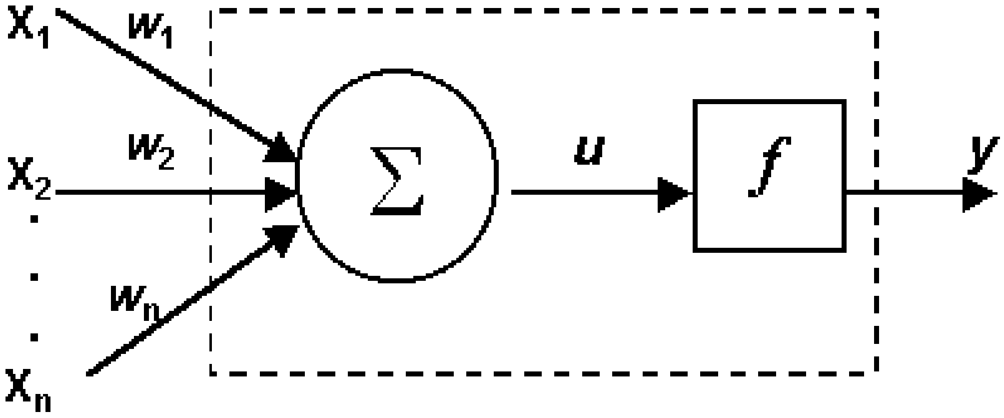
Лекция 2. Однослойные НС.

Буряк Д.Ю.
к.ф.-м.н
dyb04@yandex.ru

Структура искусственного нейрона



$$u = \sum_{i=1}^{n} w_i x_i$$

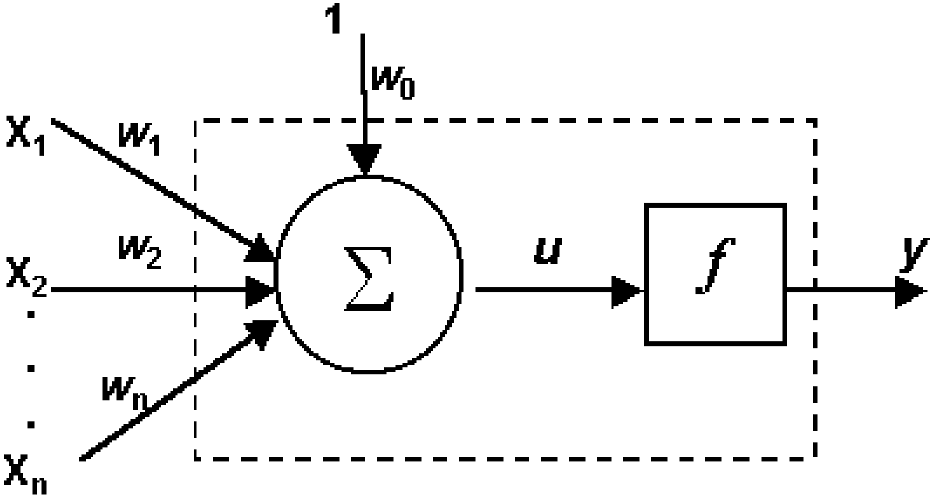
$$u = xW$$

$$y = f(u) = f(\sum_{i=1}^{n} w_i x_i)$$

f(u) - активационная функция;
f(u) - обычно нелинейная.

Бинарный нейрон

$$f(u) = \begin{cases} 1, u > T \\ 0, u \le T \end{cases}$$ - пороговая активационная функция.

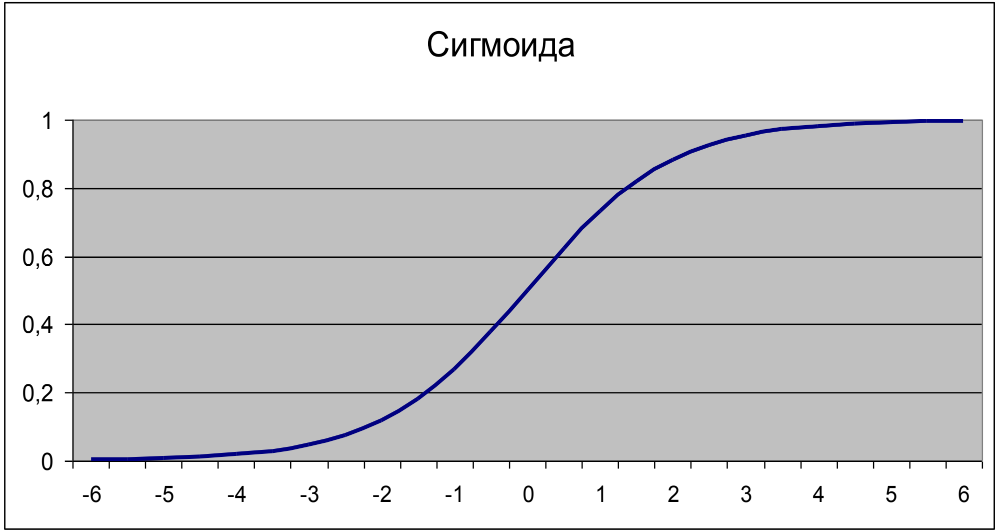


$$u = \sum_{i=1}^{n} w_i x_i + w_0$$

$$y = f(u) = \begin{cases} 1, u > 0 \\ 0, u \le 0 \end{cases}$$

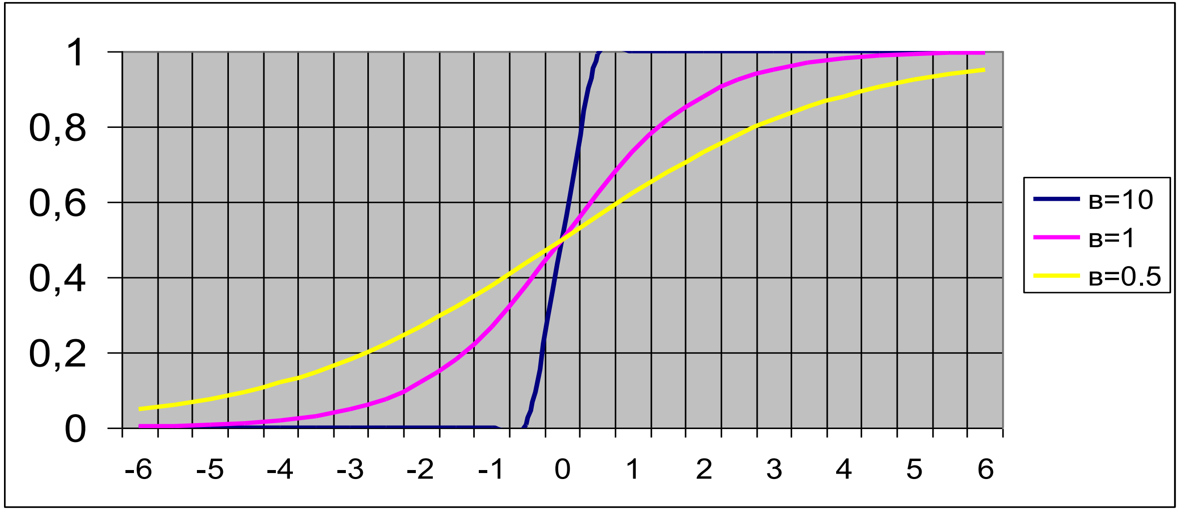
Функции активации. Сигмоида

Сигмоидальная (S-образная) функция: $$f(u) = \frac{1}{1 + e^{-\beta u}}$$



Функции активации. Сигмоида

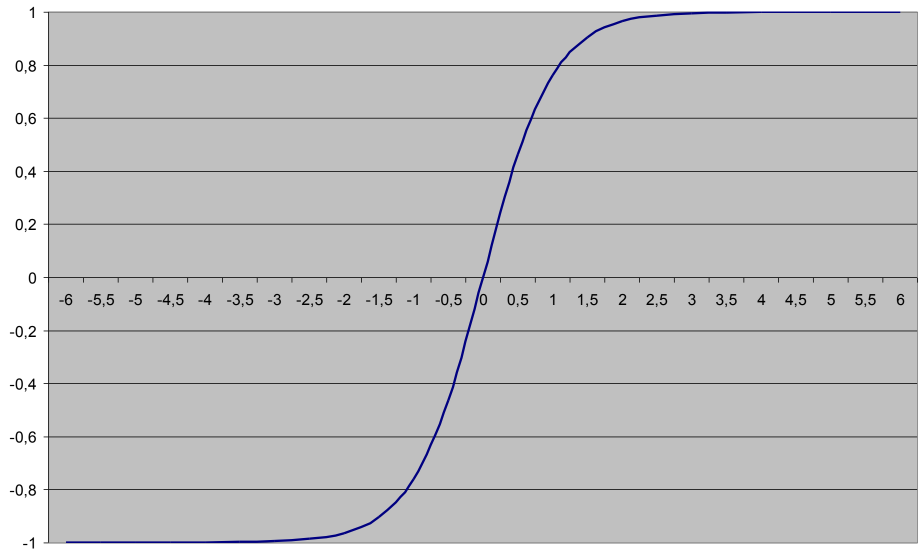
Влияние параметра β на форму кривой.



При $\beta \to \infty$ сигмоидальная функция превращается в пороговую.

Функции активации. Гиперболический тангенс

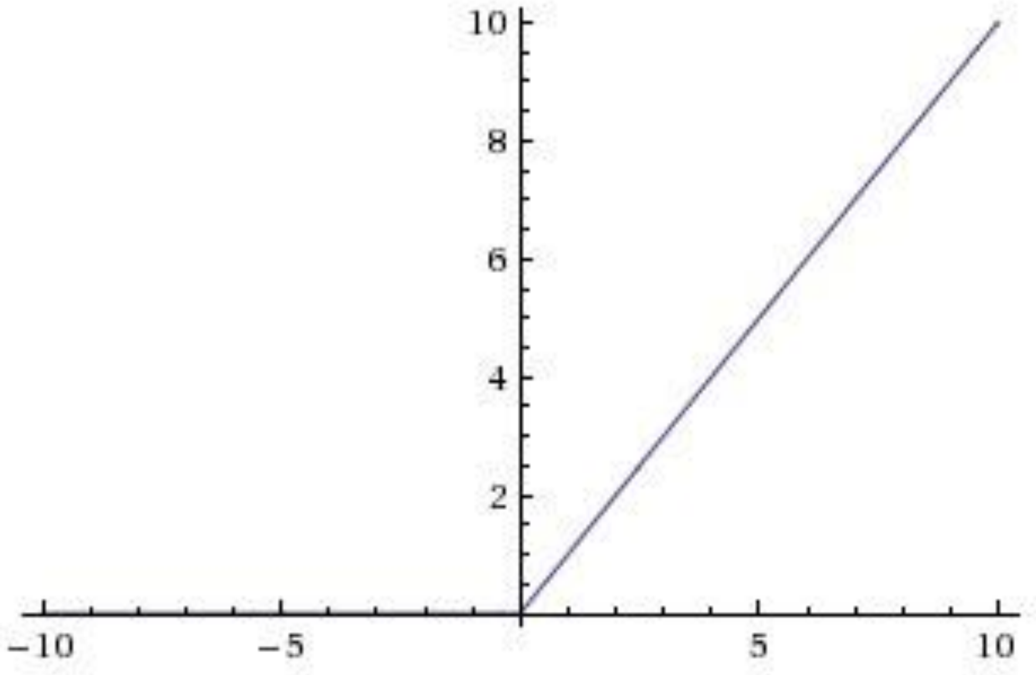
Функция гиперболический тангенс: $$f(u) = th(\beta u) = \frac{e^{\beta u} - e^{-\beta u}}{e^{\beta u} + e^{-\beta u}}$$



Функции активации. ReLU

ReLU – Rectified Linear Unit:

$$f(u) = \max(0, u)$$

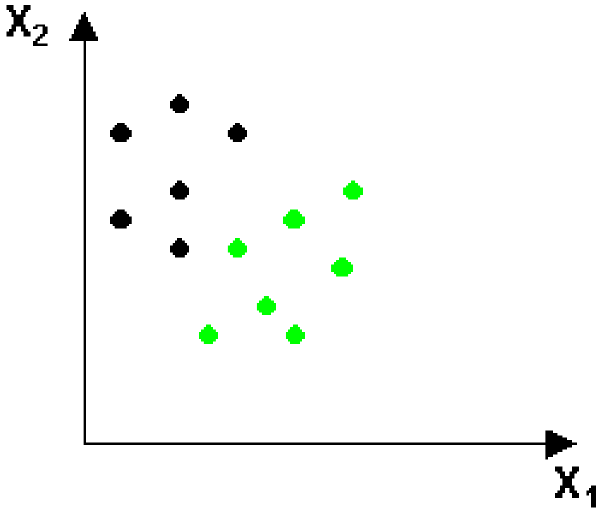


Задача классификации

Рассмотрим множества точек на плоскости:

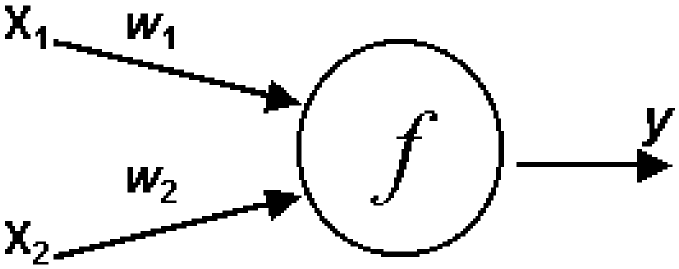
$$S_1 = \{(x_{1i}^1, x_{2i}^1) \mid i = 1,2,\ldots,K_1\}$$

$$S_2 = \{(x_{1j}^2, x_{2j}^2) \mid j = 1,2,\ldots,K_2\}$$



Можно ли построить НС, классифицирующую точки из S_1 и S_2?

Однослойный персептрон



- Персептрон
  - модель МакКаллока-Питса
  - 1 слой
  - алгоритм обучения.

$$u = w_1 x_1 + w_2 x_2 + w_0$$

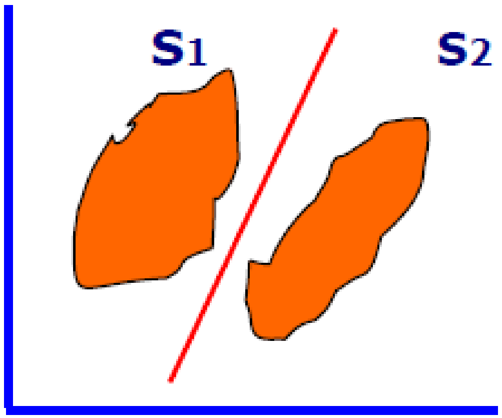
$$y = f(u) = \begin{cases} 1, u > 0 \\ 0, u \le 0 \end{cases}$$

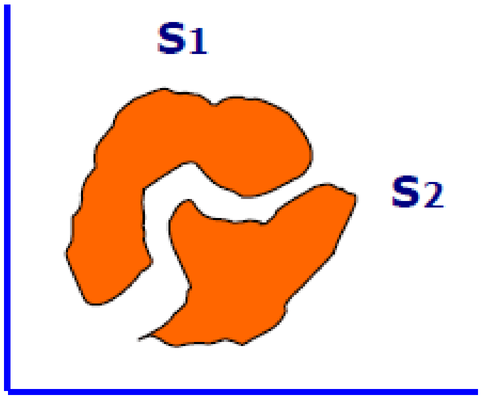
Знак выражения $u = w_1 x_1 + w_2 x_2 + w_0$ определяет класс, к которому будет отнесена точка (x_1, x_2)

Линейная разделимость

$w_1x_1 + w_2x_2 + w_0 = 0$ - разделяющая прямая.

Однослойная НС позволяет решать задачи классификации, в которых образы линейно разделимы.

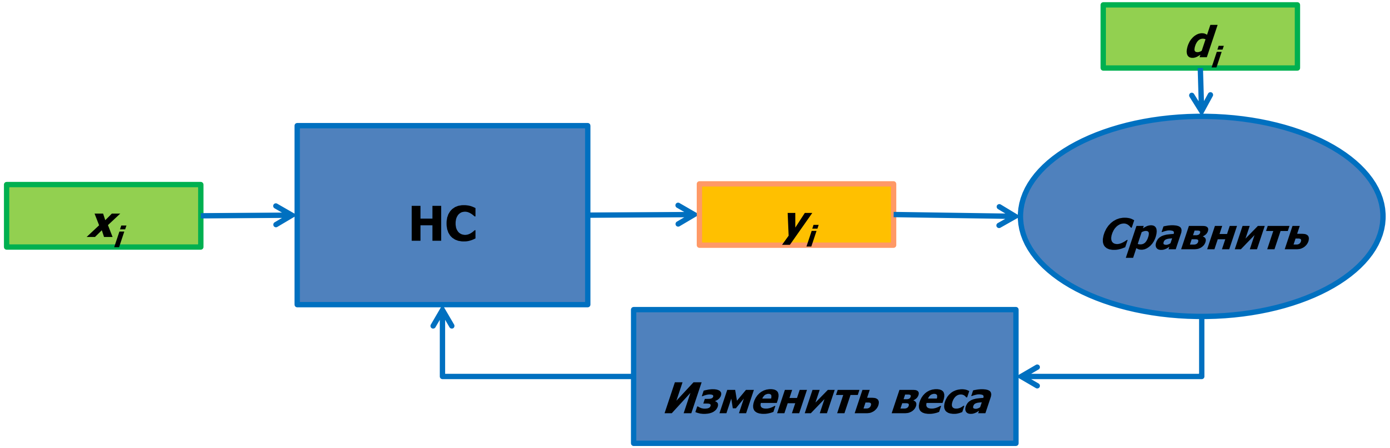




Обучение НС

- Цель обучения - вычисление весов синаптических связей сети.
- Обучение проводится на примерах.
- Обучающая выборка: $S=\{(x_i,d_i)|i=1,2,\ldots,N\}$.

  x_i – входной вектор; d_i – выходной вектор.
- Критерий окончания – суммарная ошибка на всех векторах из S



Обучение бинарного нейрона

- Бинарный нейрон $$u = \sum_{i=1}^{n} w_i x_i + w_0 \qquad y = f(u) = \begin{cases} 1, u > 0 \\ 0, u \le 0 \end{cases}$$
- Обучающая выборка: (x_i, d_i)
- Правило персептрона (обучение с учителем):
  1. Если y совпадает с ожидаемым значением d, то веса не изменяются.
  2. Если y=0, d=1, то $$w_i(t+1) = w_i(t) + x_i$$
  3. Если y=1, d=0, то $$w_i(t+1) = w_i(t) - x_i$$

Правило Видроу-Хоффа:

$$w_i(t+1) = w_i(t) + \Delta w_i$$

$$\Delta w_i = x_i(d - y)$$

Обучение сигмоидального нейрона

Функция активации: $$f(u) = \frac{1}{1+e^{-\beta u}}$$

Принцип обучения - минимизация целевой функции: $$E = \frac{1}{2}(y-d)^2$$

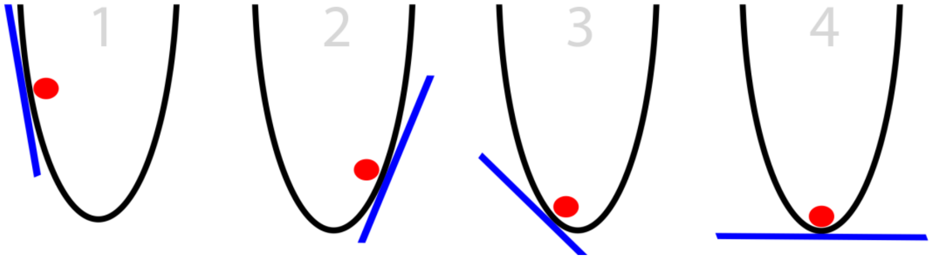
d - желаемый выход

y - реальный выход

$$y = f(u) = f\left(\sum_{i=1}^{n} w_i x_i\right)$$

Метод обучения - градиентный спуск.

$$\nabla_i E = \frac{dE}{dw_i} = e x_i \frac{df(u)}{du}$$



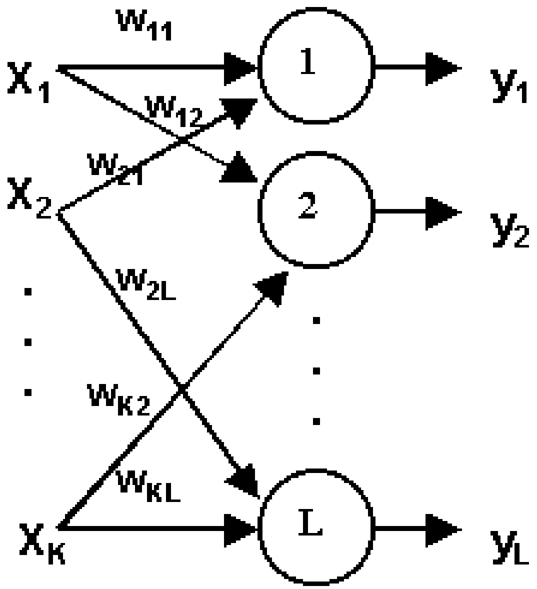
$$\delta = e \frac{df(u)}{du} \qquad e = (y-d)$$

$$w_i(t+1) = w_i(t) - \eta \nabla_i E = w_i(t) - \eta \delta x_i$$

η- коэффициент обучения, выбирается из интервала (0,1).

Обучение сигмоидального нейрона (2)

Обучающая выборка: $S=\{(X^i,d^i)|i=1,2,\ldots,N\}$.



$$E=\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{L}(y_j^i-d_j^i)^2$$

$$y_j^i=f(u_j^i)=f(\sum_m w_{mj}x_m)$$

$$\delta_j=e_j\frac{df(u_j)}{du} \qquad e_j=(y_j-d_j)$$

$$w_{mj}(t+1)=w_{mj}(t)-\eta\delta_j x_m$$

Обучение сигмоидального нейрона (3)

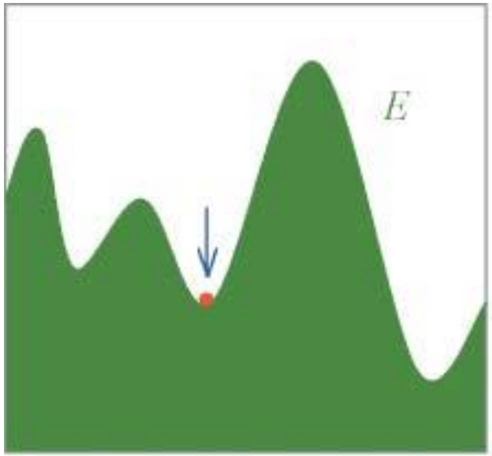
- Особенности сигмоидальных функций
  - сигмоида:

$$\frac{df(u)}{du} = \beta f(u)(1 - f(u))$$

  - гиперболический тангенс:

$$\frac{df(u)}{du} = \beta(1 - f^2(u))$$

- Проблема градиентного метода - достижение локального минимума.



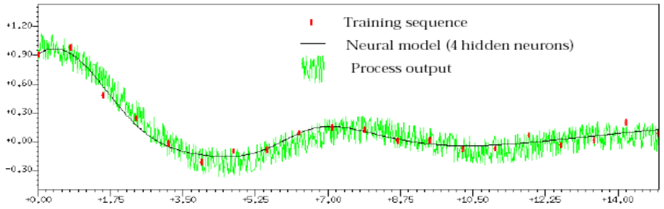
- Модификация градиентного метода обучения:

$$w_i(t+1) = w_i(t) - \eta\delta x_i + \alpha\Delta w_i(t)$$

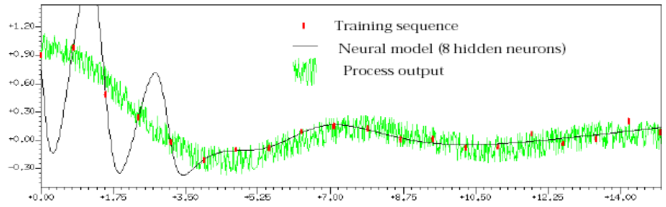
Критерии окончания обучения

Переобучение

❑ «Привыкание» к примерам из обучающей выборки



❑ Использование подтверждающей выборки



❑ Критерии окончания:

- по количеству проведенных итераций
- по ошибкам на обучающей и подтверждающей выборках

Проблемы обучения

- Выбор алгоритма обучения
  - скорость сходимости
  - качество сходимости
  - вычислительные ресурсы
- Построение обучающей выборки
  - репрезентативность
  - размер
- Начальная инициализация весов.
- Определение момента окончания процесса обучения
  - переобучение