



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

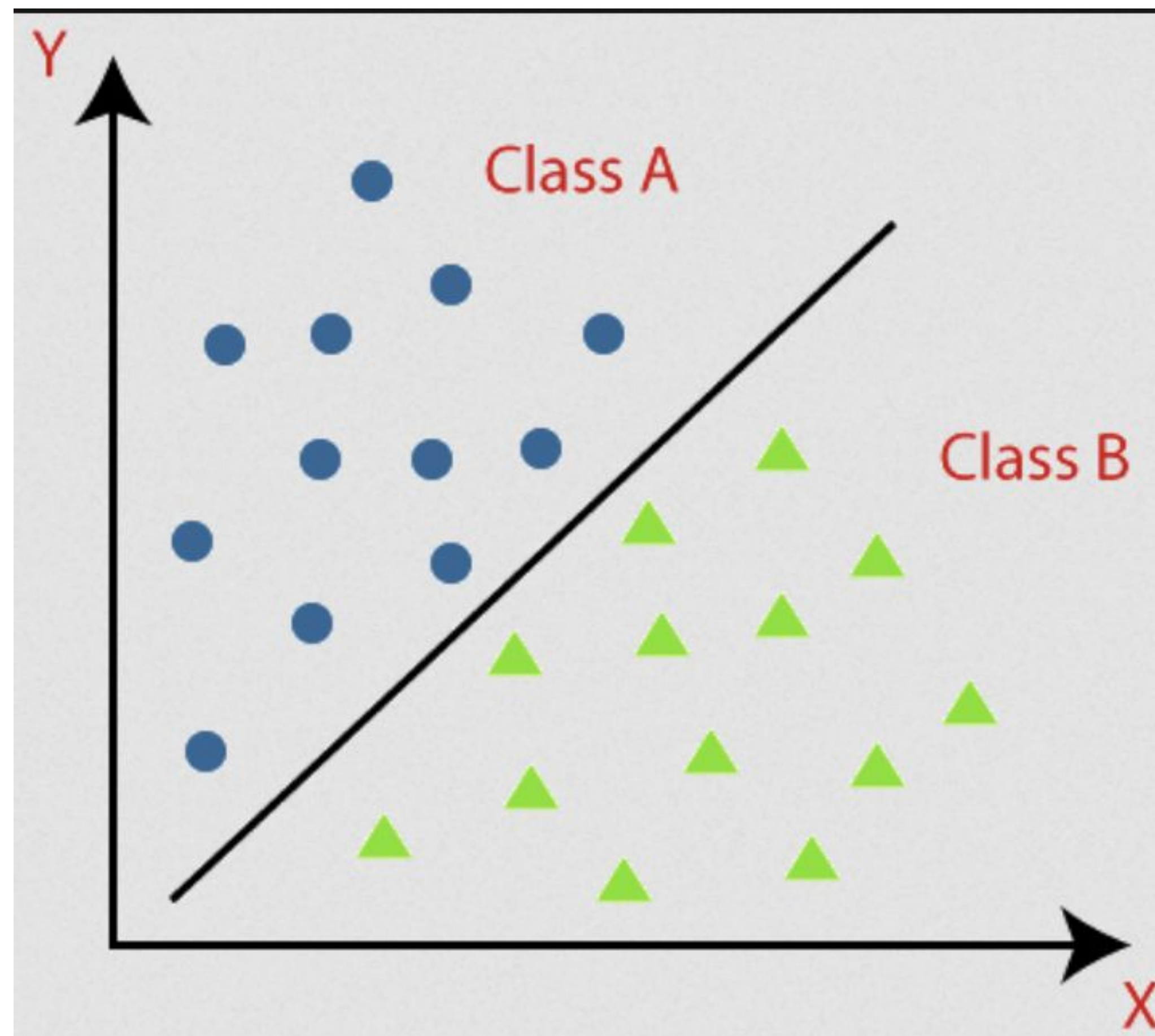
Факультет социальных наук

# Бинарная классификация. Подходы и метрики

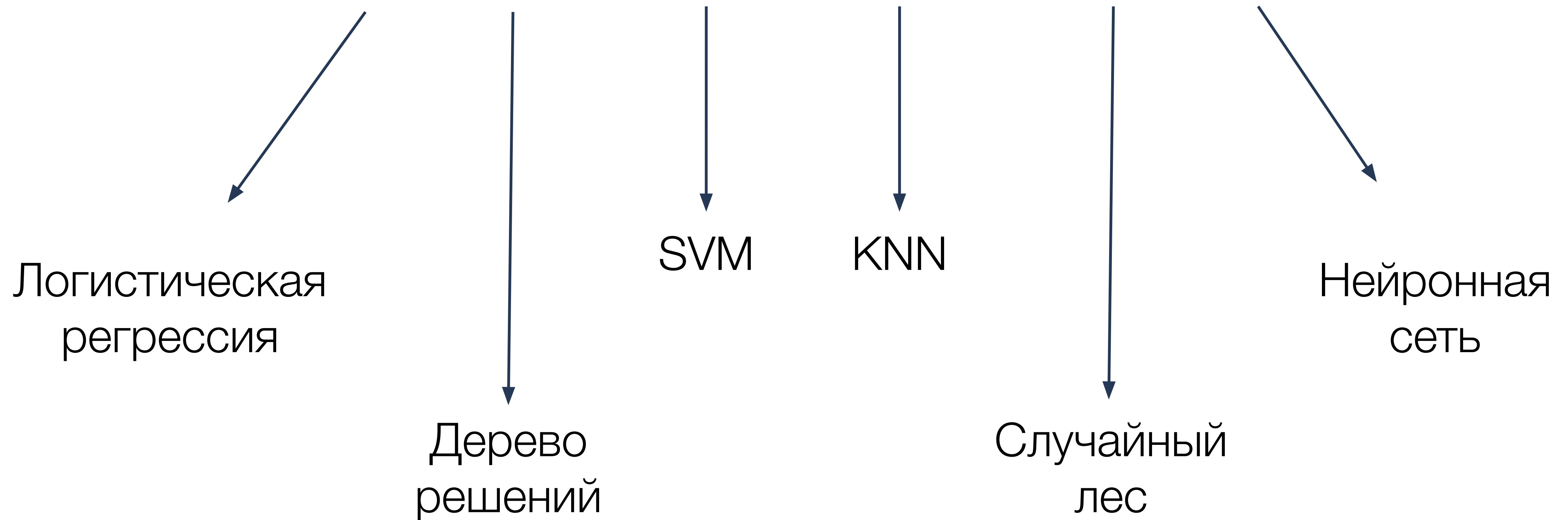
Москва, 2023



# Задача бинарной классификации



# Модели бинарной классификации





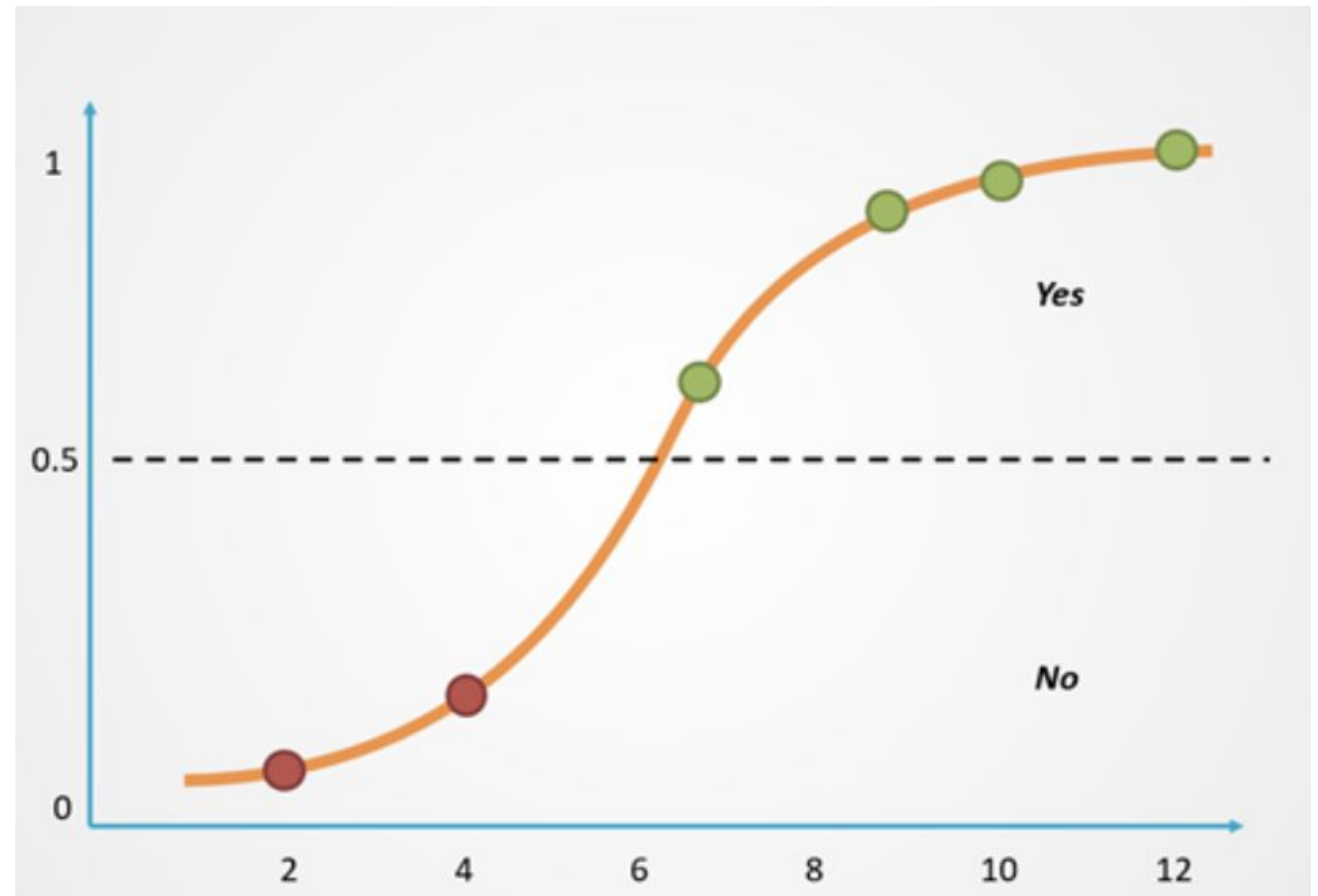
# Логистическая регрессия

Предсказание модели:

$$\frac{1}{1 + \exp(-\langle w, x \rangle)}$$

$w$  - веса модели

$x$  - входные параметры



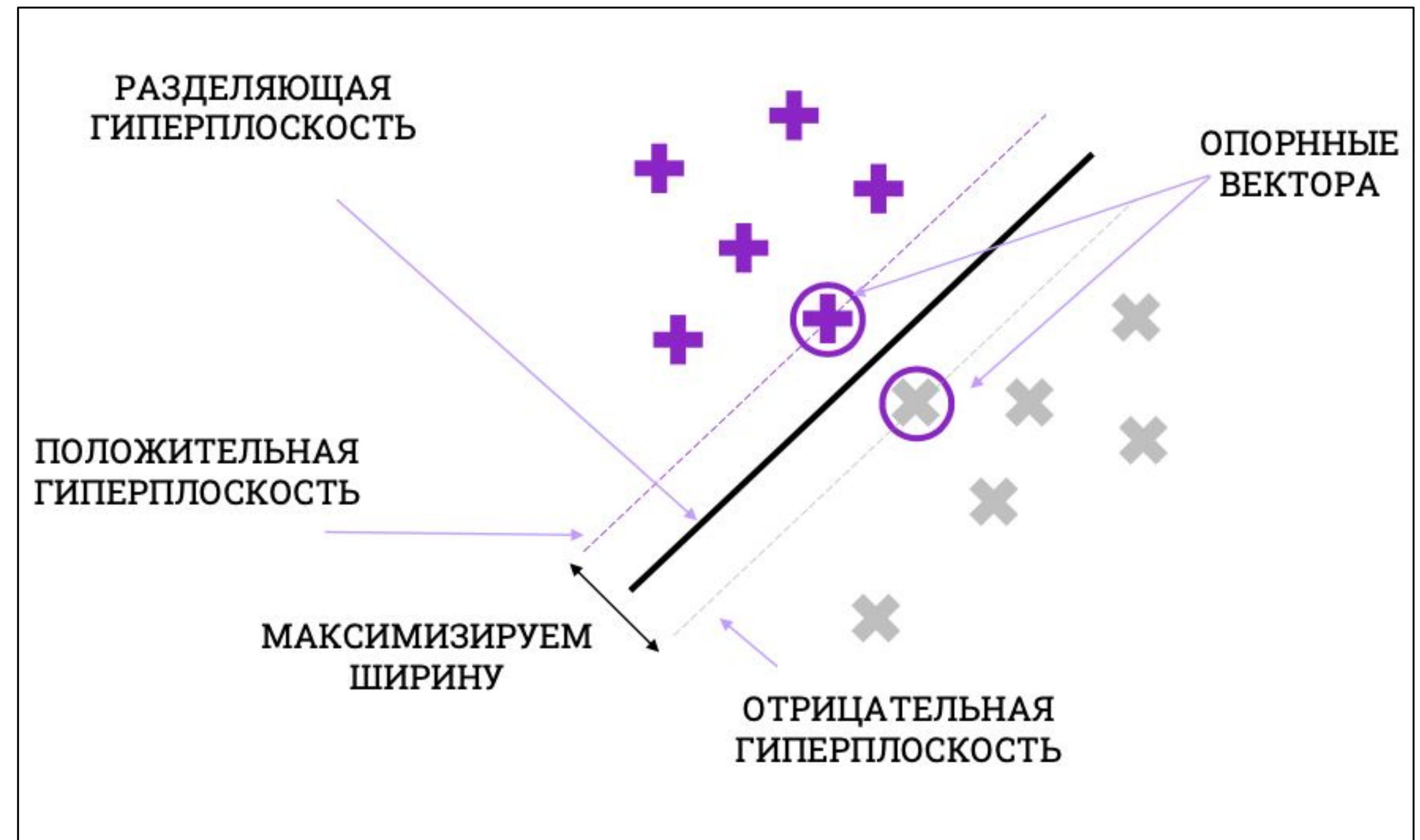
# SVM

Предсказание модели:

$$\text{sign}(\langle w, x \rangle + b)$$

$w$  - веса модели

$x$  - входные параметры





# Дерево решений

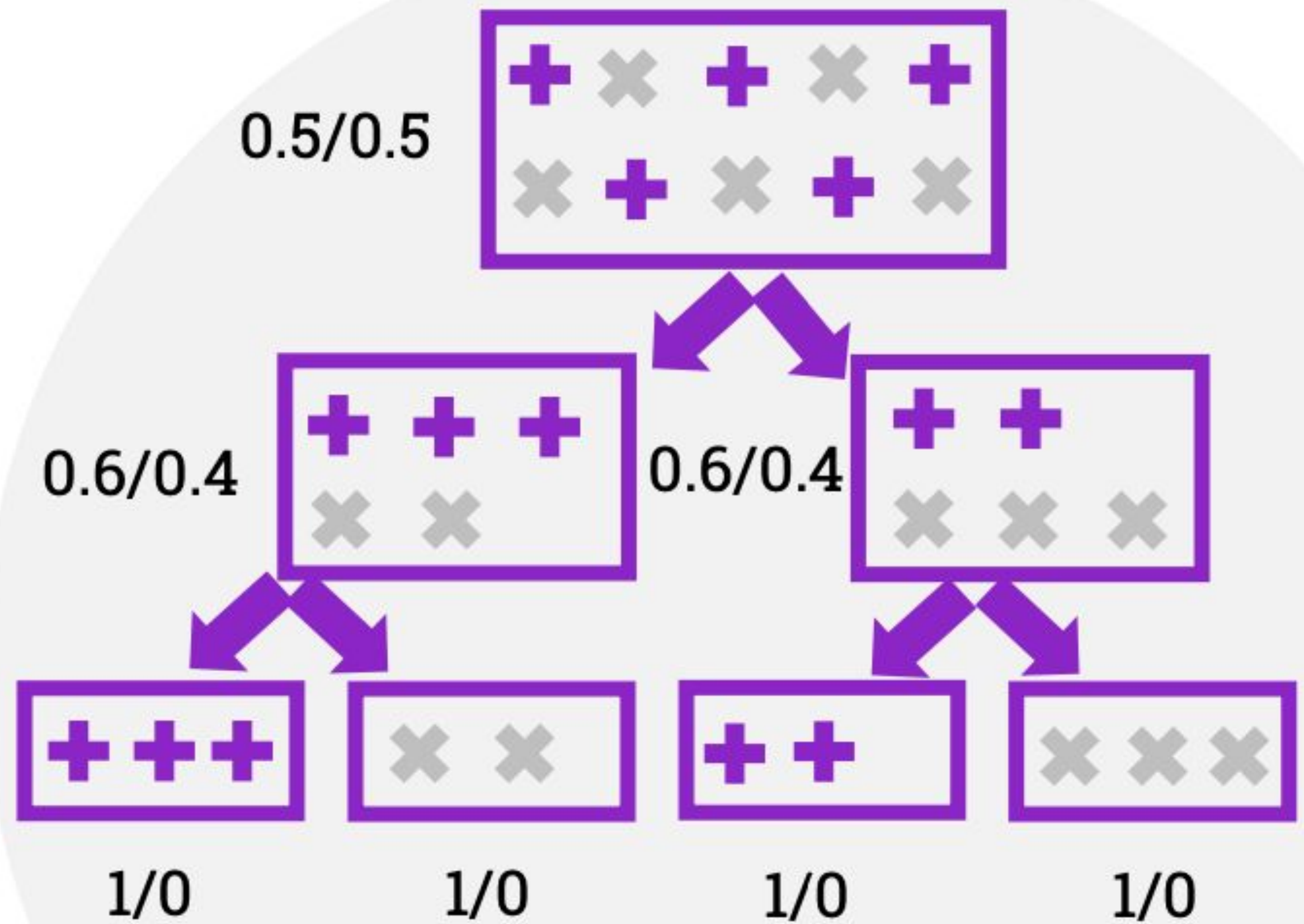
## МИНИМИЗИРУЕМ

$$E = \sum_{i=1}^c -p_i * \log(p_i)$$

**E1:**  $-0.5 * \log(0.5) + -0.5 * \log(0.5) = 1$

**E2:**  $-0.6 * \log(0.6) + (-0.4 * \log(0.4)) = 0.94$

**E3:**  $1 * \log(1) + 0 * \log(0) = 0$



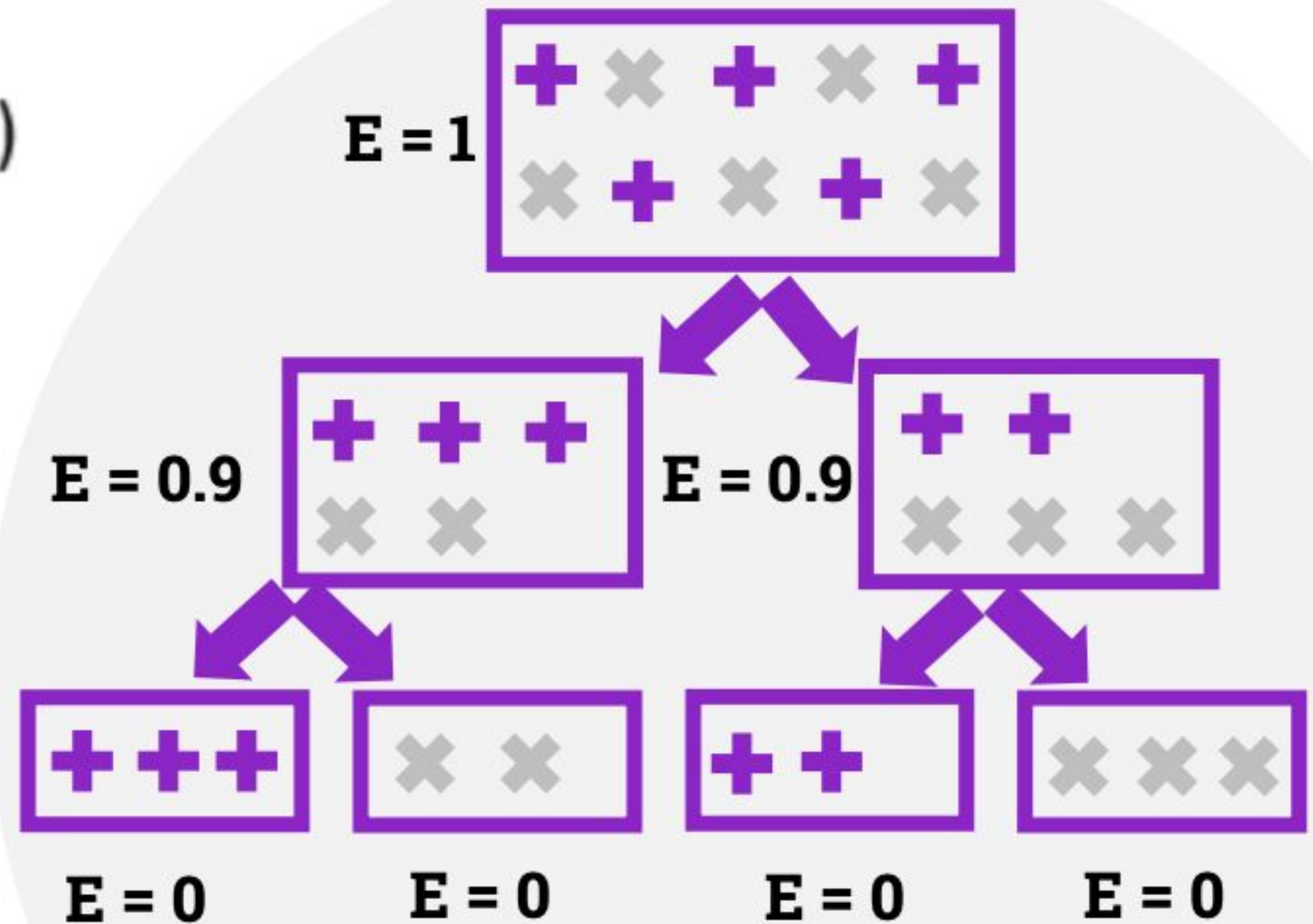
# Дерево решений

## МАКСИМИЗИРУЕМ

$$\text{Information Gain} = E(Y) - E(Y|X)$$

$$\text{IG1: } ((1-0.9)+(1-0.9))/2 = 0.1$$

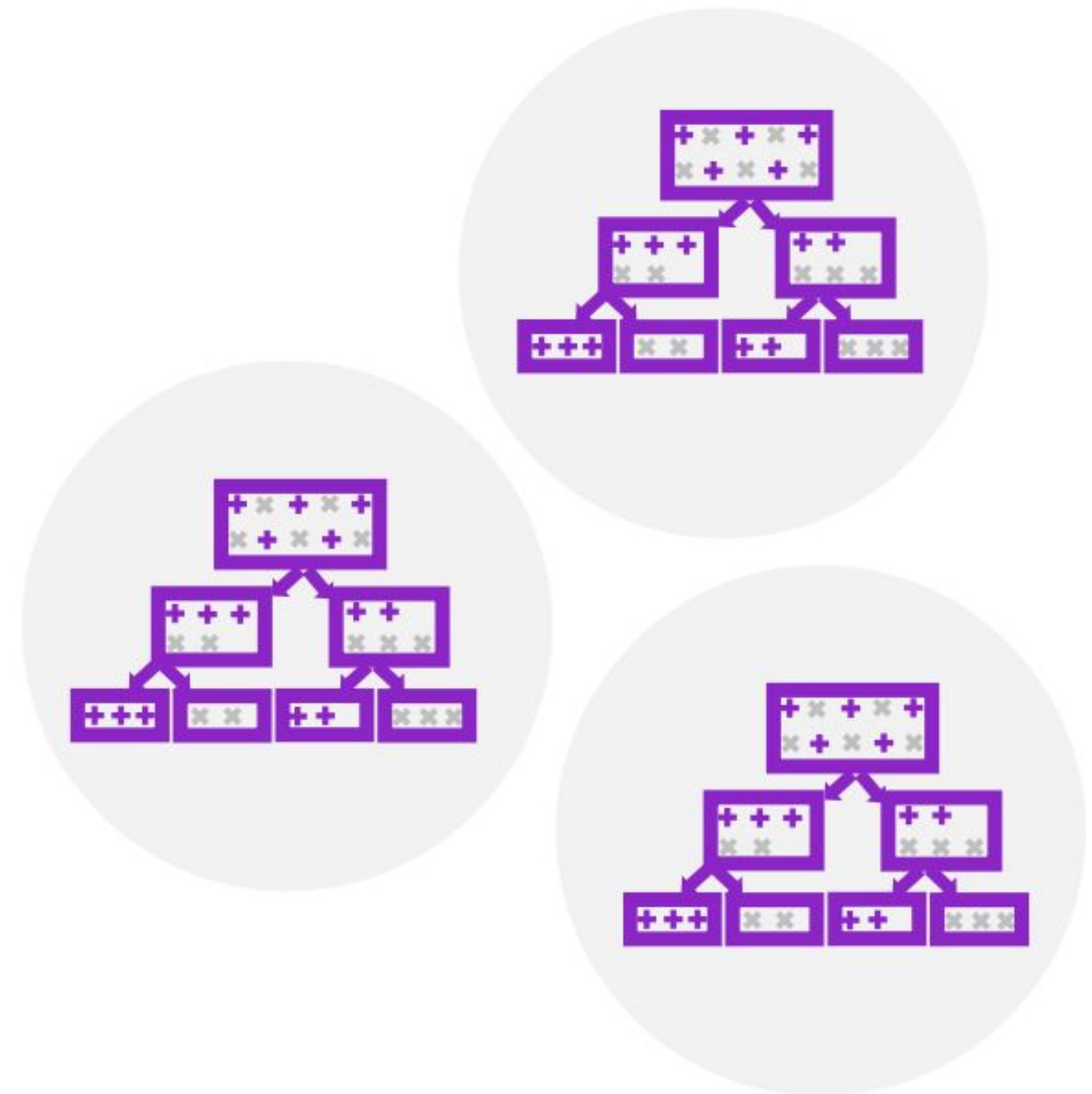
$$\text{IG2: } ((0.9-0)+(0.9-0))/2 = 0.9$$





# Случайный лес

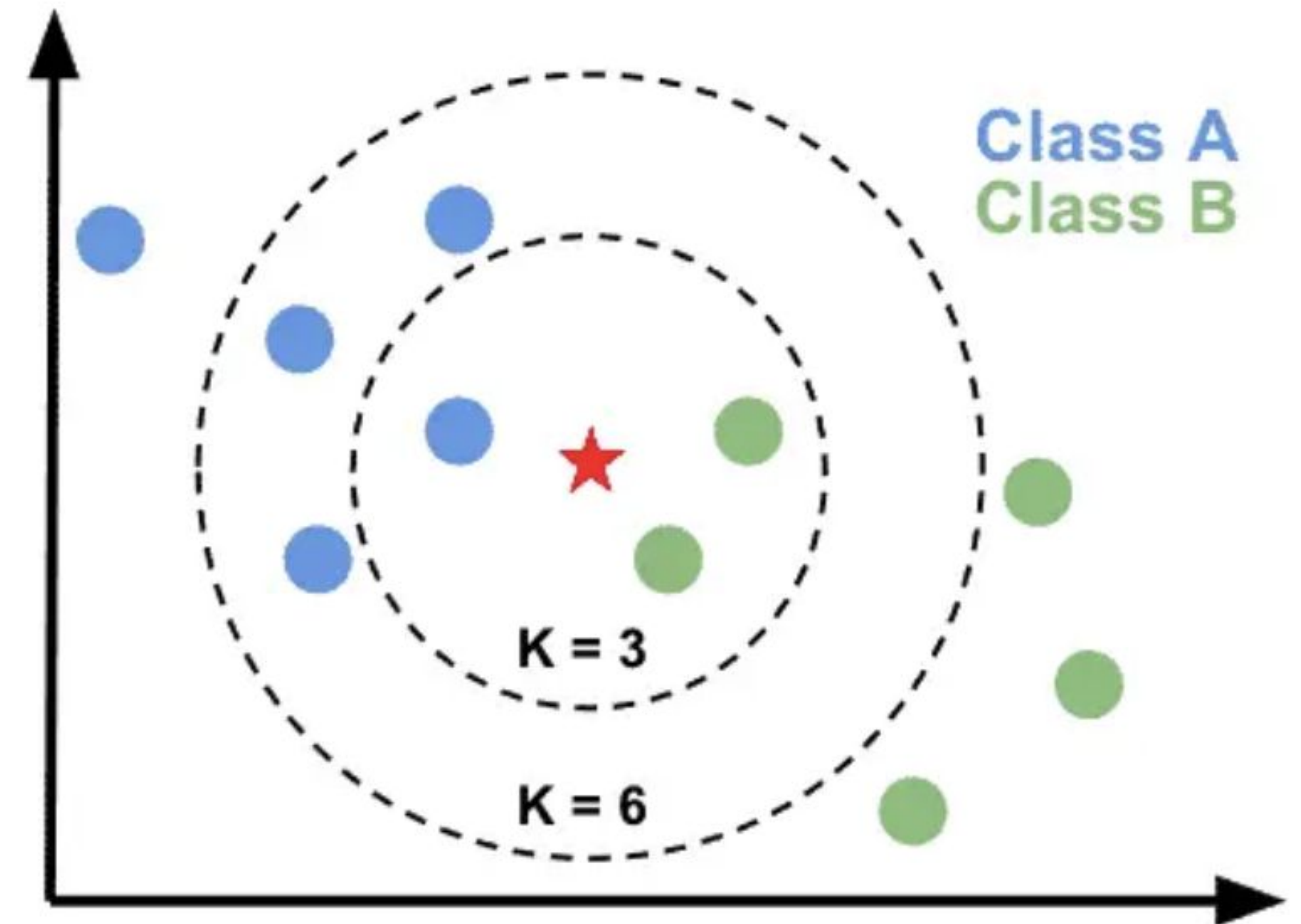
1. Разделяем выборку на подвыборки
2. Обучаем на каждой подвыборке дерево
3. Решение – по мнению большинства



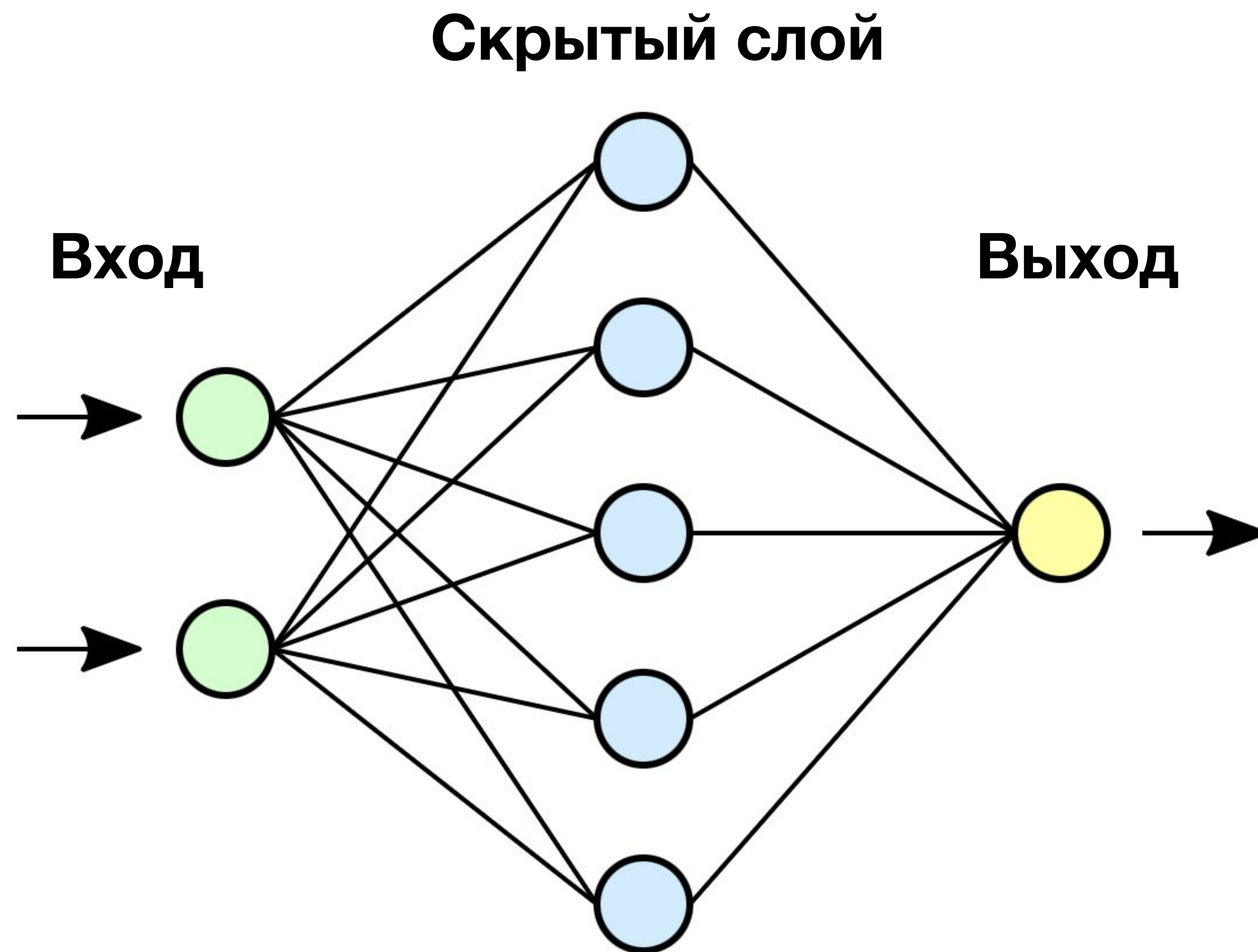


# KNN

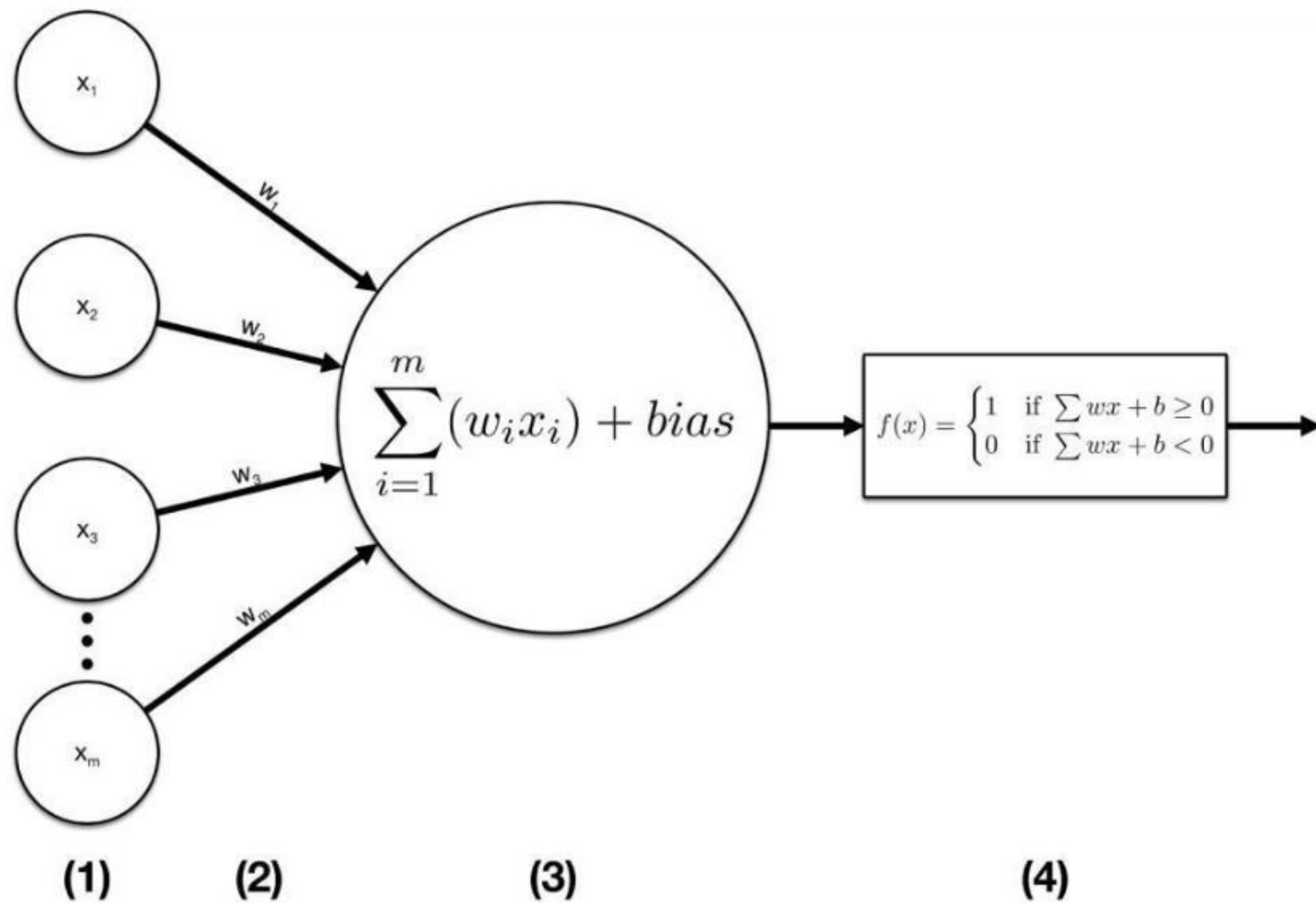
1. Считаем расстояние всех объектов
2. Находим ближайшие
3. Решение – по мнению большинства



# Нейронная сеть



# Нейронная сеть





Как измерить качество ?





# Матрица ошибок

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive (TP)	False Negative (FN) <b>Type II Error</b>
	Negative	False Positive (FP) <b>Type I Error</b>	True Negative (TN)

# Accuracy

Смотрим на совпадение меток (процент правильных)



100%



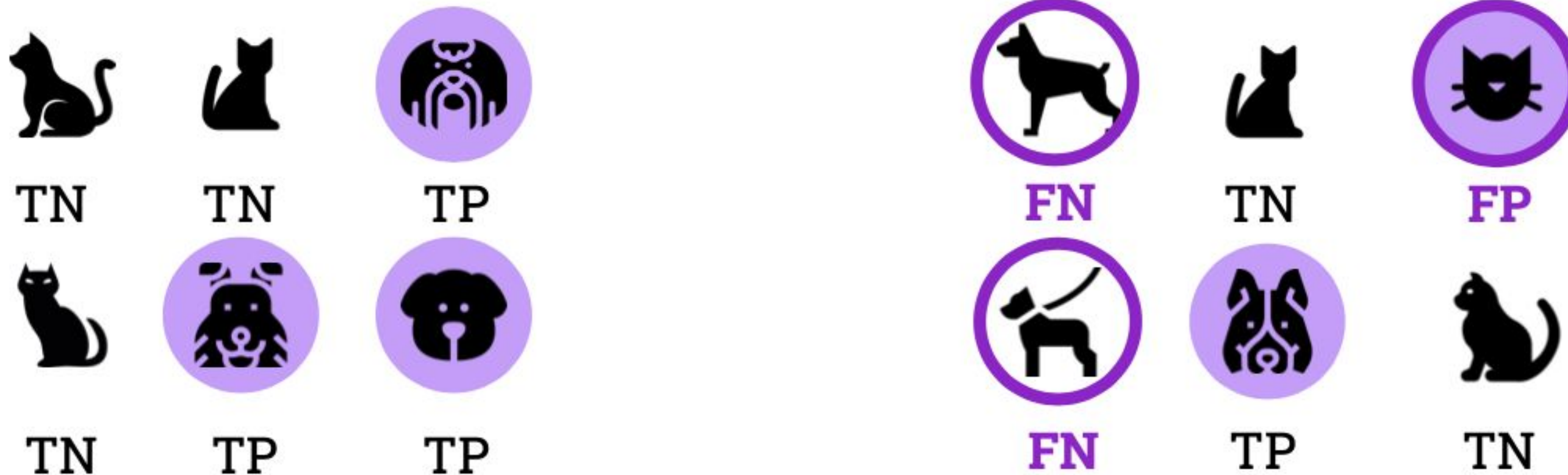
50%



## Recall (полнота)

Какая часть положительных предсказана верно

$$\text{TP}/(\text{TP}+\text{FN})$$



100%

33%

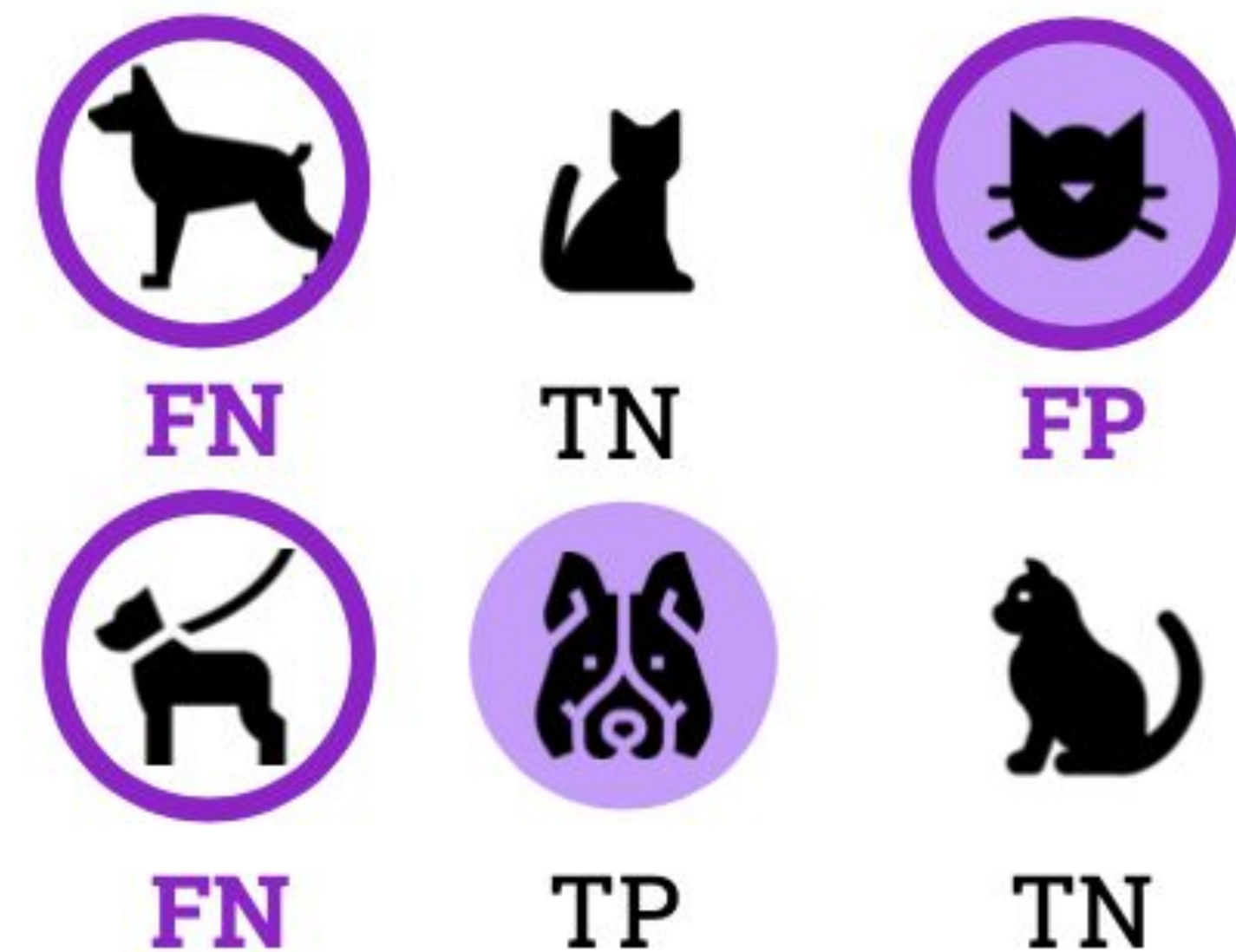
# Precision (точность)

Какая доля предсказанных положительных действительно положительные

$$\text{TP}/(\text{TP}+\text{FP})$$



100%



50%



# F-score

Общий критерий качества на основе полноты и точности

$$F_{\beta} = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}$$

**ПРИ БЕТА = 1:**

$$F = 2 \cdot \frac{(0.5 \cdot 0.33)}{(2 \cdot 0.5) + 0.33} = 2 \cdot \frac{0.165}{1.33} =$$

**0.25**



**R -33%; P – 50%**

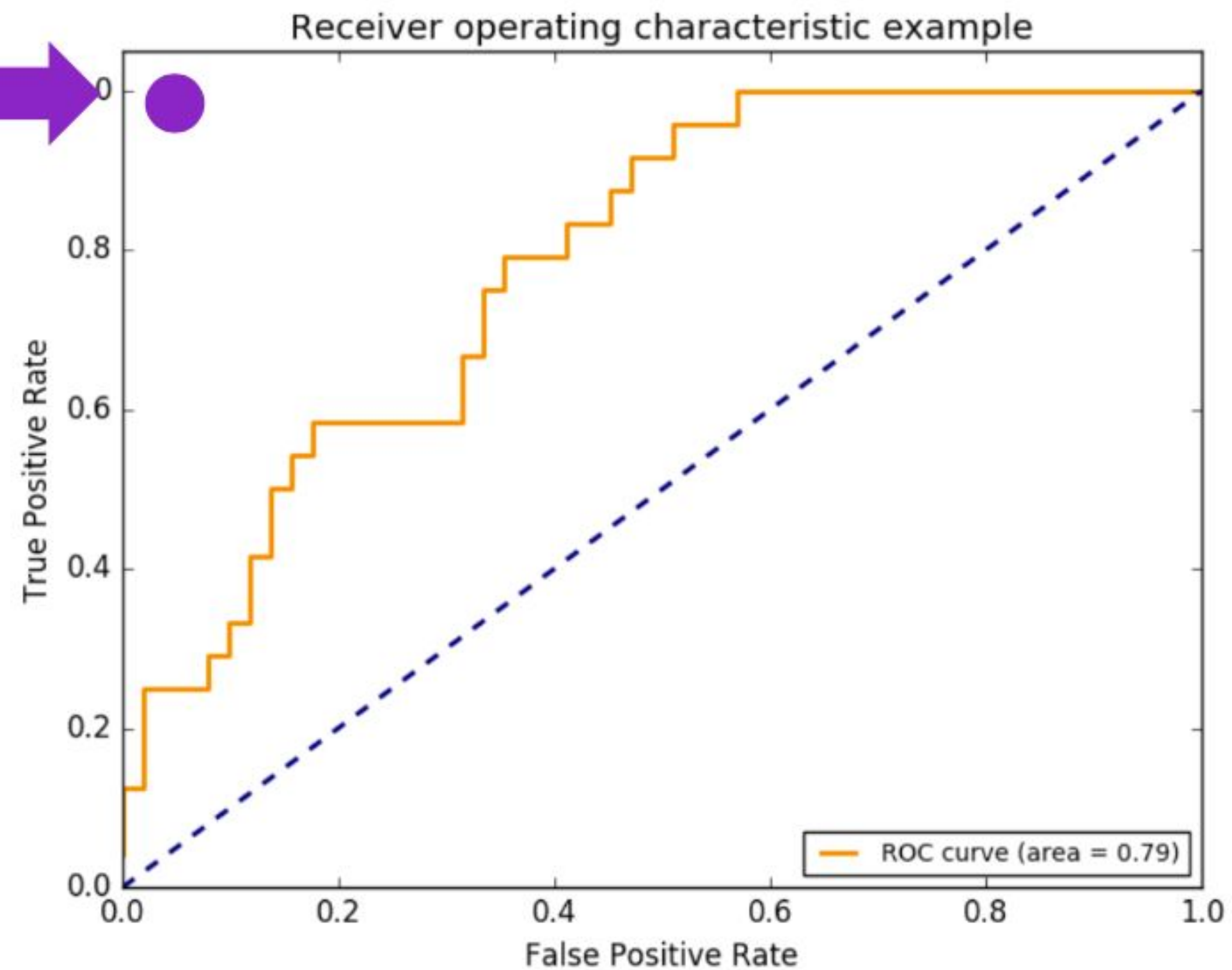
# AUC-ROC

**AUC**

ПЛОЩАДЬ ПОД  
КРИВОЙ

ИДЕАЛЬНО

МАКСИМИЗИРУЕМ



МИНИМИЗИРУЕМ