

# ЗНАКОМСТВО С МАШИННЫМ ОБУЧЕНИЕМ

---

# ПОНРАВИТСЯ ЛИ ПОЛЬЗОВАТЕЛЮ ФИЛЬМ?

---



# ПОНРАВИТСЯ ЛИ ПОЛЬЗОВАТЕЛЮ ФИЛЬМ?

---

- › Просмотреть
- › Дать случайный ответ
- › Спросить у психолога-киномана
- › Воспользоваться машинным обучением!

# ОБЪЕКТЫ И ОТВЕТЫ

---

- »  $x$  — объект
- »  $\mathbb{X}$  — пространство объектов
- »  $y = y(x)$  — ответ
- »  $\mathbb{Y}$  — пространство ответов

# ПРИЗНАКИ

---

- ›  $x = (x^1, x^2, \dots, x^d)$  — признаковое описание
- ›  $x^i$  — число, элемент множества, ...

# ОБУЧАЮЩАЯ ВЫБОРКА

---

- › Машинное обучение — извлечение закономерностей из примеров
- ›  $X = (x_i, y_i)_{i=1}^\ell$

# АЛГОРИТМ

---

- ›  $a(x)$  — алгоритм, модель
- › Функция из  $\mathbb{X}$  в  $\mathbb{Y}$

# АЛГОРИТМ

---

- › Пример: линейный алгоритм
- ›  $a(x) = \text{sign}(w_0 + w_1 x^1 + \dots + w_d x^d)$

# ФУНКЦИОНАЛ ОШИБКИ

---

- ›  $a(x) = 1$  — не подходит
- ›  $Q(a, X)$  — ошибка алгоритма  $a$  на выборке  $X$
- › Пример: доля неправильных ответов

# ОБУЧЕНИЕ

---

›  $Q(a, X) \rightarrow \min$

# ОБУЧЕНИЕ

---

›  $Q(a, X) \rightarrow \min_{a \in ?}$

# ОБУЧЕНИЕ

---

- ›  $Q(a, X) \rightarrow \min_{a \in \mathcal{A}}$
- ›  $\mathcal{A}$  — семейство алгоритмов
- › Решающие пни:  $\mathcal{A} = \{[x^j < t] \mid \forall j, t\}$

# МАШИННОЕ ОБУЧЕНИЕ

---

- › Каким выбрать функционал ошибки?
- › Какое взять семейство алгоритмов?
- › Как обучить алгоритм?
- › Как сформировать признаки?
- › Как подготовить данные?

# **ДАЛЕЕ В ПРОГРАММЕ**

---

- › Постановки задач
- › Виды признаков

# ОБУЧЕНИЕ НА РАЗМЕЧЕННЫХ ДАННЫХ

---

# МАШИННОЕ ОБУЧЕНИЕ

---

- ›  $X = (x_i, y_i)_{i=1}^\ell$  — обучающая выборка  
(размеченные данные)
- ›  $Q(a, X) \rightarrow \min_{a \in \mathcal{A}}$  — обучение с учителем
- › Какие бывают ответы?

# КЛАССИФИКАЦИЯ

---

- »  $\mathbb{Y} = \{0, 1\}$  — бинарная классификация



# КЛАССИФИКАЦИЯ

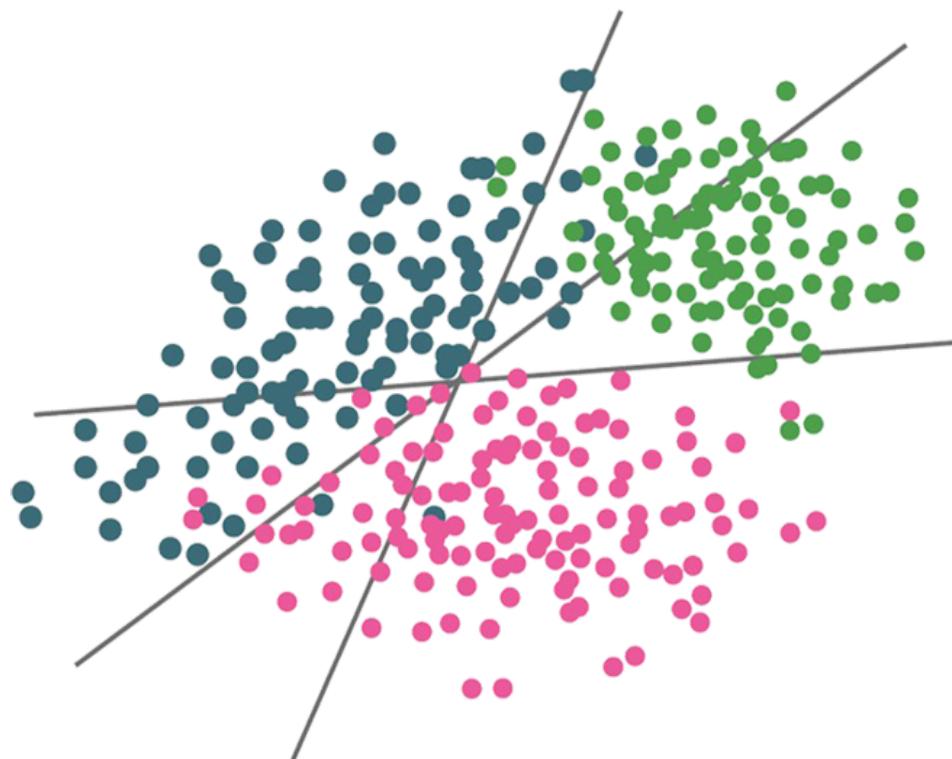
---

- ›  $\mathbb{Y} = \{0, 1\}$  — бинарная классификация
- › Понравится ли пользователю фильм?
- › Вернет ли клиент кредит?
- › Нужно ли делать пациенту операцию?
- › Качественное ли вино?

# КЛАССИФИКАЦИЯ

---

- »  $\mathbb{Y} = \{0, 1, \dots, K\}$  — многоклассовая классификация



# КЛАССИФИКАЦИЯ

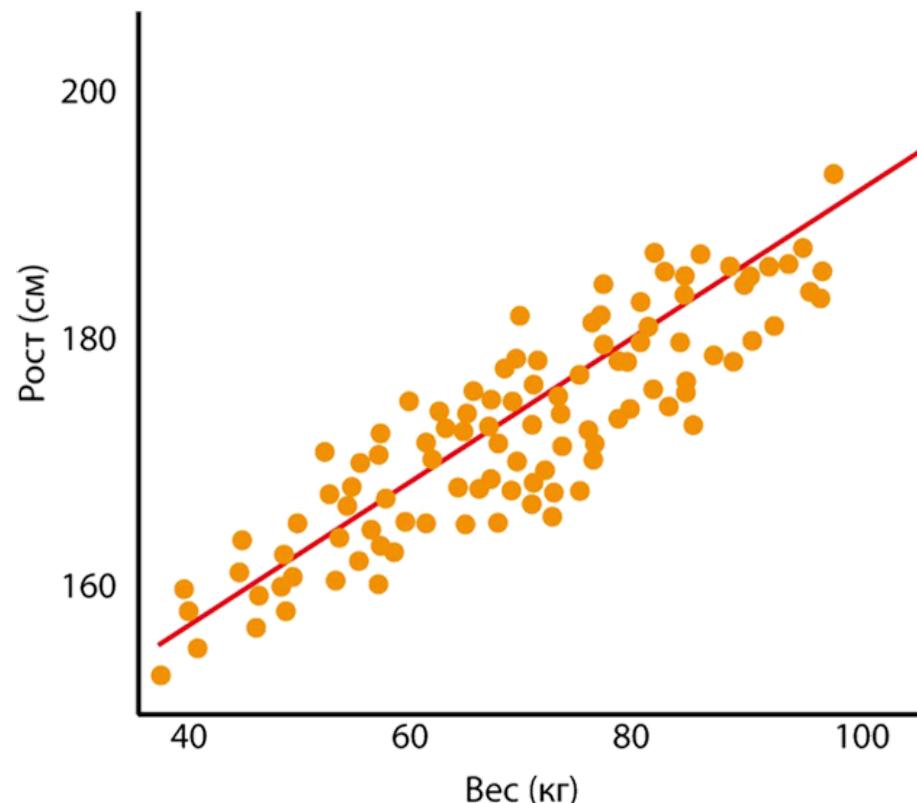
---

- ›  $\mathbb{Y} = \{0, 1, \dots, K\}$  — многоклассовая классификация
- › Из какого сорта винограда сделано вино?
- › На какую тему статья?
- › Какой тип имеет машина на фото?

# РЕГРЕССИЯ

---

- $\mathbb{Y} = \mathbb{R}$
- Предсказание роста по весу



# РЕГРЕССИЯ

---

- ›  $\mathbb{Y} = \mathbb{R}$
- › Предсказание роста по весу
- › Какая завтра будет температура?
- › Сколько денег заработает магазин за год?
- › Сколько лет человеку на фото?

# РАНЖИРОВАНИЕ

Яндекс

картинки с котиками — 5 млн ответов



Найти

Поиск

[Картинки с кошками | Fun Cats — Забавные коты](#)

[funcats.by > pictures/](#) ▾

Картинки с кошками. Прикольные коты. 777 изображений. ... 32 изображения. Кошки Стамбула. 41 изображение. Веселые котята.

Видео

[Уморные котики \(57 фото\) » Бяки.нет | Картинки](#)

[byaki.net > Картинки > 14026-umornye-kotiki-57...](#) ▾

Бяки нет! . NET. Уморные котики (57 фото). 223. Коментариев:9Автор:4ertonok  
Просмотров:161 395 Картинки28-10-2008, 00:03.

Карты

Маркет

Ещё

[Смешные картинки кошек с надписями | Полкот.Ру](#)

[lolkot.ru](#) ▾

Смешные картинки для новых приключений! Сделать свой прикол очень просто. ... Котик верит в чудеса. Он в носке подарок ищет...

[Красивые картинки и фото кошек, котят и котов](#)

[foto-zverey.ru > Кошки](#) ▾

Фото и картинки кошек и котят потрясающей красоты и нежности. Здесь мы собрали такие изображения, которые всегда вызывают море положительных эмоций...

[Обои для рабочего стола Котята | картинки на стол Котята](#)

[7fon.ru > Чёрные обои и картинки > Обои котята](#) ▾

Картинки Котята с 1 по 15. Обои для рабочего стола Котята. ... Скачать Картинки Котята на рабочий стол бесплатно.

# РЕЗЮМЕ

---

- › Классификация: бинарная, многоклассовая
- › Регрессия
- › Ранжирование

# РЕЗЮМЕ

---

- › Классификация: бинарная, многоклассовая
- › Регрессия
- › Ранжирование
- › Далее: обучение без учителя

# ОБУЧЕНИЕ БЕЗ УЧИТЕЛЯ

---

# ВИДЫ ОБУЧЕНИЯ

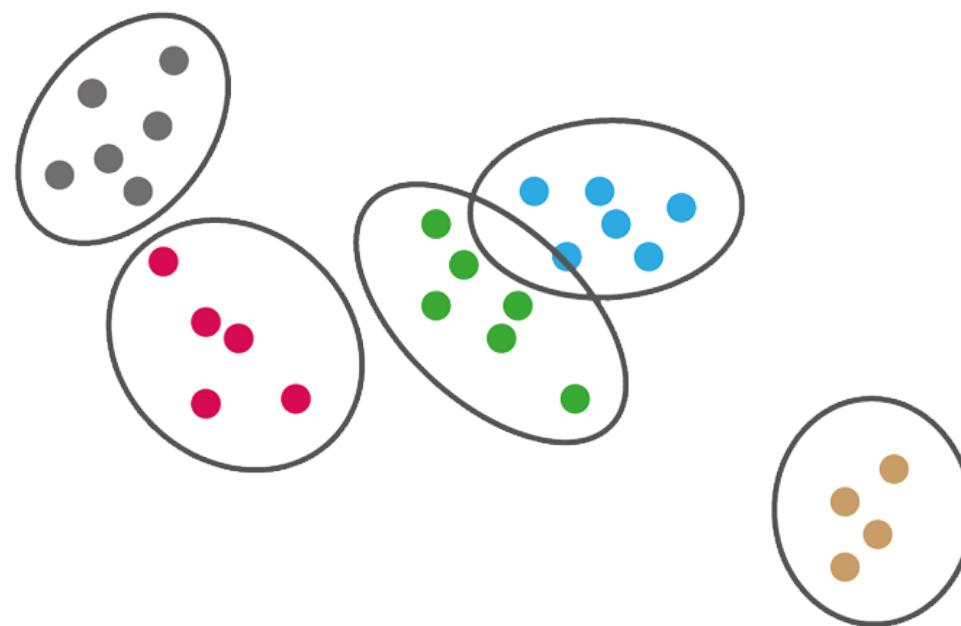
---

- › Обучение с учителем — объекты и ответы
- › Обучение без учителя — только объекты
- › Частичное обучение — объекты, некоторые с ответами
- › Активное обучение — на каком объекте получить ответ?

# КЛАСТЕРИЗАЦИЯ

---

- › Найти группы похожих объектов
- › Не знаем количество кластеров
- › Не знаем правильных ответов



# КЛАСТЕРИЗАЦИЯ

---

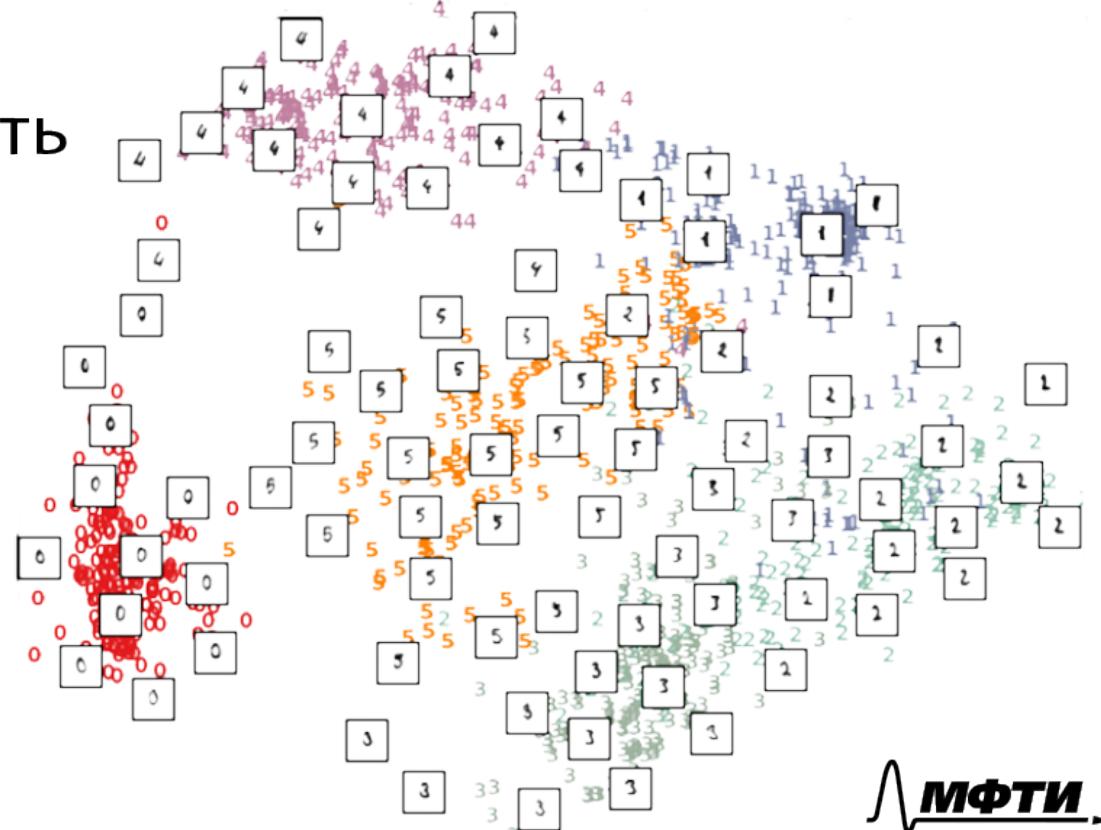
- › Сегментация пользователей
- › Поиск схожих пользователей в социальных сетях
- › Поиск генов с похожими профилями экспрессии

# ВИЗУАЛИЗАЦИЯ

› Задача: нарисовать многомерную выборку

› Должна отражать структуру объектов

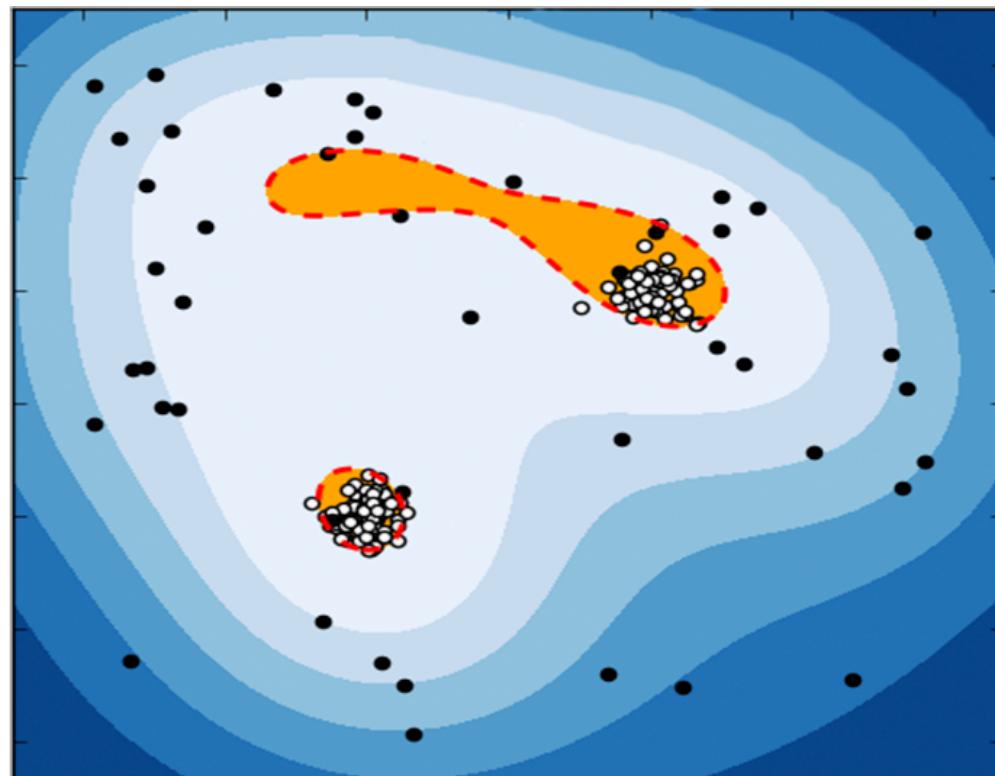
› Должно быть  
красиво



# ПОИСК АНОМАЛИЙ

---

- › Определить объект, который не похож на большинство
- › Примеров аномалий — мало или вообще нет



# ПОИСК АНОМАЛИЙ

---

- › Определение поломки в системах самолета
- › Определение поломки интернет-сайта
- › Выявление проблем в модели машинного обучения

# РЕЗЮМЕ

---

- › Кластеризация
- › Визуализация
- › Поиск аномалий
- › Курс «Поиск структуры в данных»

# ПРИЗНАКИ В МАШИННОМ ОБУЧЕНИИ

---

# ПРИЗНАКИ В МАШИННОМ ОБУЧЕНИИ

---

# ПРИЗНАКИ

---

- › Описывают объекты
- ›  $D_j$  — множество значений  $j$ -го признака

# БИНАРНЫЕ ПРИЗНАКИ

---

- ›  $D_j = \{0, 1\}$
- › Доход клиента выше среднего по городу?
- › Цвет фрукта — зеленый?

# ВЕЩЕСТВЕННЫЕ ПРИЗНАКИ

---

- ›  $D_j = \mathbb{R}$
- › Возраст
- › Площадь квартиры
- › Количество звонков в колл-центр

# КАТЕГОРИАЛЬНЫЕ ПРИЗНАКИ

---

- ›  $D_j$  — неупорядоченное множество
- › Цвет глаз
- › Город
- › Образование (может быть упорядоченным)

# ПОРЯДКОВЫЕ ПРИЗНАКИ

---

- ›  $D_j$  — упорядоченное множество
- › Роль в фильме (первого плана, второго плана, массовка)
- › Тип населенного пункта
- › Образование (может быть неупорядоченным)

# МНОЖЕСТВОЗНАЧНЫЕ ПРИЗНАКИ

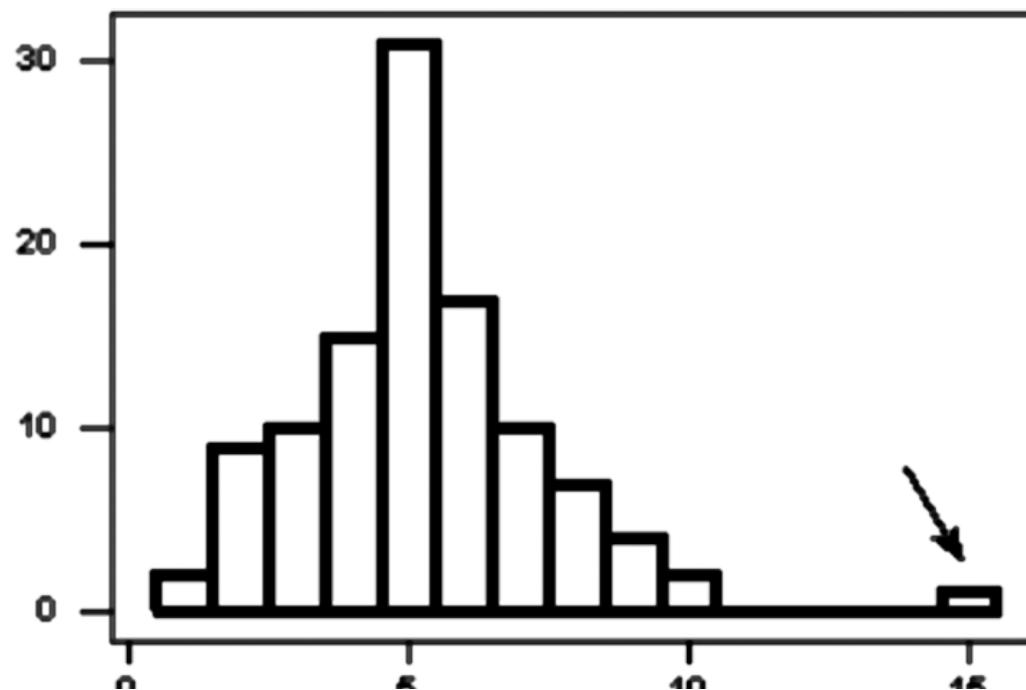
---

- › (set-valued)
- ›  $D_j$  — множество всех подмножеств некоторого множества
- › Какие фильмы посмотрел пользователь?
- › Какие слова входят в текст?

# ВЫБРОСЫ

---

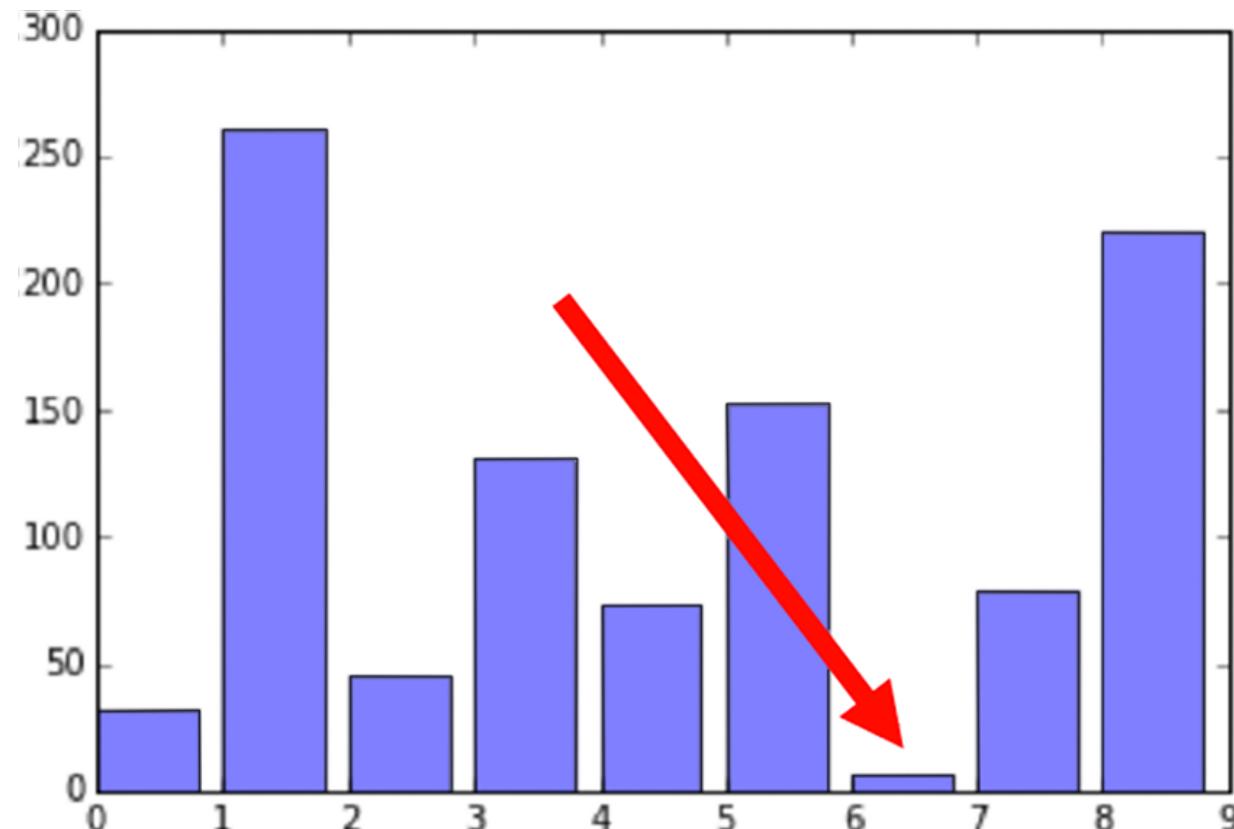
- › Значения признаков, не похожие на большинство
- › Могут сильно мешать поиску закономерности



# РАСПРЕДЕЛЕНИЕ ПРИЗНАКА

---

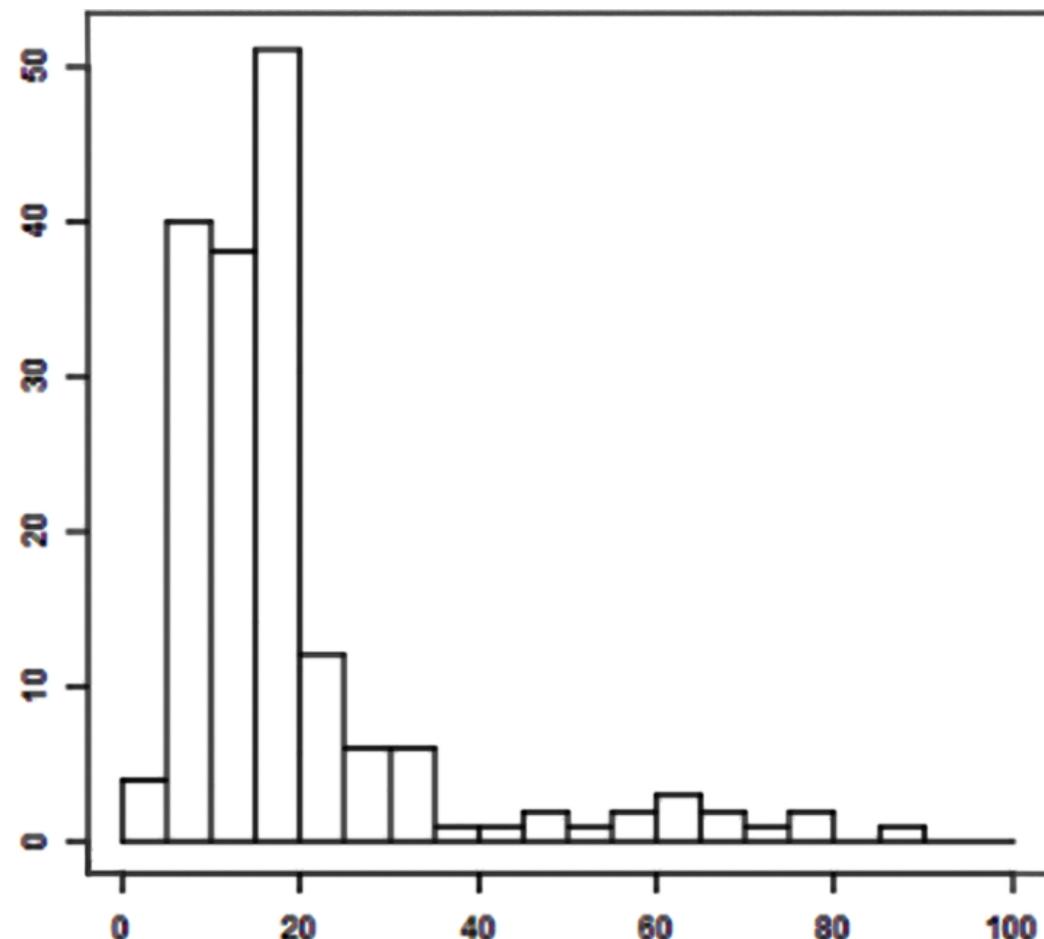
- › Признаки могут иметь распределение, которое мешает работе с ними



# РАСПРЕДЕЛЕНИЕ ПРИЗНАКА

---

- › Признаки могут иметь распределение, которое мешает работе с ними



# РЕЗЮМЕ

---

- › Бинарные признаки
- › Вещественные признаки
- › Категориальные и порядковые признаки
- › Set-valued признаки
- › Выбросы и «плохие» распределения