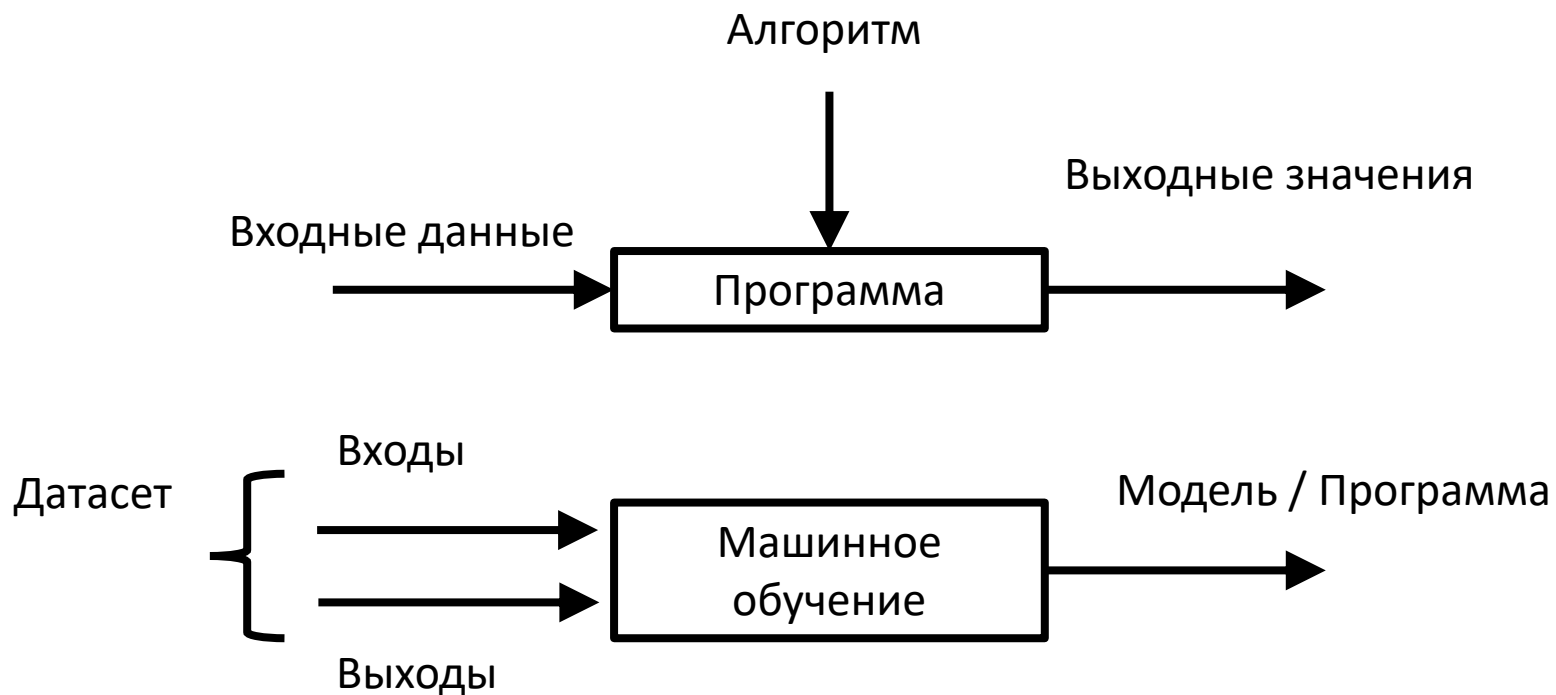


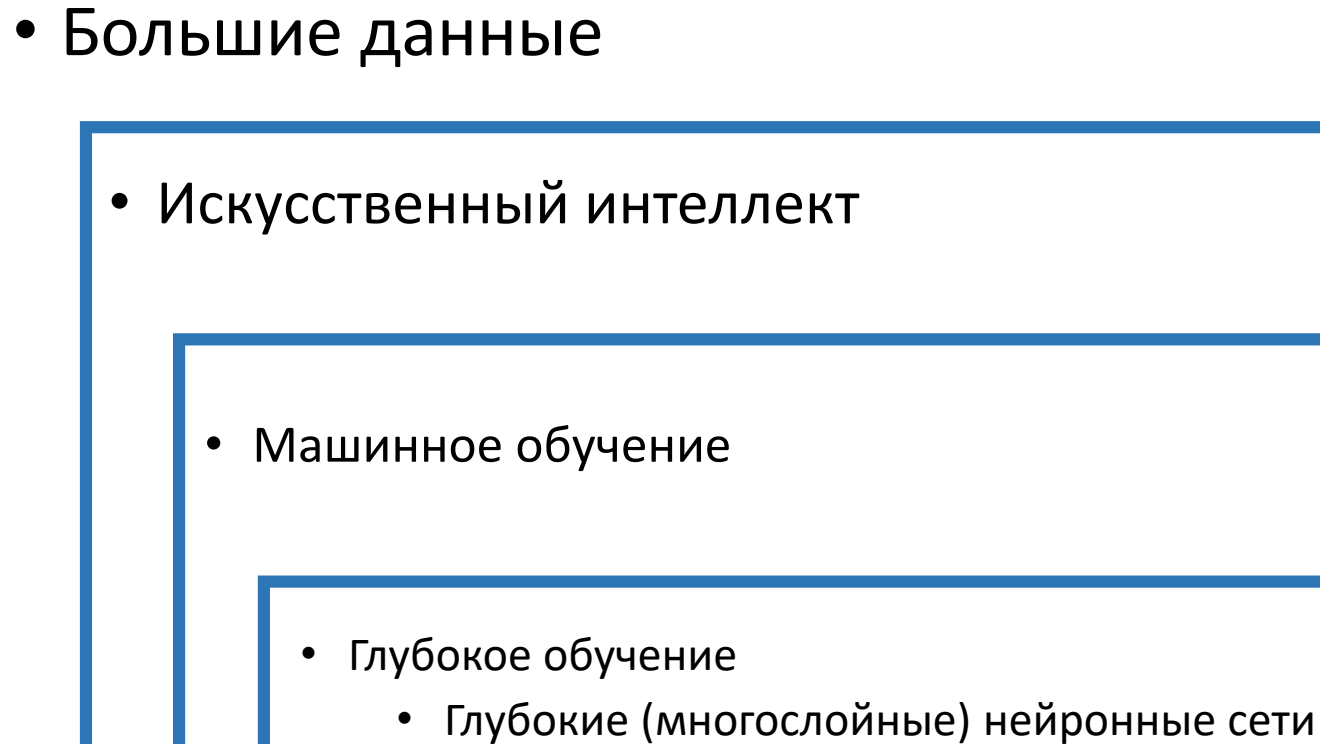
# Машинное обучение

## Machine Learning

# Программирование / МО



# Место машинного обучения

- 
- Большие данные
    - Искусственный интеллект
      - Машинное обучение
        - Глубокое обучение
          - Глубокие (многослойные) нейронные сети

# Программа

- Программирование
  - Алгоритм – программист/разработчик
- Машинное обучение
  - Программист ставит задачу
  - Машина сама составляет алгоритм
    - «Обучается» по набору исходных данных «датасету»
    - Подбирает параметры алгоритма
  - Развертывание модели: «деплой в продакшн»
    - Прогнозирование значений

# Цена квартиры

- Яндекс Недвижимость / ДомКлик ...
- Dataset – датасет - выборка
- X - Статистические признаки – Feature
  - Площадь
  - Число комнат
  - Район (Калининский, Орджоникидзевский...)
  - Год постройки
  - Вторичная/новостройка
  - Лифт (есть/нет)
- Y – Цена – Target value – целевое значение
- Машинная (компьютерная) модель
  - Прогноз цены квартиры – уравнение:
  - $Y = f(x_1, x_2, \dots)$

# Dataset – Датасет – Выборка



realty



Файл Правка Вид Вставка Формат

	A	B	C
1	<b>Площадь</b>	<b>Комнат</b>	<b>Цена</b>
2	39	1	5100000
3	18,5	1	1570000
4	35,1	1	2745000
5	47,5	1	5600000
6	37,2	1	3170000

# ETL / Data Engineering

- **E – Extract**

- Извлечение данных из разных источников

- **T – Transform** – Преобразование данных

- Data Cleaning / Cleansing - Очистка данных

- **L – Load** – Загрузка данных, подготовленных для анализа, хранилище данных

- Формат ячеек – тип данных

4.258.240 руб.
4258240
2 ноября 2022
02.11.2022
11.02.2022

# Задачи машинного обучения

- Постановка задачи: Дано... Найти...
- Обучение с учителем ( $Y$  известны)
  - Supervised Learning
  - Регрессия
  - Классификация
- Обучение без учителя ( $Y$  неизвестны)
  - Unsupervised Learning
  - Кластеризация
  - Понижение размерности
- Обучение с подкреплением (итоги игры)
  - Reinforcement Learning

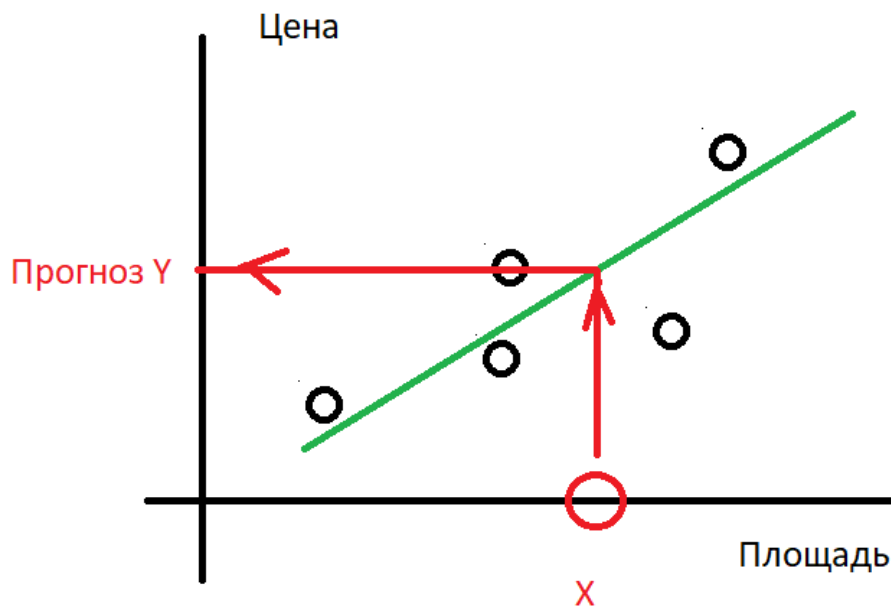


# Регрессия

- Прогноз значения для непрерывной величины

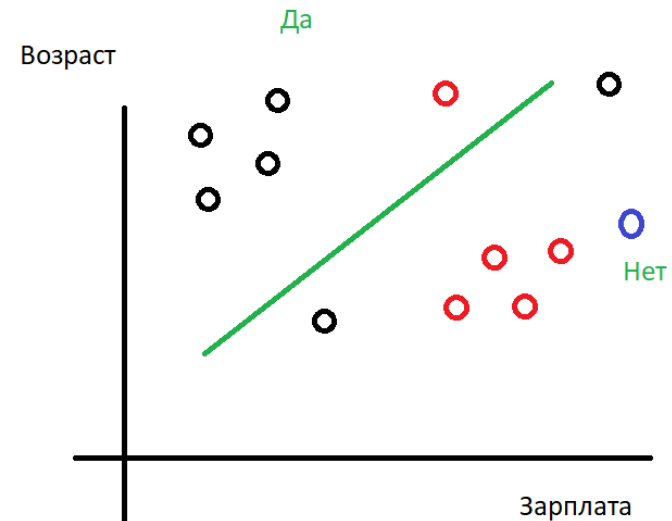
$$y = a + b * X \pm \Delta y$$

- Difference – Разность/разница –  $dy \Delta y$



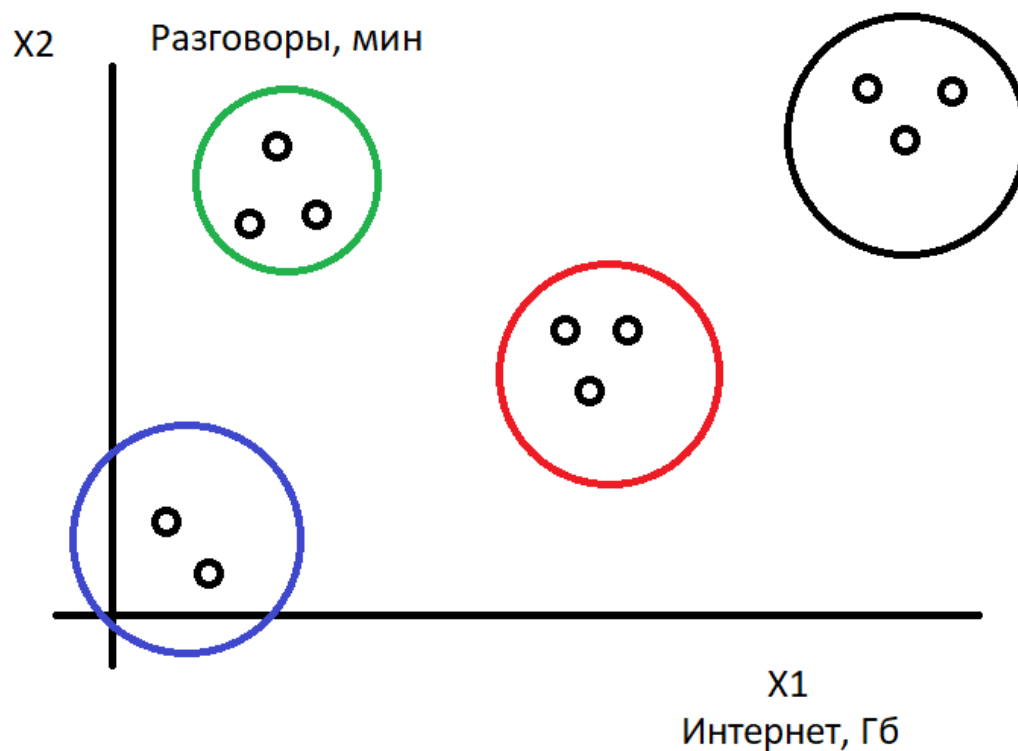
# Классификация

- Прогноз дискретной величины: да/нет: 0 или 1
  - Вероятность от 0 до 1
- Кредит одобрить / не одобрить
- Кредит погасил или не вернул в срок



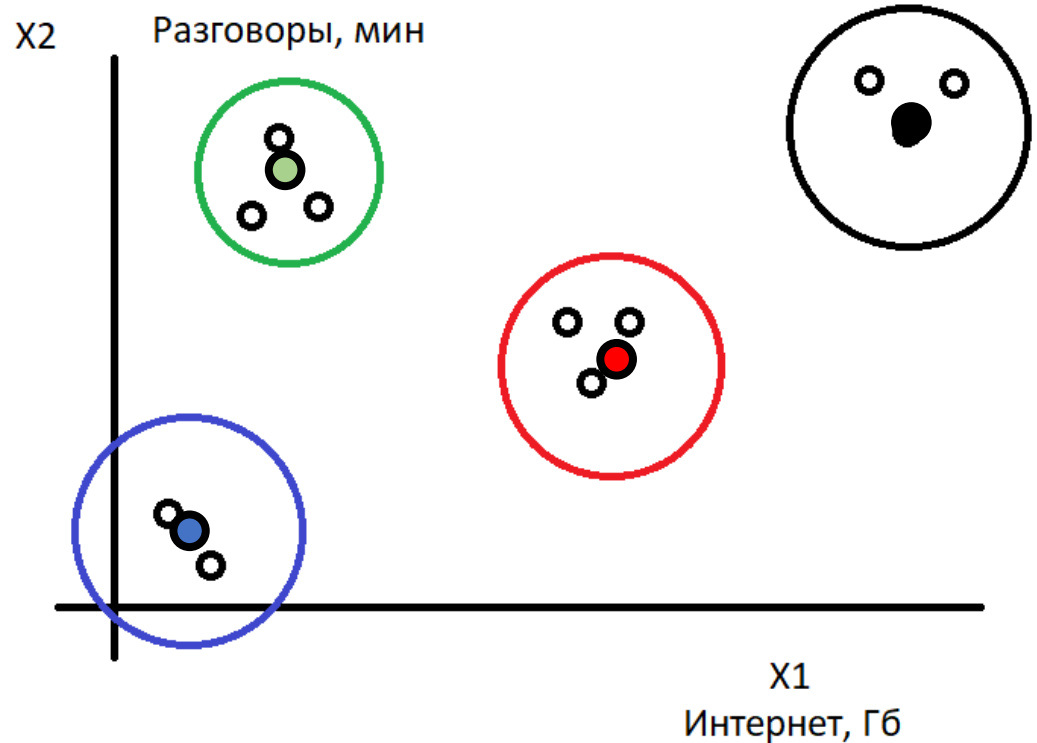
# Кластеризация

- Группы похожих клиентов



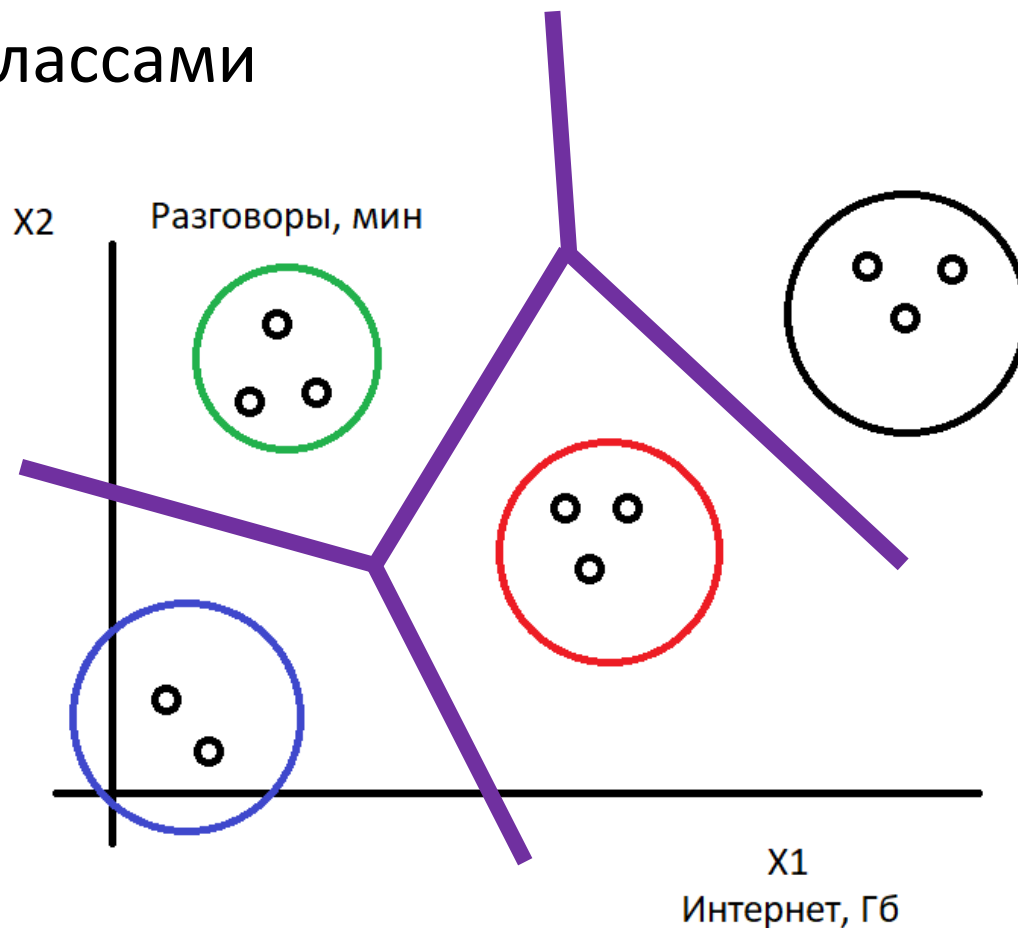
# Кластеризация

- Группы похожих клиентов
  - Центроиды



# Кластеризация

- Границы между классами



# Понижение размерности

- Dimensionality reduction:  $f(x_1, x_2) \rightarrow f(z_1)$

