слоя можно будет, решив уравнение:

$$\boldsymbol{w} = V^{-1} \boldsymbol{f}$$

где V^{-1} - матрица 9x9, обратная *V* .

Пусть в качестве точек обучающей выборки будут следующие:

X = [-0,87; -0,64; -0,22; 0,07; 0,16; 0,42; 0,5; 0,77; 0,96]

Вектор требуемой реакции сети для каждой из этих точек (значение функции $f(x)$)

f = [226,94; 176,6; 107,55; 73,70; 65,04; 44,01; 38,56; 22,90; 13,78]

Также для всех значений вектора ***x*** вычисляется матрица *V* и обратная ей.

Умножая обратную матрицу на вектор ***f***, получим вектор весовых коэффициентов

w = [280,24; -122,16; 200,44; -59,18; 91,31; -14,13; 42,63; -5,23; 19,23]

Тогда, с учётом полученных весовых коэффициентов, радиальные функции будут иметь вид:

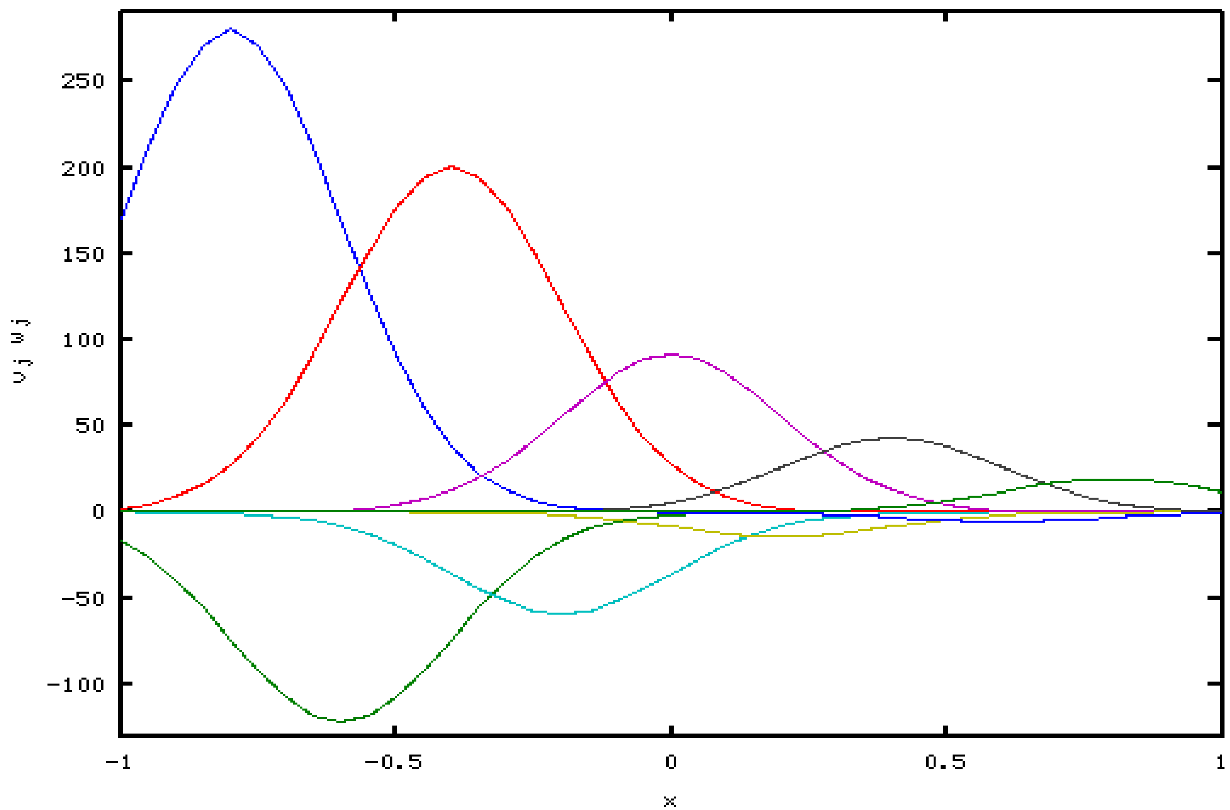


Рис. 5. Масштабированные радиальные базисные функции

Сумма этих функций представляет собой искомую аппроксимирующую кривую

Аппроксимирующая кривая проходит через все точки обучающей выборки.

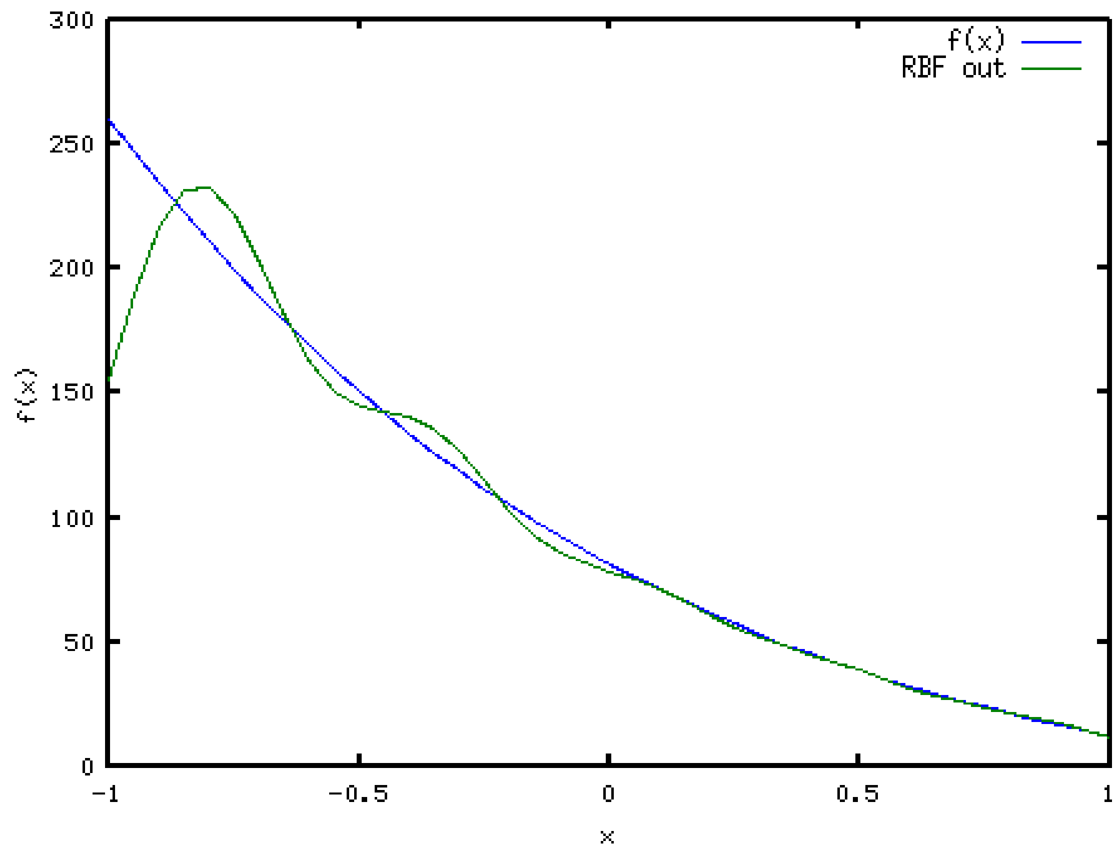


Рис. 6. Исходная функция f(x) и аппроксимирующая её кривая

Обучение также можно выполнить по правилу Видроу-Хоффа, согласно которого вычисляется ошибка выхода

$$\delta = d - y$$

а коррекция весовых коэффициентов выполняется по дельта-правилу:

$$\Delta w_j = \eta \delta v_j$$

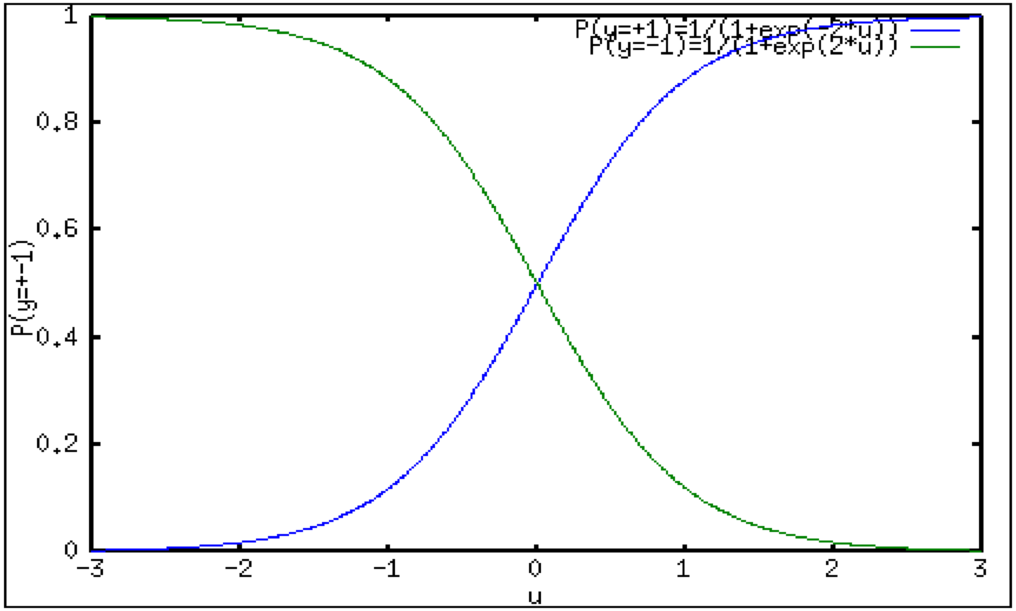
где η - коэффициент скорости обучения, v_j - выход *j* - го радиального нейрона

Стохастическая модель нейрона

В стохастической модели нейрона его выход зависит не только от взвешенной суммы сигналов, но и от некоторой **случайной переменной**.

В стохастической модели нейрона выходной сигнал y_i принимает значение ± 1 с вероятностью

$$P(y_i = \pm 1) = \frac{1}{1 + e^{\mp 2 \beta u_i}}$$



где u_i - взвешенная сумма входных сигналов *i* -го нейрона, β - положительная константа, которая обычно равна 1.

Процесс обучения состоит из следующих этапов

- Расчёт взвешенной суммы $u_i = \sum_{j=0}^{N} w_{ij} x_j$ для каждого нейрона сети.
- Расчёт вероятности того, что y_i принимает значение ± 1 в соответствии с формулой:

$$P(y_i = \pm 1) = \frac{1}{1 + e^{\mp 2\beta u_i}}$$

- Генерация значений случайной переменной $R \in (0, 1)$ и формирование выходного сигнала:

$$\begin{cases} y_i = \pm 1, & \text{если } R < P(y_i = \pm 1) \\ y_i = \mp 1, & \text{в противном случае} \end{cases}$$

- Этот процесс осуществляется на случайно выбранной группе нейронов, вследствие чего их состояние модифицируется в соответствии с этим правилом.
- После фиксации состояния отобранных нейронов их весовые коэффициенты модифицируются по применяемому правилу уточнения весов.

Например, уточнение весов при обучении с учителем можно выполнять по правилу Видроу-Хоффа:

$$\Delta w_{ij} = \eta x_j (d_i - y_i)$$

Доказано, что такой способ подбора весов приводит к минимизации целевой функции (среднеквадратичная погрешность):

$$E = \frac{1}{2} \sum_{k=1}^{p} \sum_{i=1}^{n} (d_i^{(k)} - y_i^{(k)})^2$$

Которая рассчитывается по всем *n* нейронам и *p* обучающим выборкам.