

# Линейный классификатор

Елена Кантонистова

Skillbox

# Елена Кантонистова

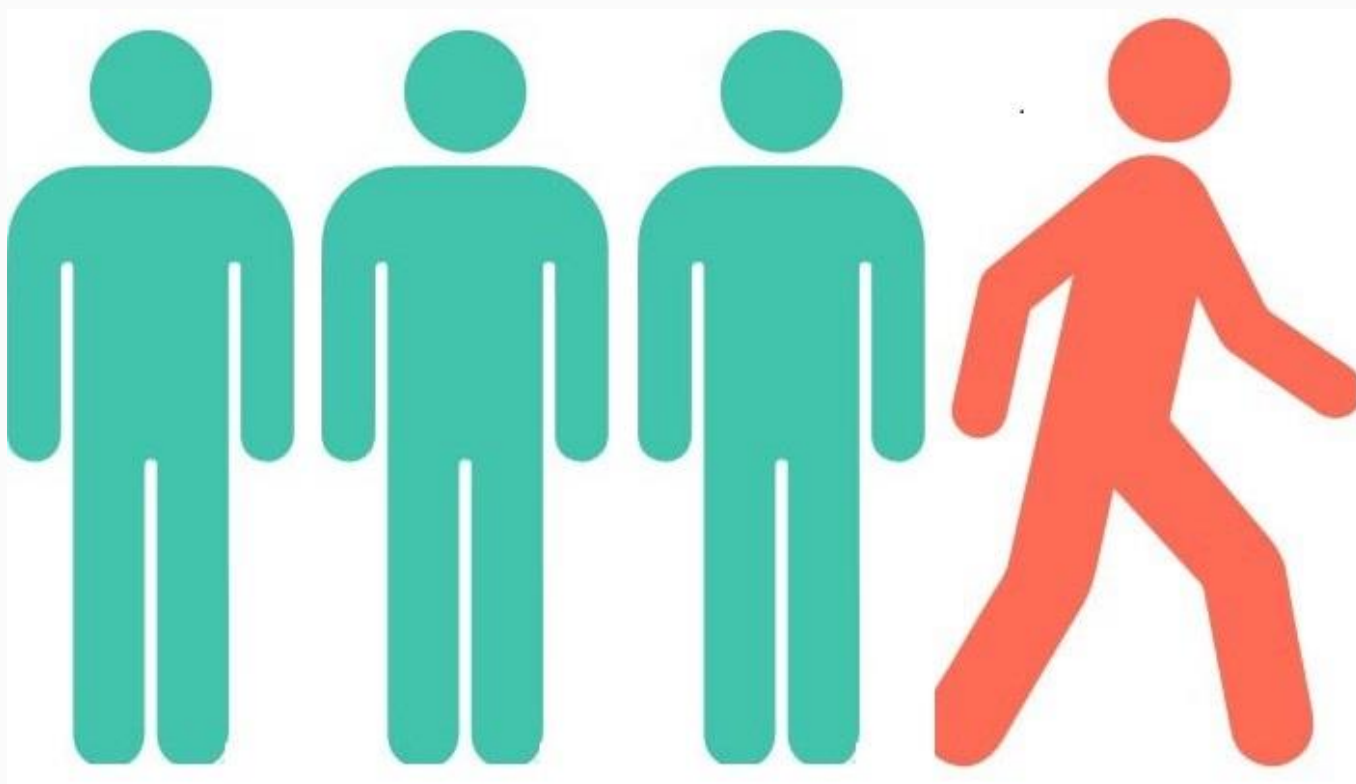
- Кандидат физико-математических наук
- Выпускница Школы Анализа Данных (ШАД) Яндекса
- Доцент факультета компьютерных наук ВШЭ
- Академический руководитель магистратуры «Машинное обучение и высоконагруженные системы» ФКН ВШЭ

Предыдущий индустриальный опыт:

- Data scientist в Raxel Telematics
- Data scientist в United Consulting Group

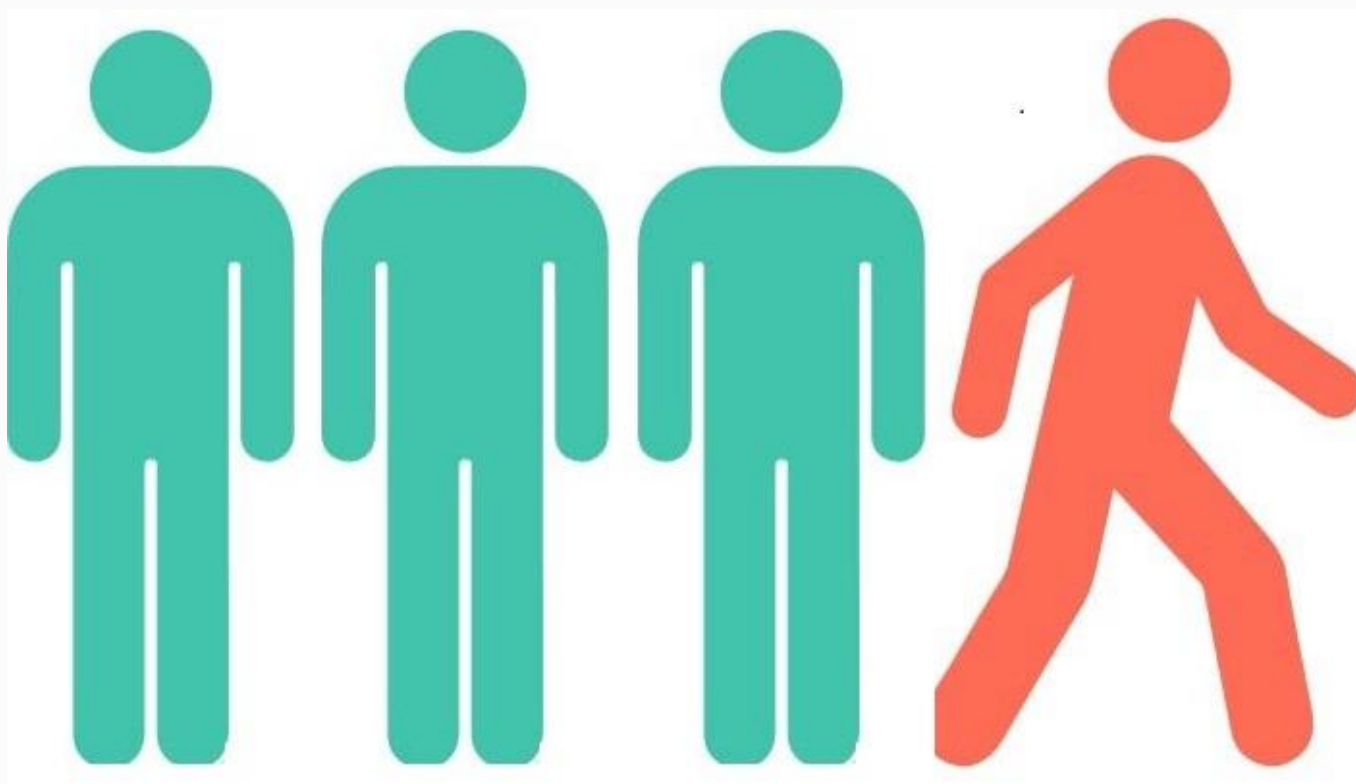
# Модель оттока

Модель оттока — это классификатор, который для каждого клиента в выбранный момент времени предсказывает, уйдёт ли клиент из компании и с какой вероятностью.



# ОТТОКОВЫЕ КЛИЕНТЫ

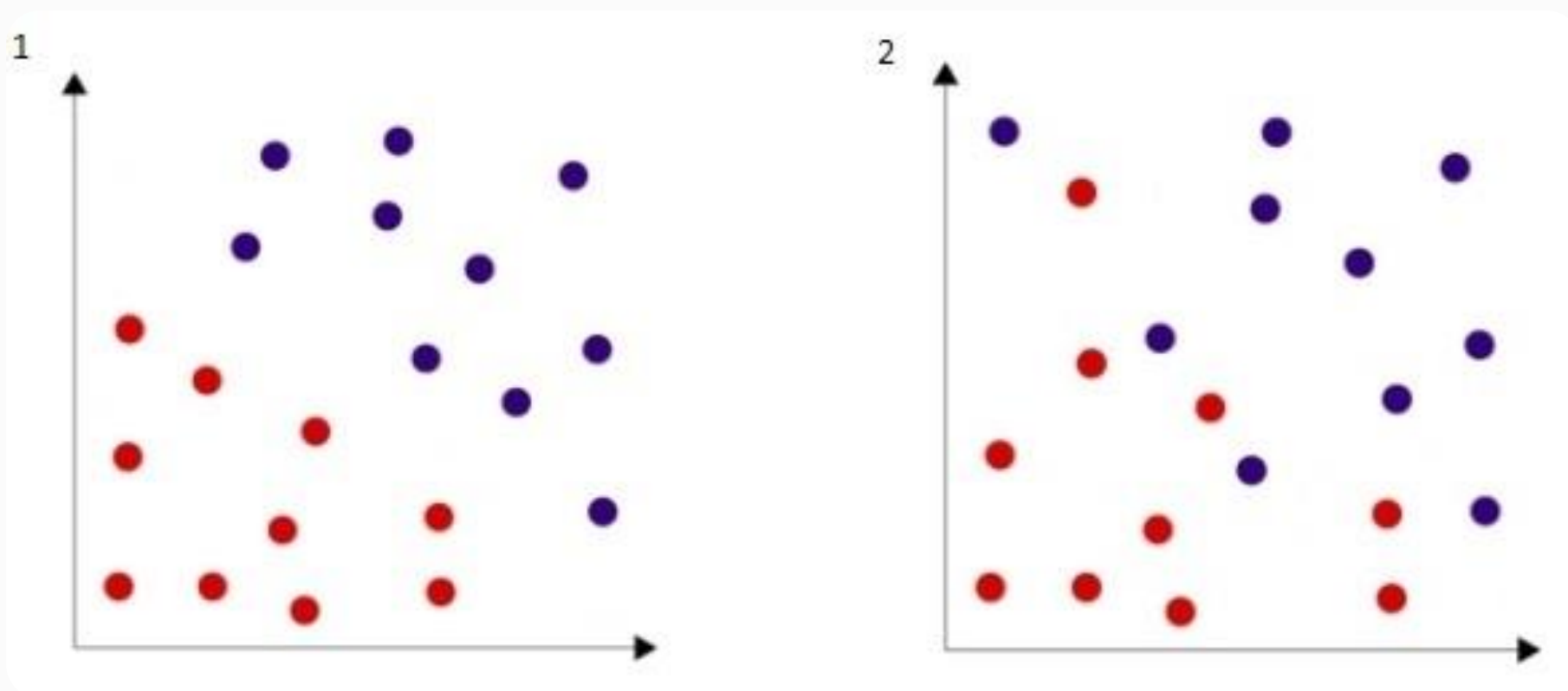
- Есть данные о клиентах компании
- Хочется построить модель предсказания оттока
- Метод ближайших соседей может показать неудовлетворительные результаты



# В этом модуле

- 1 Линейный классификатор
- 2 Логистическая регрессия
- 3 Градиентный спуск

# Два примера задач

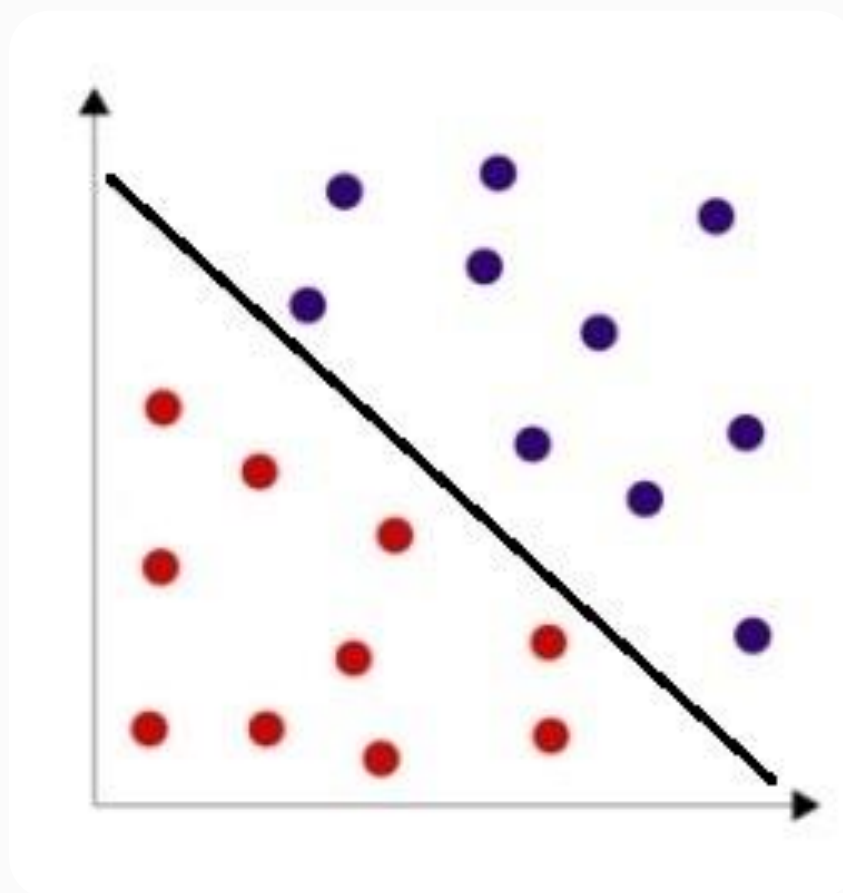


Изображение: [Два множества, разделённых линейно \(ru.wikipedia.org\)](http://ru.wikipedia.org)

Линейный классификатор

# Первая выборка

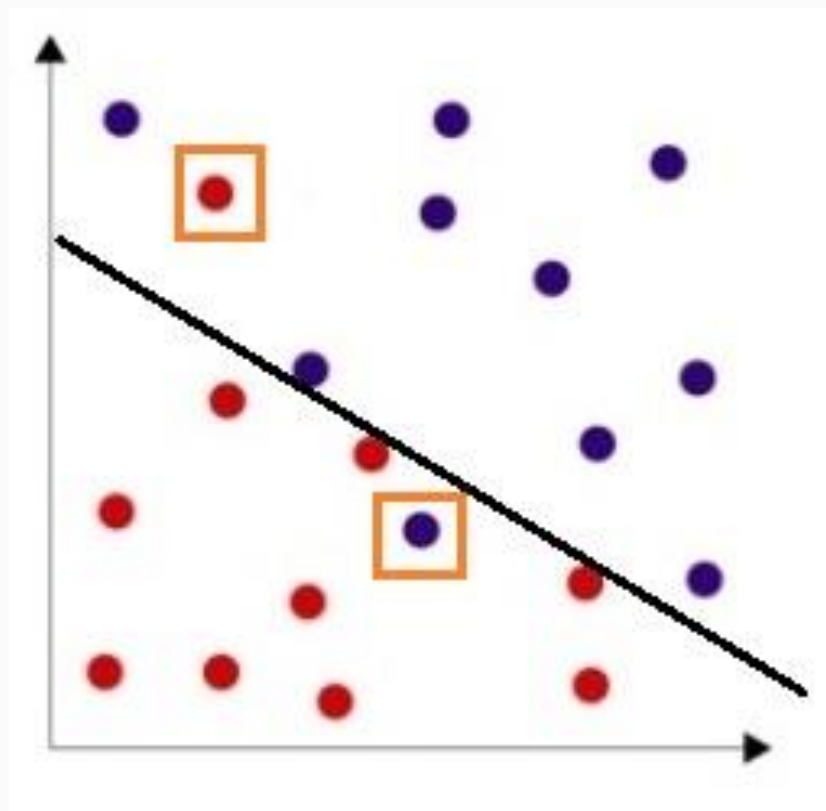
Линейно разделимая выборка



Изображение: [Два множества, разделённых линейно \(ru.wikipedia.org\)](https://ru.wikipedia.org)

# Вторая выборка

Не является линейно разделимой



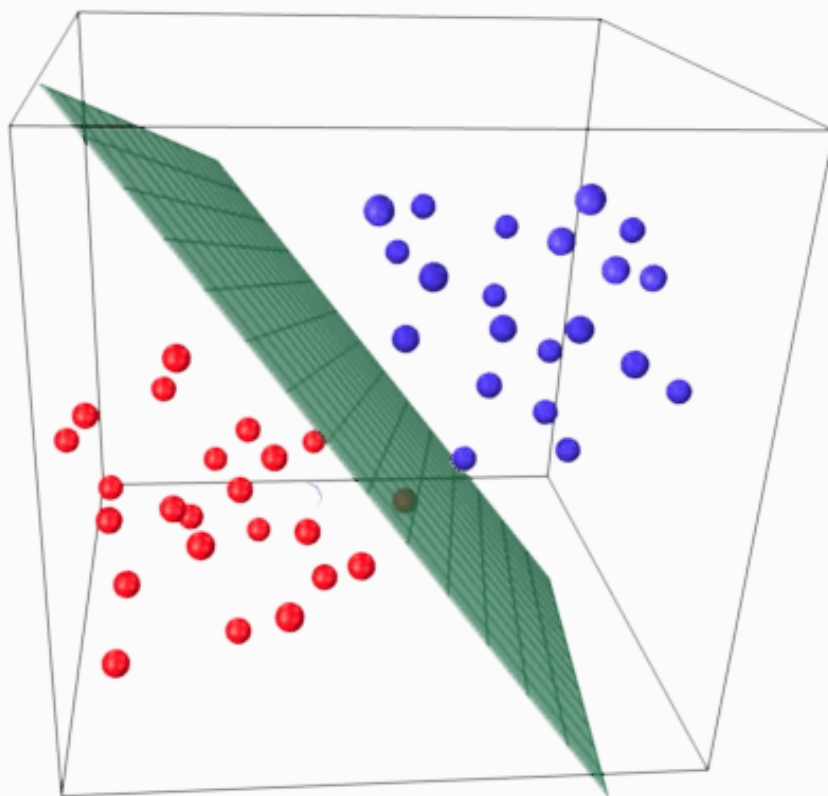
Изображение: [Два множества, разделённых линейно \(ru.wikipedia.org\)](http://ru.wikipedia.org)



Линейный классификатор

# Многомерный случай

Линейно разделимая выборка



Изображение: [Линейно разделимая выборка \(blog.sairahul.com\)](http://blog.sairahul.com)

# Формула линейной регрессии

$$a(x) = w_0 + w_1x_1 + w_2x_2 + \dots + w_dx_d,$$

- $a(x)$  — предсказание модели
- $x_1, x_2, \dots$  — признаки объекта
- $w_0, w_1, w_2, \dots$  — веса модели

# Формула линейной регрессии

$$a(x) = (w, x)$$

# Бинарный классификатор

$$a(x) = \textit{sign}(w, x)$$

Комментарий:

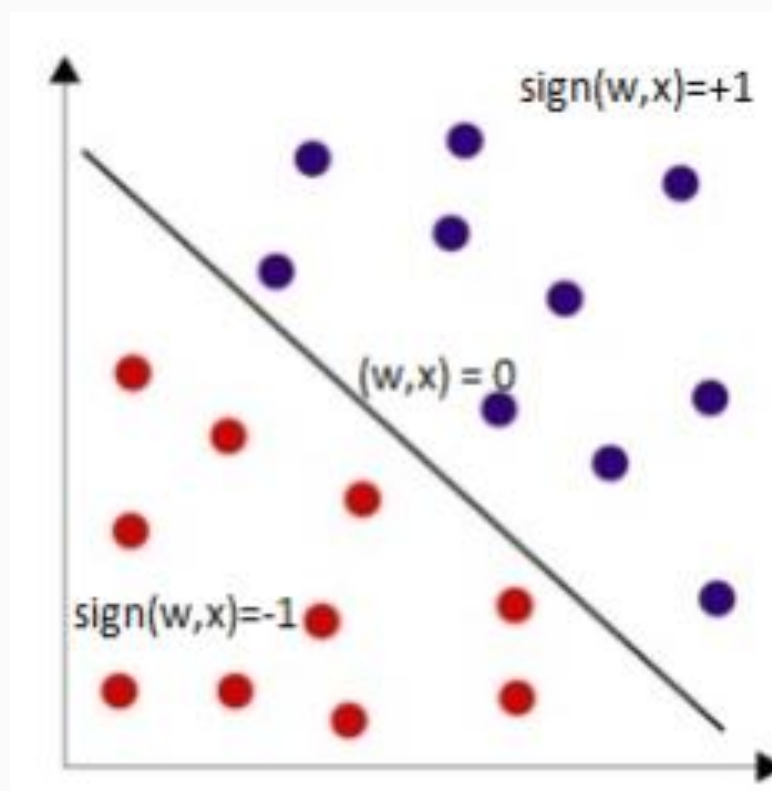
$$\textit{sign}(w, x) = +1, \text{ если } (w, x) > 0$$

и

$$\textit{sign}(w, x) = -1, \text{ если } (w, x) < 0$$

# Линейный классификатор

$$a(x) = \text{sign}(w, x)$$



Изображение: [Два множества, разделённых линейно \(ru.wikipedia.org\)](http://ru.wikipedia.org)

# Бинарный классификатор

$$a(x) = [(w, x) > 0], [.] \text{ — индикатор}$$

- Если  $(w, x) > 0$ , то  $a(x) = 1$
- Если  $(w, x) \leq 0$ , то  $a(x) = 0$

# Пример

Признаки клиента:

- $x_1$  — возраст
- $x_2$  — количество минут разговоров в месяц
- $x_3$  — сумма трат на мобильную связь в месяц



Классификатор:

$$a(x) = \text{sign}(w_0 + w_1x_1 + w_2x_2 + w_3x_3)$$

**4 параметра:**  $w_0, w_1, w_2, w_3$