

Лекция 5

Обобщенная модель нейрона.

Персептрон (структура, алгоритмы обучения).

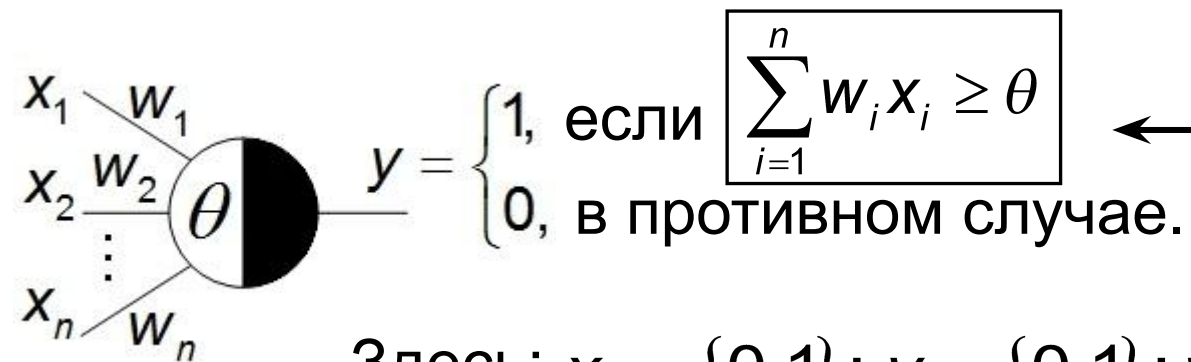
Решение задачи распознавания образов с помощью НС.



1943 – формальный нейрон (МакКаллок, Питтс):



1958 – расширение модели (МакКаллок):



Здесь: $x_i \in \{0, 1\}$; $y \in \{0, 1\}$; $w_i \in \{-1, 1\}$

$w_i = -1$ «тормозящие» входы; $w_i = 1$ «возбуждающие» входы

условие возбуждения

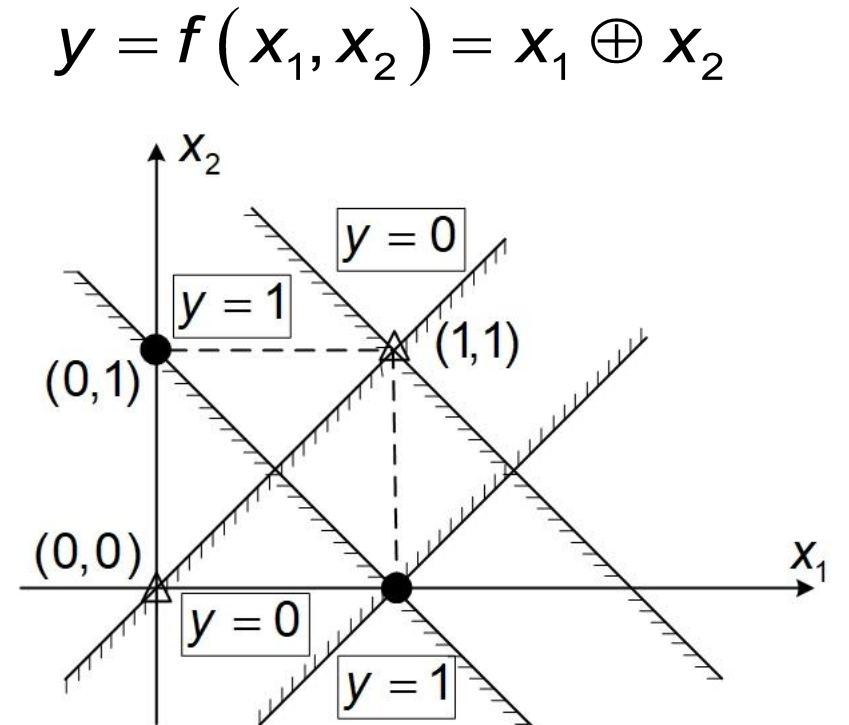
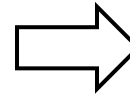
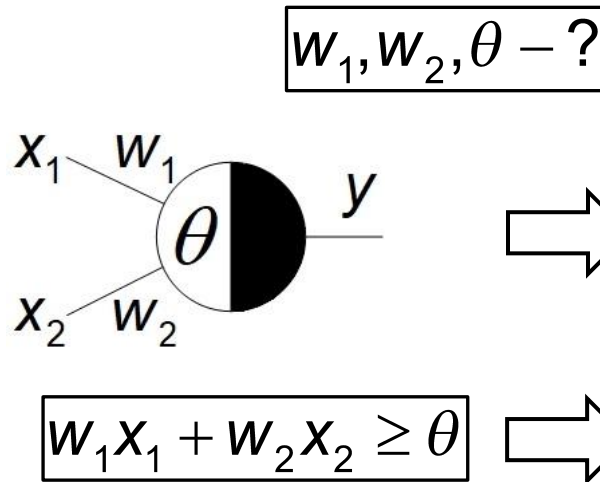
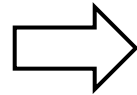
(θ – порог возбуждения)

ПРОБЛЕМА ИСКЛЮЧАЮЩЕГО ИЛИ (Exclusive OR, XOR)

→ [М. Мински, С. Пейперт, 1969]

Таблица истинности
функции XOR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0



⇒ XOR = линейно неразделимая функция

Решение проблемы – 2-хслойная НС:

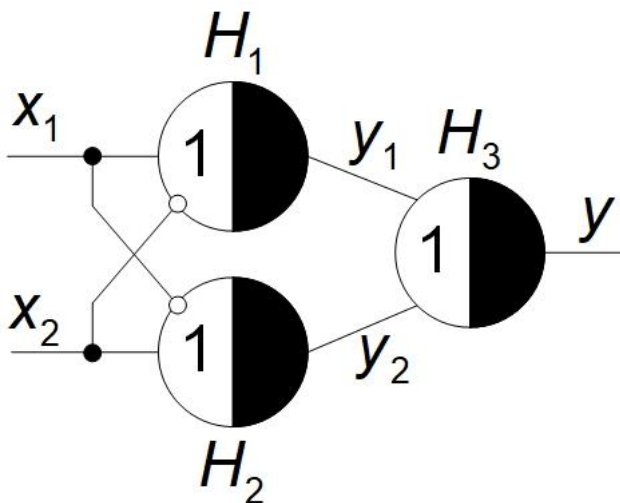
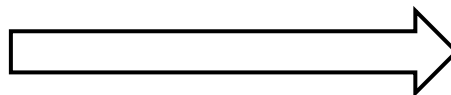


Таблица
истинности:



x_1	x_2	y_1	y_2	y
0	0	0	0	0
0	1	0	1	1
1	0	1	0	1
1	1	0	0	0

= !

Условия возбуждения нейронов:

$$H_1 : x_1 - x_2 \geq 1;$$

$$H_2 : -x_1 + x_2 \geq 1;$$

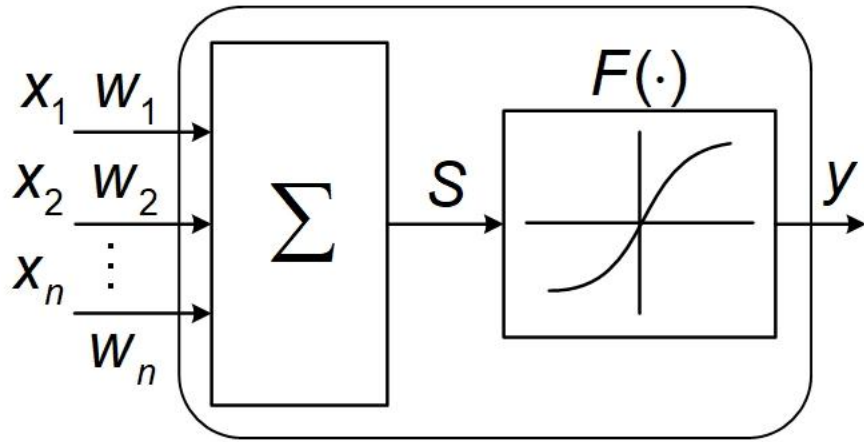
$$H_3 : y_1 + y_2 \geq 1.$$

ОБОБЩЕННАЯ МОДЕЛЬ НЕЙРОНА

Уравнение нейрона:

$$y = F(S), \quad (1)$$

где $S = \sum_{i=1}^n w_i x_i$ – суммарное возбуждение нейрона;



$x_i \in [-1, 1]$ – входы нейрона ($i = 1, 2, \dots, n$);

$y \in [-1, 1]$ – выход нейрона;

w_i ($i = 1, 2, \dots, n$) – вещественные числа, веса синаптических связей.

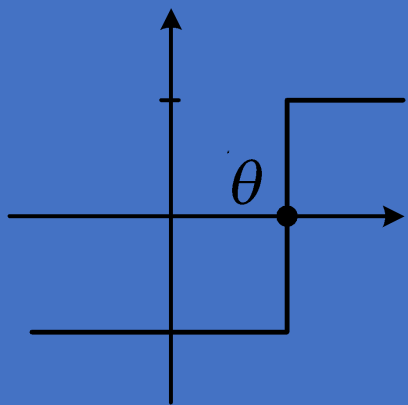
$F(S)$ – функция активации (характеристическая функция),
удовлетворяющая 2-м условиям:

а) $|F(S)| \leq 1$;

б) $F(S)$ – монотонно возрастающая (или неубывающая) функция S .

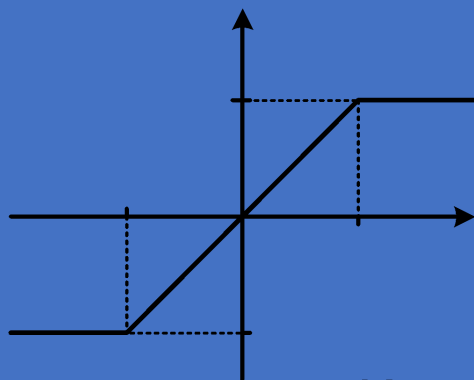
ВИДЫ ФУНКЦИИ АКТИВАЦИИ

а) логическая



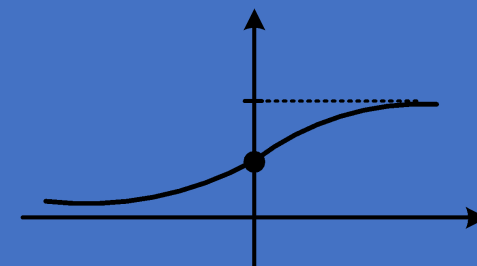
$$y = \text{sign}(S - \theta)$$

б) линейная



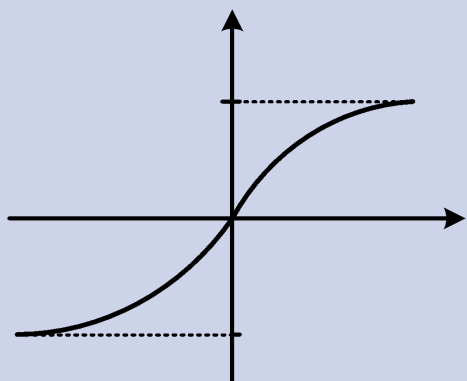
$$y = \begin{cases} S, & |S| \leq 1; \\ \text{sign}(S), & |S| > 1 \end{cases}$$

в) сигмоидная



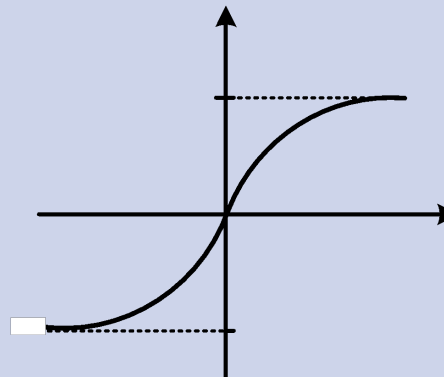
$$y = \frac{1}{1 + e^{-aS}}, \quad a > 0$$

г) 2-полярная сигмоида
(гиперболический тангенс)



$$y = \frac{1 - e^{-aS}}{1 + e^{-aS}} = \text{th}\left(\frac{S}{2}\right), \quad a > 0$$

д) рациональная сигмоида (функция
Эллиота)



$$y = \frac{S}{c + |S|}, \quad c > 0$$

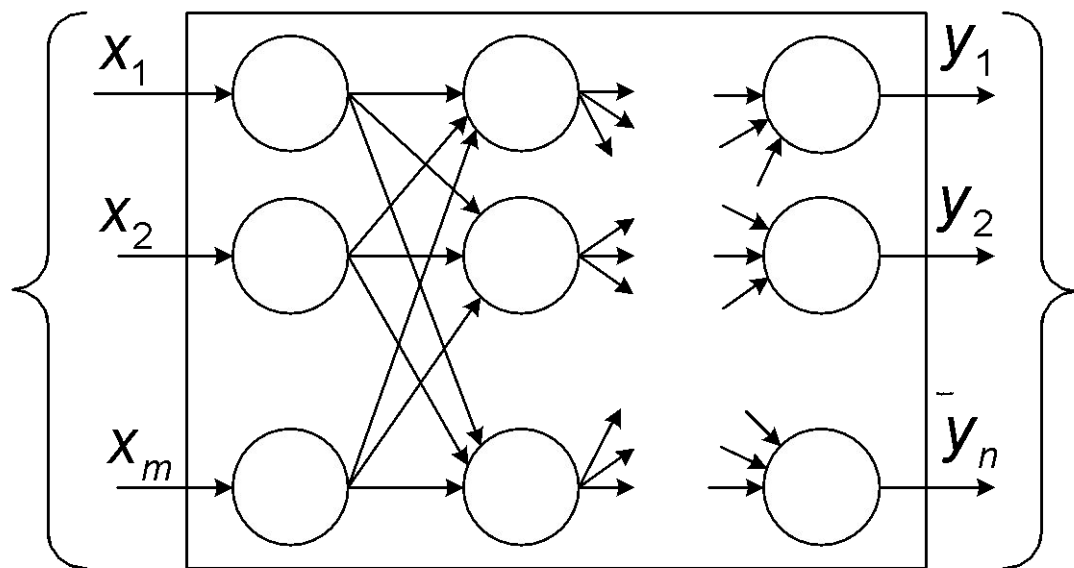
ПЕРСЕПТРОН (структура)

1958 – однослойный персептрон (Ф. Розенблат)

→ (от англ. Perception – «восприятие»)

1958 – 1-й нейрокомпьютер «Марк-1» (Ф. Розенблат)

Структура многослойного перспетрона



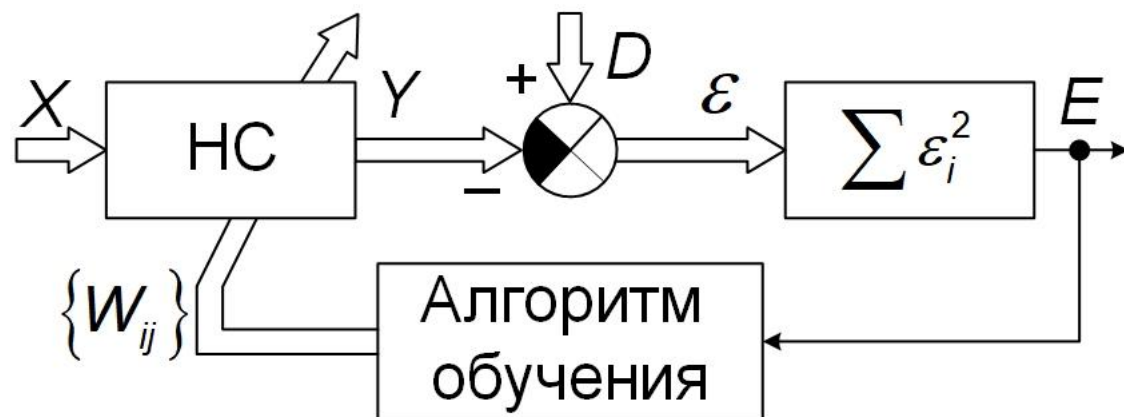
$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \text{ – вектор входов, } x_i \in [-1, 1];$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \text{ – вектор выходов, } y_j \in [-1, 1].$$

2 режима работы:

- обучение (настройка весов НС);
- эксплуатация.

ПЕРСЕПТРОН (процедура обучения)



Цель обучения: $E \rightarrow \min$,

где $E = \left| \boldsymbol{\varepsilon} \right|^2$ — суммарная

квадратичная ошибка обучения (СКО) НС.

$$D = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} \text{ — вектор эталонов, (желаемых выходов) НС;}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \text{ — вектор ошибки обучения НС;}$$

$$\boldsymbol{\varepsilon} = D - Y$$

ОБУЧАЮЩАЯ ВЫБОРКА

– это набор пар $\{(X^{(r)}, D^{(r)})\}$, $(r = 1, 2, \dots, R)$, соответственно входной вектор и вектор желаемых выходов НС r -м обучающем примере; R – число обучающих примеров (объем обучающей выборки).

r	Входы НС				Желаемые выходы НС			
	x_1	x_2	\dots	x_m	d_1	d_2	\dots	d_n
1								
2								
\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots
R								

АЛГОРИТМЫ ОБУЧЕНИЯ ПЕРСЕПТРОНА

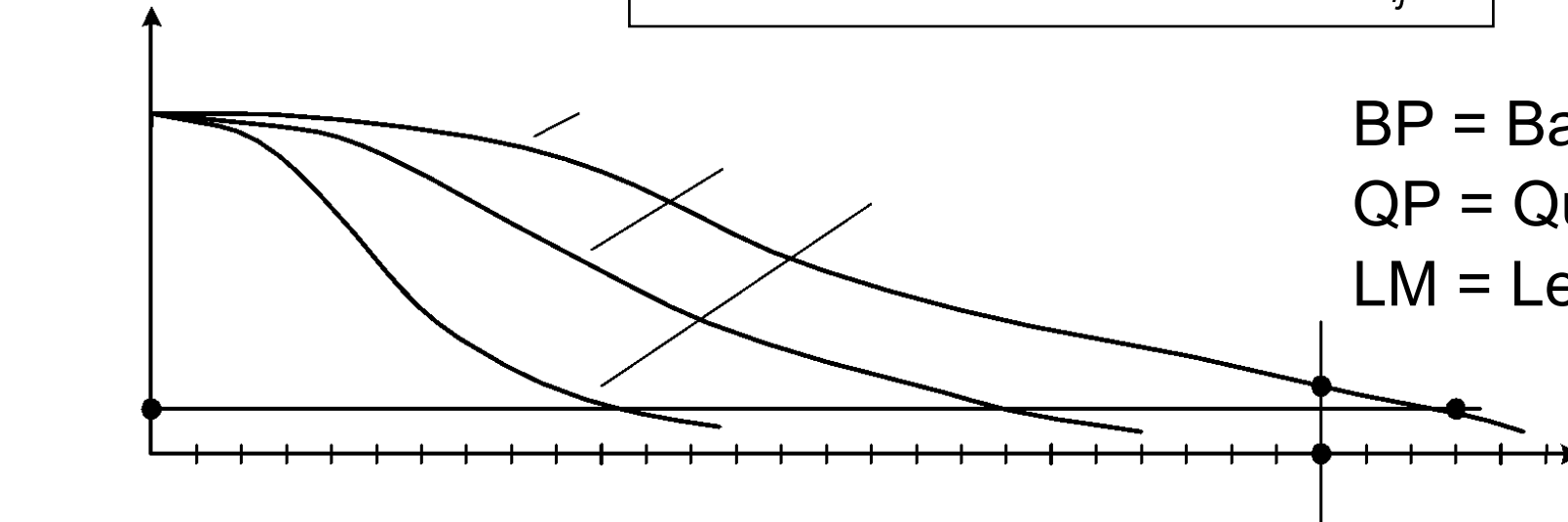
СКО: $E = \sum_{r=1}^R \sum_{i=1}^n [d_i^{(r)} - y_i^{(r)}]^2 \rightarrow \min,$

где n – число выходов НС; R – объем обучающей выборки.

Градиентный алгоритм обучения:

$$W_{ij}(k+1) = W_{ij}(k) - \alpha \frac{\partial E(k)}{\partial W_{ij}}$$

α – скорость обучения
($0 < \alpha < 1$)



BP = Back Propagation

QP = Quick Prop

LM = Levenberg- Marquardt

Критерий останова: а) $k = K_{\text{зад}}$; б) $E = E_{\text{зад}}$

СЛОЖНОСТИ ПРОЦЕССА ОБУЧЕНИЯ

1) «зависание» в локальных минимумах

→ метод «тяжелого» шарика:

$$W_{ij}(k+1) = W_{ij}(k) + \Delta W_{ij}(k) + \beta \cdot \Delta W_{ij}(k-1)$$

где $\Delta W_{ij}(k) = -\alpha \frac{\partial E(k)}{\partial W_{ij}}$ момент; $\alpha < 1$ – коэффициент

2) регуляризация процесса обучения:

$$E = \sum_{r=1}^n \sum_{i=1}^n [d_i^{(r)} - y_i^{(r)}]^2 + \gamma \cdot \sum_{i,j} W_{ij}^2 \rightarrow \min$$

где γ – коэффициент «штрафа» весов НС;

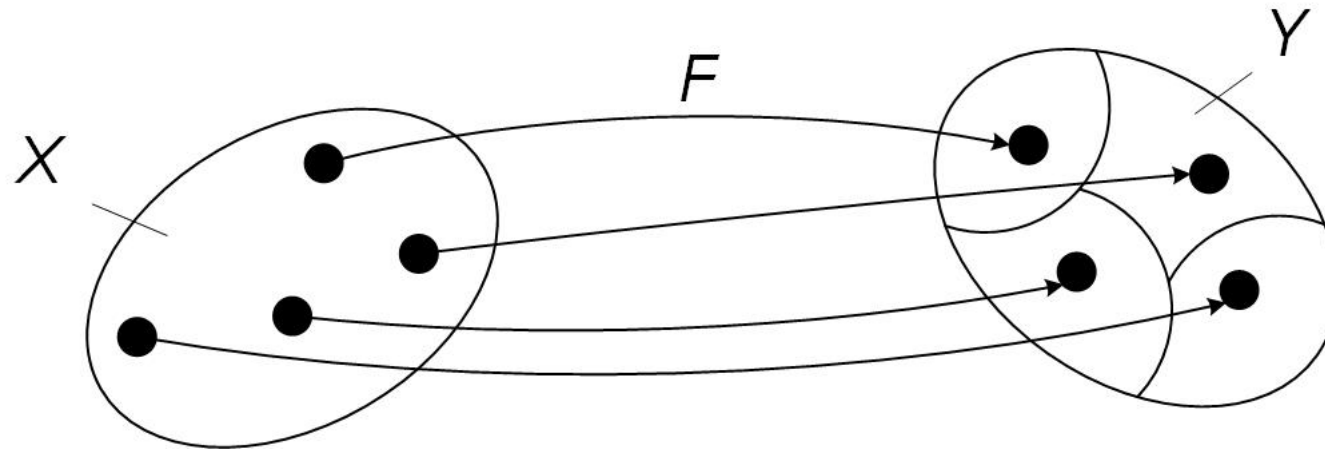
3) эффект «переобучения» (over-learning) НС

→ выбор оптимальной сложности НС (количества нейронов в скрытом слое); тестирование / валидация НС.

РЕШЕНИЕ ЗАДАЧИ РАСПОЗНАВАНИЯ ОБРАЗОВ С ПОМОЩЬЮ ПЕРСЕПТРОНА

Задача классификации (распознавания образов):

требуется определить, к какому из M известных классов относится предъявляемый объект (образ), представленный вектором $X = (x_1, x_2, \dots, x_m)^T$.

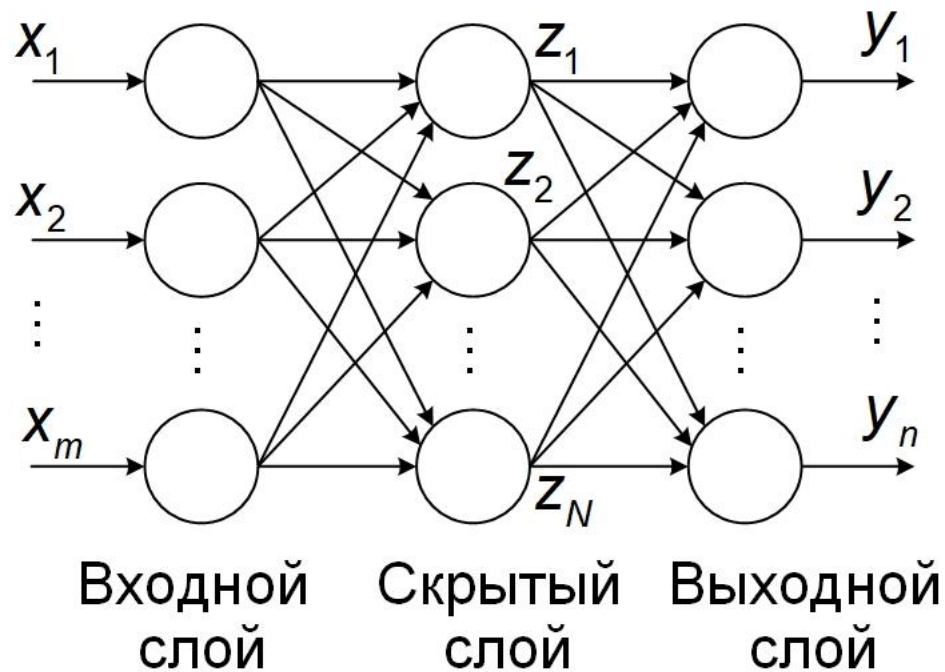


Множество входных
векторов (признаков)

Множество решений
(классов принадлежности)

Решение:

Персептрон с 1 скрытым слоем:



$$n = \lceil \log_2 M \rceil; R \geq M$$

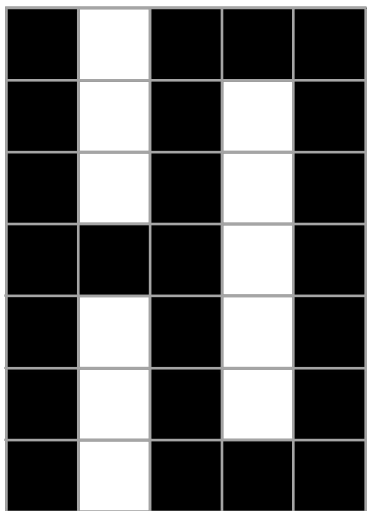
Количество нейронов в скрытом слое:

$$N = \left\lceil \frac{R \cdot n}{m + n} \right\rceil^*$$

* Васильев В.И., Ильясов Б.Г. Интеллектуальные системы управления. Теория и практика, 2009. С. 168-170.

ПРИМЕР

Дано: 7x5 матрица входных образов (изображений):

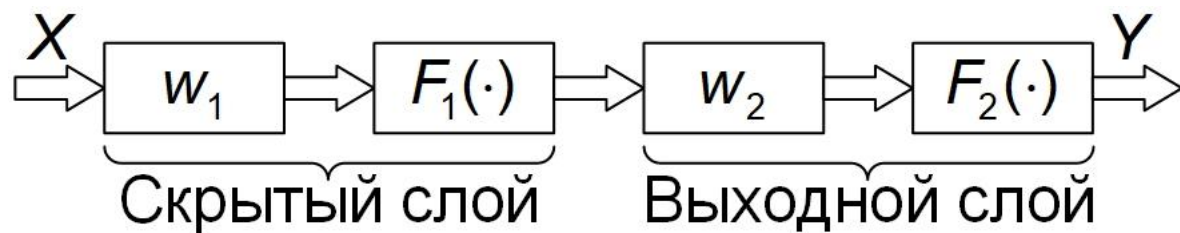


$M = 64$ (буквы, цифры, символы); входы $x_i \in \{-1, 1\}$;
вектор желаемых реакций НС: $D = (d_1, d_2, \dots, d_n)$ — двоичный код /
номер образа (класса изображений).

Решение: $m = 7 \cdot 5 = 35$; $R = M = 64$;

$$n = \lceil \log_2 64 \rceil = 6; N = \left\lceil \frac{R \cdot n}{m + n} \right\rceil = \left\lceil \frac{64 \cdot 6}{35 + 6} \right\rceil = 10.$$

Структура НС: 35 – 10 – 6:



$w_1 = \|w_{ij}^{(1)}\|_{35 \times 10}$, $w_2 = \|w_{ij}^{(2)}\|_{10 \times 6}$ — матрицы
весов синаптических связей;

$F_1(\cdot)$, $F_2(\cdot)$ — сигмоидные функции
активации

Количество настраиваемых параметров (весов) НС: (КП) = $35 \cdot 10 + 10 \cdot 6 = 410$.