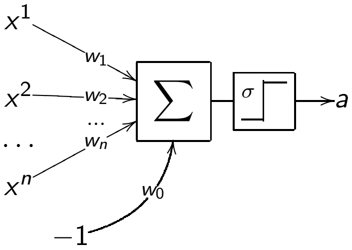
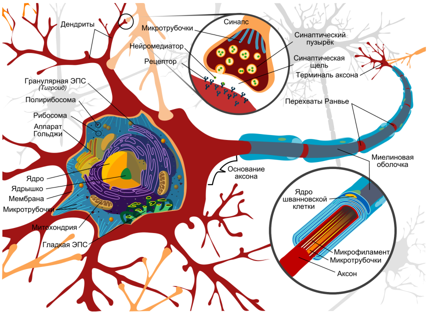
$f_j \colon X \to \mathbb{R}$, $j = 1, \dots, n$ — числовые признаки, $x^j = f_j(x)$;

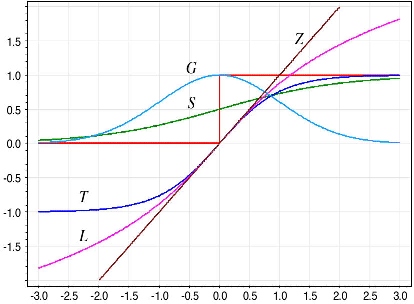
$$a(x, w) = \sigma\bigl(\langle w, x\rangle\bigr) = \sigma\left(\sum_{j=1}^{n} w_j f_j(x) - w_0\right),$$

где $w_0, w_1, \dots, w_n \in \mathbb{R}$ — веса признаков;
$\sigma(s)$ — функция активации (в частности, sign).





Часто используемые функции активации $\sigma(z)$



$\theta(z) = [z \geqslant 0]$	— пороговая функция Хевисайда;
$\sigma(z) = (1 + e^{-z})^{-1}$	— сигмоидная функция (S);
$\mathrm{th}(z) = 2\sigma(2z) - 1$	— гиперболический тангенс (T);
$\ln(z + \sqrt{z^2 + 1})$	— логарифмическая функция (L);
$\exp(-z^2/2)$	— гауссовская функция (G);
z	— линейная функция (Z);

Задача классификации: $Y = \{\pm 1\}$, $a(x, w) = \operatorname{sign}\langle w, x_i \rangle$;

$$Q(w; X^\ell) = \sum_{i=1}^{\ell} \mathscr{L}\bigl(\underbrace{\langle w, x_i \rangle y_i}_{M_i(w)}\bigr) \to \min_w;$$

Задача регрессии: $Y = \mathbb{R}$, $a(x, w) = \sigma(\langle w, x_i \rangle)$;

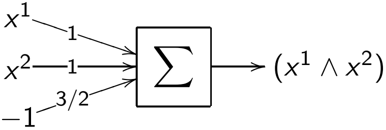
$$Q(w; X^\ell) = \sum_{i=1}^{\ell} \bigl(\sigma(\langle w, x_i \rangle) - y_i\bigr)^2 \to \min_w;$$

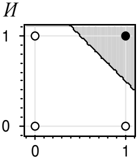
Насколько богатый класс функций реализуется нейроном?
А сетью (суперпозицией) нейронов?

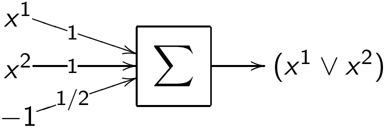
Нейронная реализация логических функций

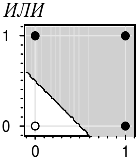
Функции И, ИЛИ, НЕ от бинарных переменных x^1 и x^2:

$$x^1 \wedge x^2 = \left[x^1 + x^2 - \tfrac{3}{2} > 0\right];$$
$$x^1 \vee x^2 = \left[x^1 + x^2 - \tfrac{1}{2} > 0\right];$$
$$\neg x^1 = \left[-x^1 + \tfrac{1}{2} > 0\right];$$



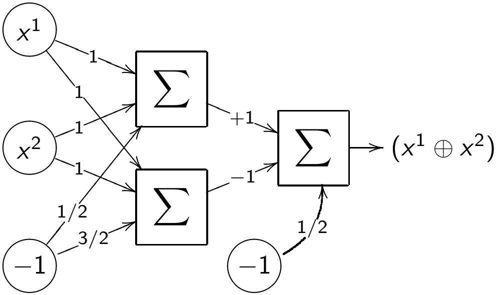


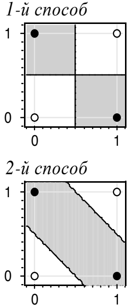




Функция $x^1 \oplus x^2 = [x^1 \neq x^2]$ не реализуема одним нейроном.
Два способа реализации:

- Добавлением нелинейного признака:
  $x^1 \oplus x^2 = \left[x^1 + x^2 - 2x^1x^2 - \frac{1}{2} > 0\right]$;
- Сетью (двухслойной суперпозицией) функций И, ИЛИ, НЕ:
  $x^1 \oplus x^2 = \left[(x^1 \vee x^2) - (x^1 \wedge x^2) - \frac{1}{2} > 0\right]$.





- Двухслойная сеть в $\{0,1\}^n$ позволяет реализовать произвольную булеву функцию (ДНФ).
- Двухслойная сеть в $\mathbb{R}^n$ позволяет отделить произвольный выпуклый многогранник.
- Трёхслойная сеть $\mathbb{R}^n$ позволяет отделить произвольную многогранную область, не обязательно выпуклую, и даже не обязательно связную.
- С помощью линейных операций и одной нелинейной *функции активации* φ можно приблизить любую непрерывную функцию с любой желаемой точностью.

Практические рекомендации:

- Двух-трёх слоёв обычно достаточно.
- Можно достраивать нейроны в произвольных местах сети по необходимости, вообще не заботясь о числе слоёв.