

# Линейные классификаторы

и немножко теории...

# Perceptron

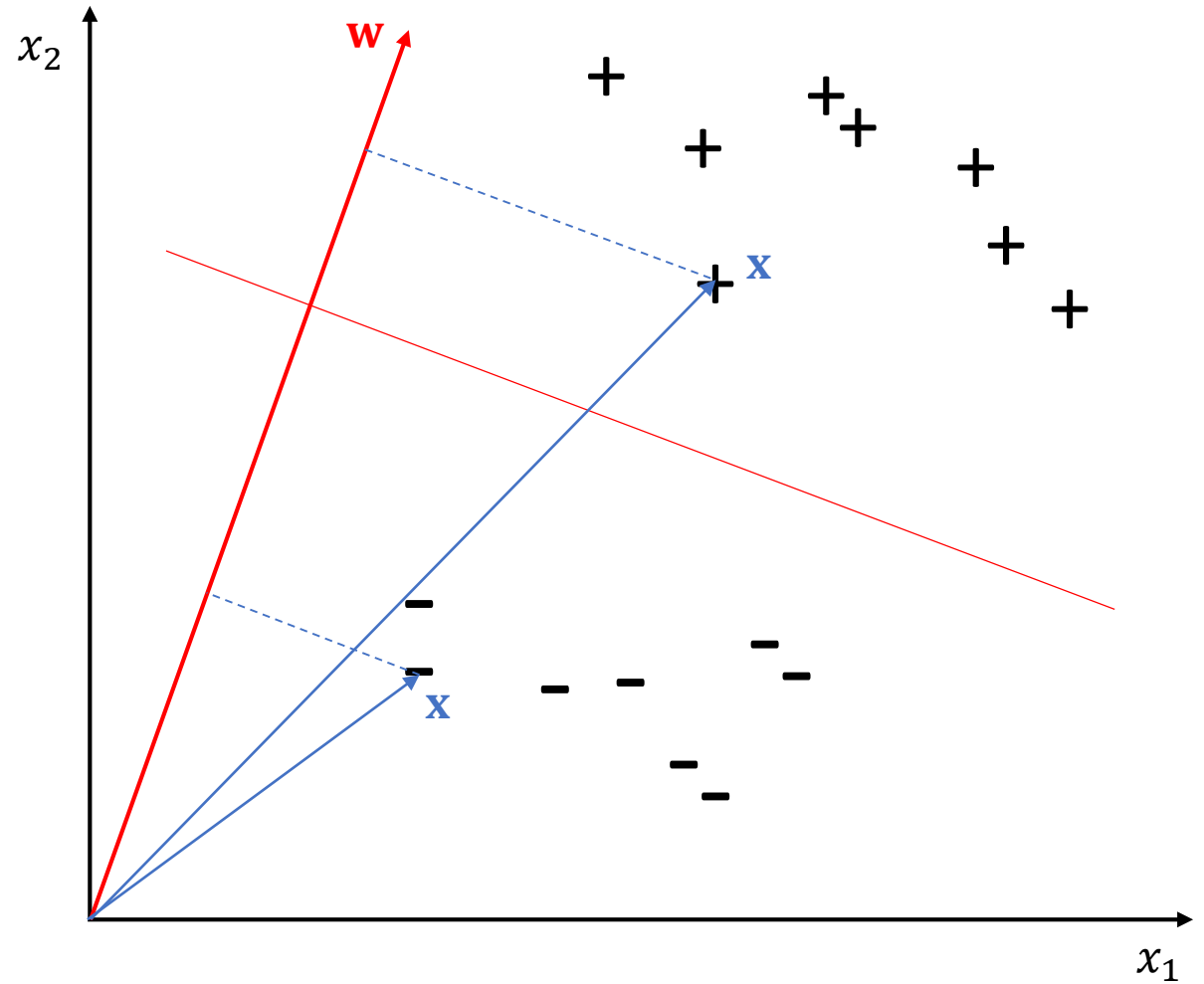
$$y \in \{-1, 1\}$$

$$h(\mathbf{x}) = \text{sign} \left( \left( \sum_{i=1}^d w_i x_i \right) - \text{threshold} \right)$$

$$h(\mathbf{x}) = \text{sign} \left( \left( \sum_{i=1}^d w_i x_i \right) - w_0 \right)$$

$$h(\mathbf{x}) = \text{sign} \left( \sum_{i=0}^d w_i x_i \right)$$

$$h(\mathbf{x}) = \text{sign}(\mathbf{w}^T \mathbf{x})$$



# Обучение перцептрона

$$h(\mathbf{x}) = \text{sign}(\mathbf{w}^T \mathbf{x})$$

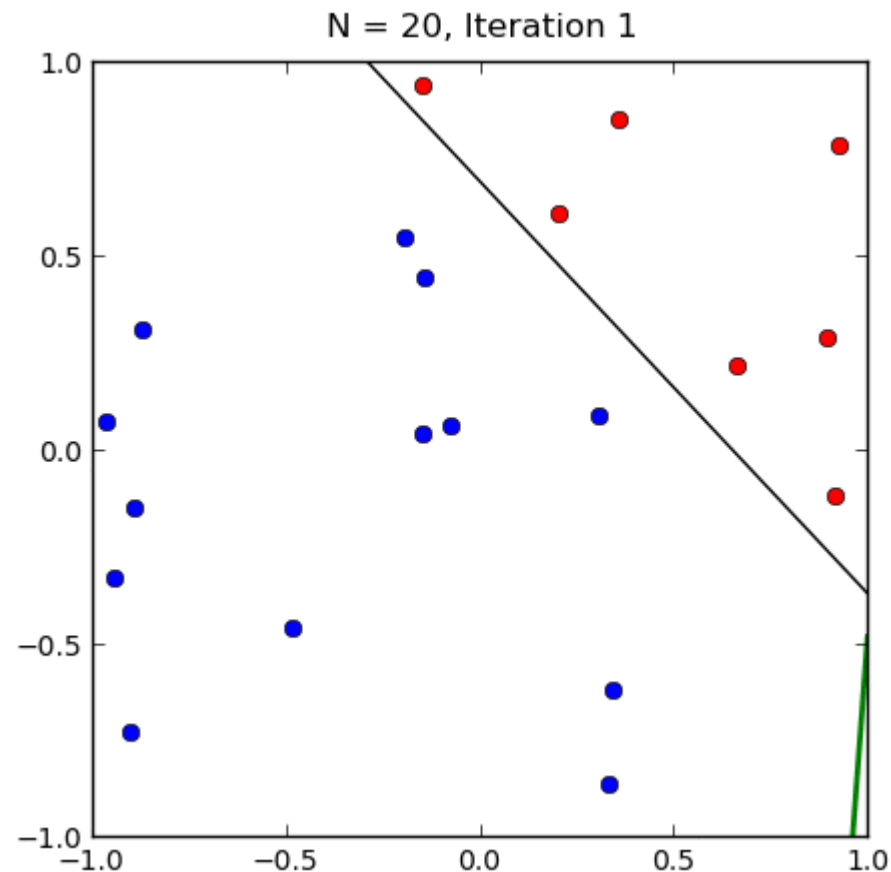
Data:  $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)$

Алгоритм:

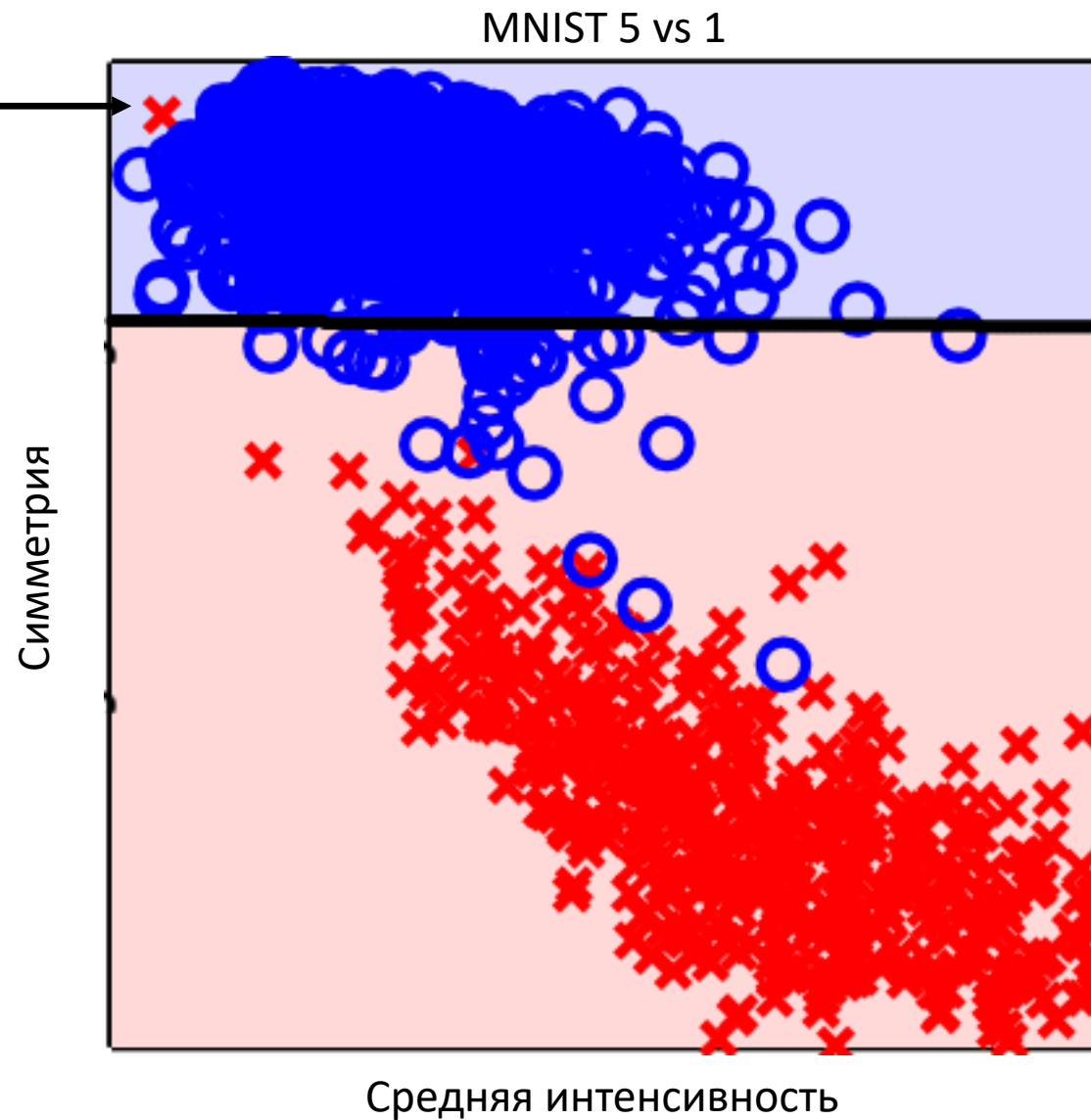
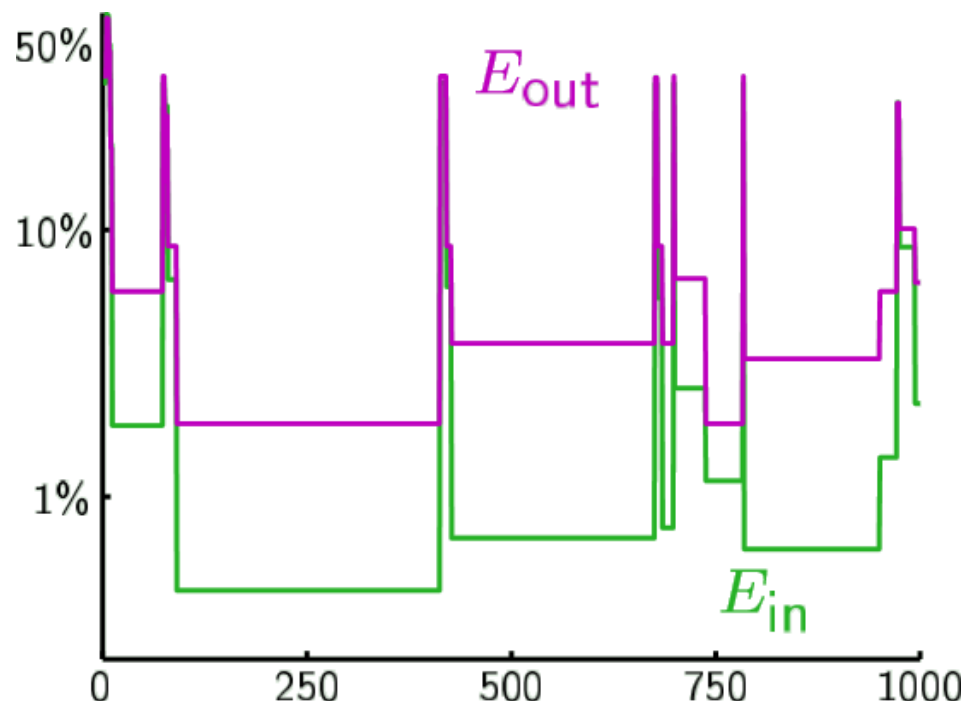
Начнем со случайного вектора  $\mathbf{w}$

Найдем такой  $\mathbf{x}_i$ , что  $h(\mathbf{x}_i) \neq y_i$

$$\mathbf{w} \leftarrow \mathbf{w} + y_i \mathbf{x}_i$$



# Перцептрон с карманом



# Подсчет ошибки

$$E_{in}(h) = \frac{1}{N} \sum_1^N e(h(x_n), f(x_n)) \quad \text{in sample (в выборке)}$$

$$E_{out}(h) = E_x[e(h(x), f(x))] \quad \text{out of sample (вне выборки)}$$

# Неравенство Хёфдинга

$$P[|E_{in}(h) - E_{out}(h)| > \epsilon] \leq 2e^{-\epsilon^2 N}$$

Ошибка обобщения  
Generalization error

Иногда ошибкой обобщения называют  $E_{out}$

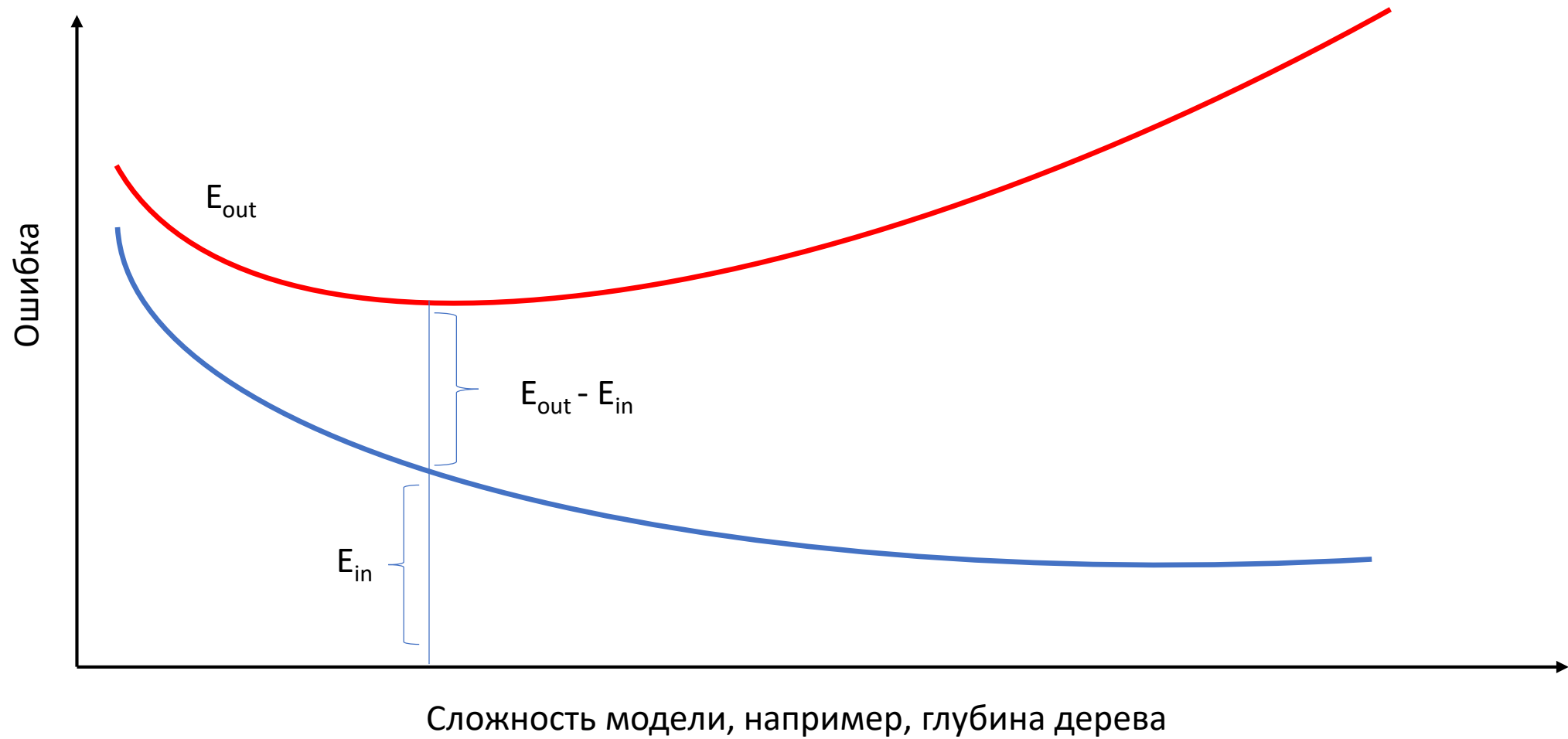
# Неравенство Хёфдинга

$$P[|E_{in}(h) - E_{out}(h)| > \epsilon] \leq 2e^{-\epsilon^2 N}$$

$$P[|E_{in}(h) - E_{out}(h)| > \epsilon] \leq M2e^{-\epsilon^2 N}$$

$M$  – количество гипотез

# Сложность модели

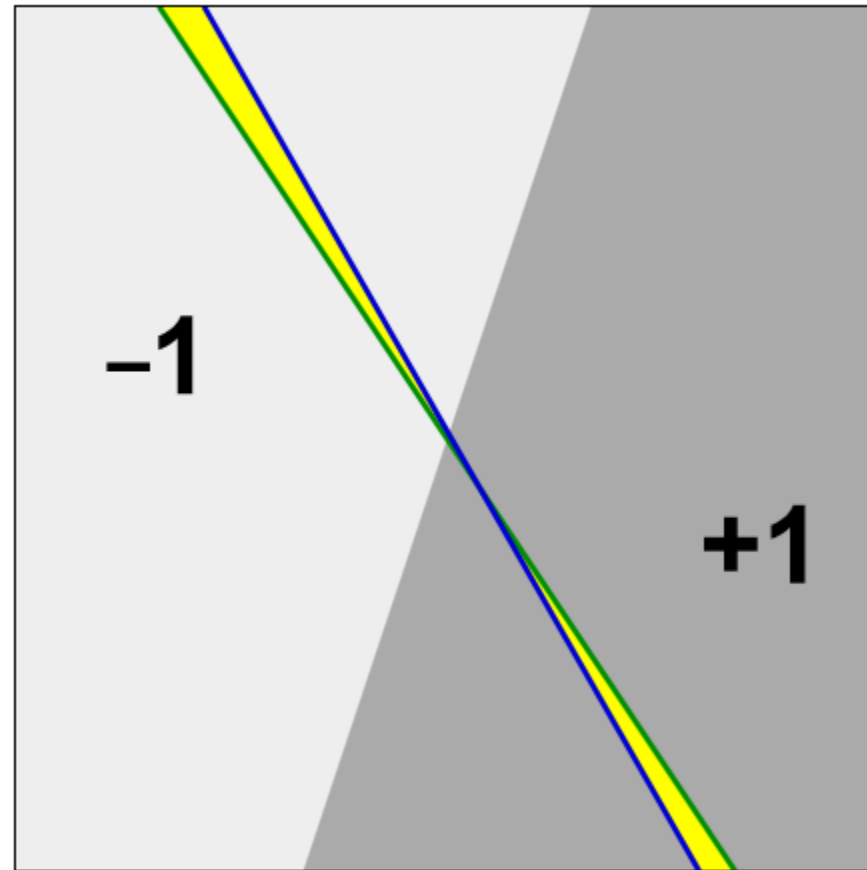




# Ближкие гипотезы

$$|E_{\text{in}}(h_1) - E_{\text{out}}(h_1)| \approx |E_{\text{in}}(h_2) - E_{\text{out}}(h_2)|$$

up



down

# От гипотез к дихотомиям

- Гипотеза:  $h: X \rightarrow \{-1, +1\}$
- Дихотомия:  $h: \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \rightarrow \{-1, +1\}$
- Максимальное количество дихотомий:  $2^N$

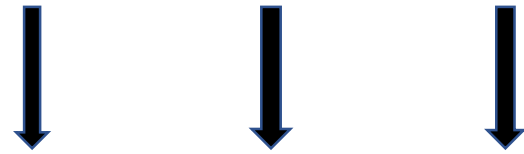
Функция роста (growth function)

$$m_H(N) = \max_{x_1, \dots, x_N} |H(\mathbf{x}_1, \dots, \mathbf{x}_N)|$$

$$m_H(N) \leq 2^N$$

# Неравенство Вапника-Червоненкиса

$$P[|E_{in}(h) - E_{out}(h)| > \epsilon] \leq M2e^{-\epsilon^2 N}$$



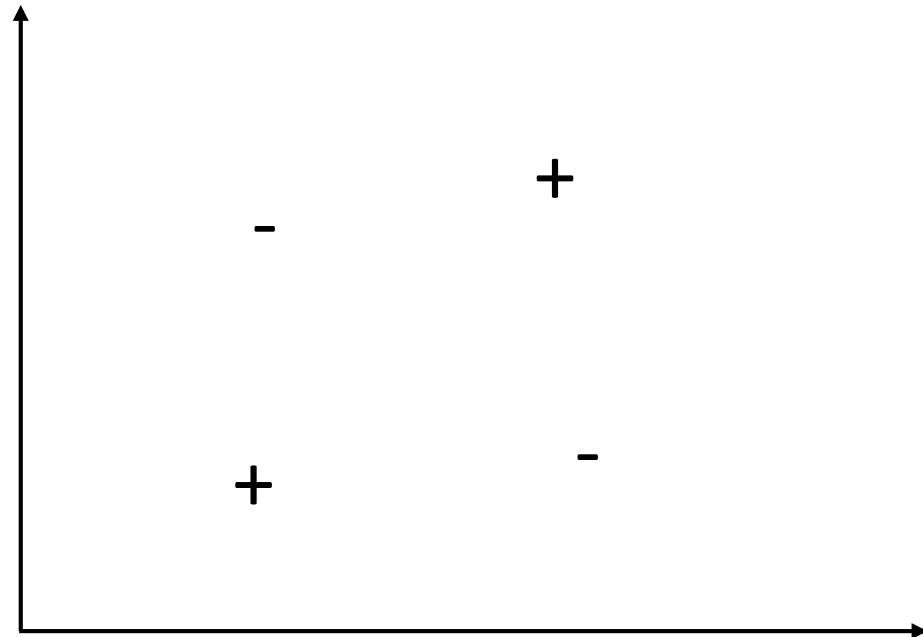
$$P[|E_{in}(h) - E_{out}(h)| > \epsilon] \leq m_H(\textcolor{red}{2}N)\textcolor{red}{4}e^{-\epsilon^{\textcolor{red}{1}\over 8}N}$$

# Точка “поломки” (Breakpoint)

Определение: если никакой набор данных размера  $k$  нельзя распределить на все возможные случаи набором гипотез  $H$ , то  $k$  – точка поломки для  $H$ .

$$\min(k: m_H(k) < 2^k)$$

Для 2D перцептрона,  $k = 4$ .



# Доказательство полиномиальности функции роста в присутствии точки поломки

- $B(N, k) = m_H(N)$  с точкой поломки  $k$
- $B(N, k) = \alpha + 2\beta$
- $\alpha + \beta \leq B(N - 1, k)$
- $\beta \leq B(N - 1, k - 1)$
- $B(N, k) \leq B(N - 1, k) + B(N - 1, k - 1)$
- Докажем, что  $B(N, k) \leq \sum_{i=0}^{k-1} C_N^i$

|          | $x_1$    | $x_2$    | $\dots$  | $x_{N-1}$ | $x_N$    |                      |
|----------|----------|----------|----------|-----------|----------|----------------------|
| $\alpha$ | +1       | +1       | $\dots$  | +1        | +1       | 1 вариант для $x_N$  |
|          | -1       | +1       | $\dots$  | +1        | -1       |                      |
|          | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$  | $\vdots$ |                      |
|          | +1       | -1       | $\dots$  | -1        | -1       |                      |
|          | -1       | +1       | $\dots$  | -1        | +1       |                      |
| $\beta$  | +1       | -1       | $\dots$  | +1        | +1       | 2 варианта для $x_N$ |
|          | -1       | -1       | $\dots$  | +1        | +1       |                      |
|          | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$  | $\vdots$ |                      |
|          | +1       | -1       | $\dots$  | +1        | +1       |                      |
|          | -1       | -1       | $\dots$  | -1        | +1       |                      |
| $\beta$  | +1       | -1       | $\dots$  | +1        | -1       |                      |
|          | -1       | -1       | $\dots$  | +1        | -1       |                      |
|          | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$  | $\vdots$ |                      |
|          | +1       | -1       | $\dots$  | +1        | -1       |                      |
|          | -1       | -1       | $\dots$  | -1        | -1       |                      |

# Индукция

$$B(N, k) \leq \sum_{i=0}^{k-1} C_N^i$$

$$B(N, 1) = 1, \quad B(1, k > 1) = 2$$

$$B(N, k) \leq B(N-1, k) + B(N-1, k-1) \leq \sum_{i=0}^{k-1} C_{N-1}^i + \sum_{i=0}^{k-2} C_{N-1}^i$$

$$= 1 + \sum_{i=1}^{k-1} C_N^i + \sum_{i=1}^{k-1} C_{N-1}^{i-1} = 1 + \sum_{i=1}^{k-1} (C_{N-1}^i + C_{N-1}^{i-1}) = 1 + \sum_{i=1}^{k-1} C_N^i = \sum_{i=0}^{k-1} C_N^i$$

# Размерность Вапника-Червоненкиса

$d_{VC}(H)$  для набора гипотез  $H$ , это наибольшее значение  $N$ ,  
для которого  $m_H(N) = 2^N$ .

$d_{VC}(H) = k - 1$ , где  $k$  — точка поломки

# Функция роста и VC-размерность

$$m_H(N) \leq \sum_{i=0}^{k-1} C_N^i$$

$$m_H(N) \leq \sum_{i=0}^{d_{VC}} C_N^i \leq N^{d_{VC}} + 1$$



# VC-размерность для перцептрона

Для размерности  $d$ ,  $d_{VC} = d + 1$

$$\mathbf{X} = \begin{bmatrix} -\mathbf{x}_1^T & - \\ -\mathbf{x}_2^T & - \\ -\mathbf{x}_3^T & - \\ \vdots & \\ -\mathbf{x}_d^T & - \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ & & \vdots & & & \\ 1 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

$\text{sign}(\mathbf{X}\mathbf{w}) = \mathbf{y}$

$\mathbf{X}\mathbf{w} = \mathbf{y}$

$\mathbf{w} = \mathbf{X}^{-1}\mathbf{y}$

$\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_{d+1}, \mathbf{x}_{d+2}$

$$\mathbf{x}_j = \sum_{i \neq j} \mathbf{x}_i a_i \quad \mathbf{w}^T \mathbf{x}_j = \sum_{i \neq j} \mathbf{w}^T \mathbf{x}_i a_i \quad y_i = \text{sign}(a_i) \quad y_j = -1$$

Сколько нужно данных?

$$P[|E_{in}(h) - E_{out}(h)| > \epsilon] \leq m_H(2N) 4e^{-\epsilon^{\frac{1}{8}}N}$$

$$\approx N^{d_{VC}} e^{-N}$$

$$N \geq 10 d_{VC}$$

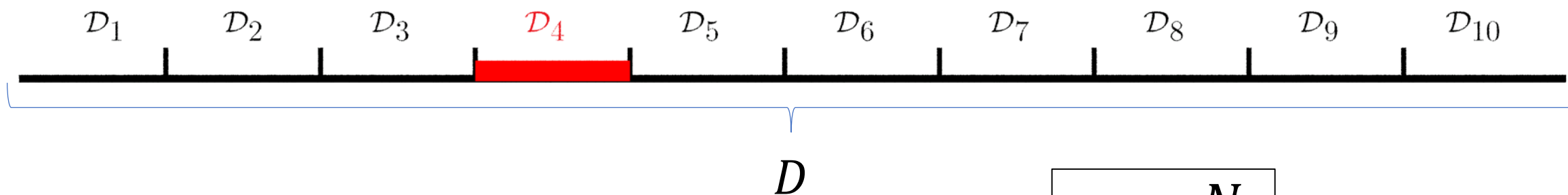
# Валидация

$$(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_K, \mathbf{y}_K) \in \mathcal{D}_{val} \quad E_{val}(h) = \frac{1}{K} \sum_1^K e(h(\mathbf{x}_k), f(\mathbf{x}_k))$$

$$P[|E_{val}(h) - E_{out}(h)| > \epsilon] \leq 2e^{-\epsilon^2 N}$$

$$K = \frac{N}{5}$$

# Кросс-валидация



$$K = \frac{N}{10}$$

# Train-Val-Test

- Обучаем алгоритм на **train**
- Оптимизируем алгоритм (гиперпараметры) на **val (cross-val)**
- Проверяем алгоритм на **test**