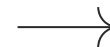
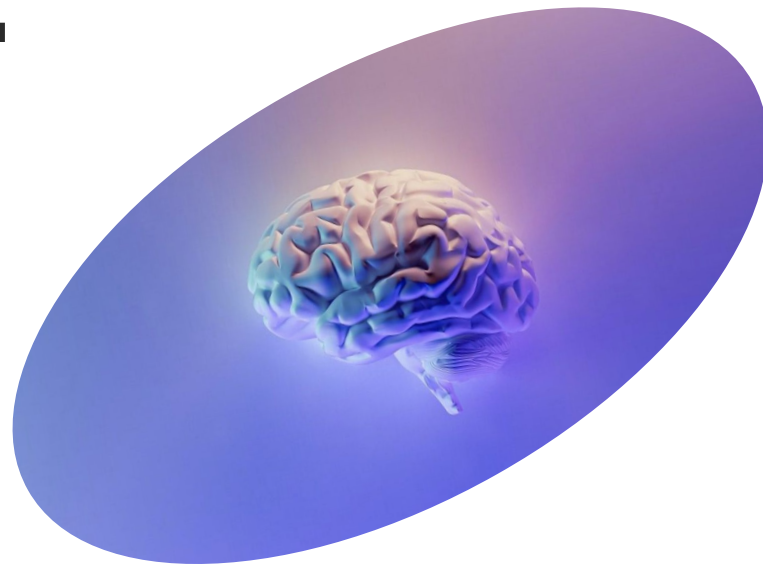


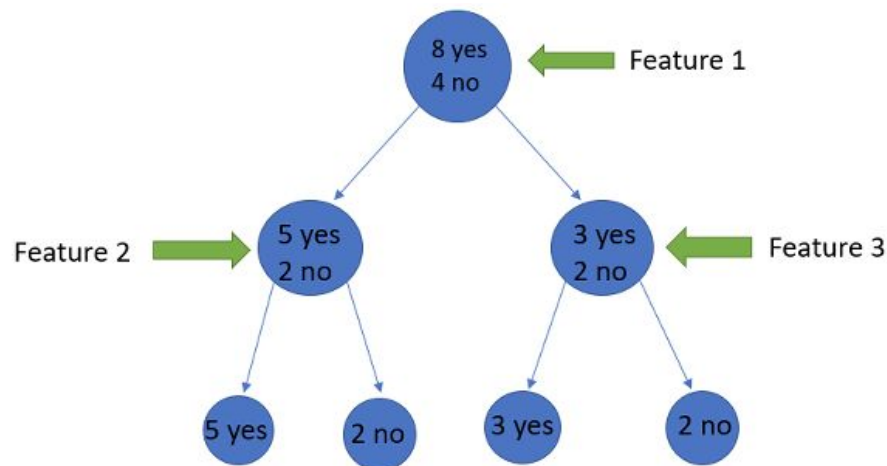
ВВЕДЕНИЕ В ML



Как представляют

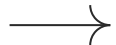
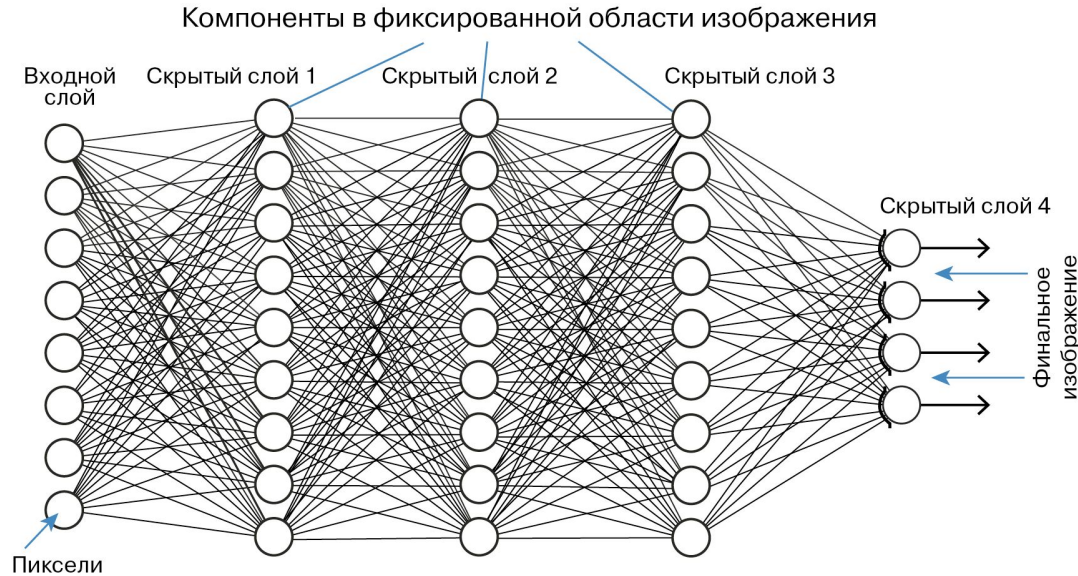


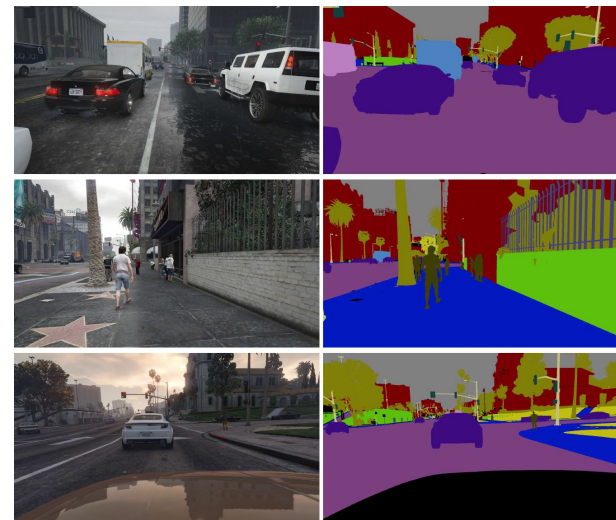
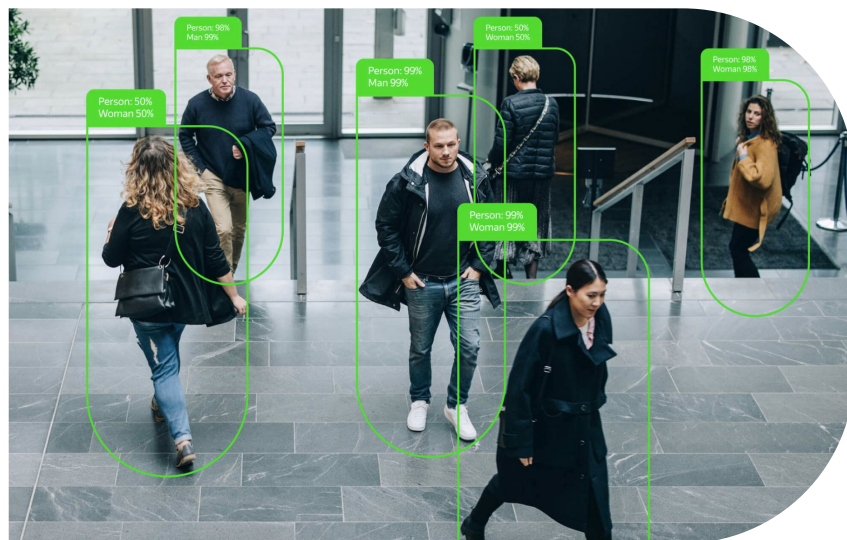
На самом деле



ЧТО ТАКОЕ ИИ НА САМОМ ДЕЛЕ

Мат. модель с определенной архитектурой





Варианты стилизации:

Без стиля

Теории заговора

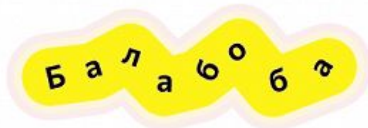
ТВ-репортажи

Тосты

Пацанские цитаты

P

18+



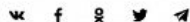
iXbt.com лучше всех пишет о технике

Перевод строки: Shift+Enter
Отправить: Enter

Набалабобить

Нейросеть подражает текстам в интернете. Не относитесь к написанному серьёзно

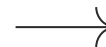
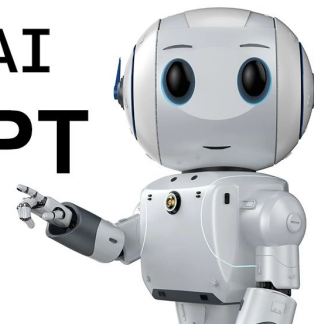
iXbt.com лучше всех пишет о технике, так что не буду повторяться и поделюсь обзором на новый смартфон Huawei — P20 Lite. В данном обзоре не будет каких-либо откровений и сюрпризов — всё те же плюсы и минусы, что и у старшей версии.



Скопировать ссылку

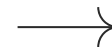
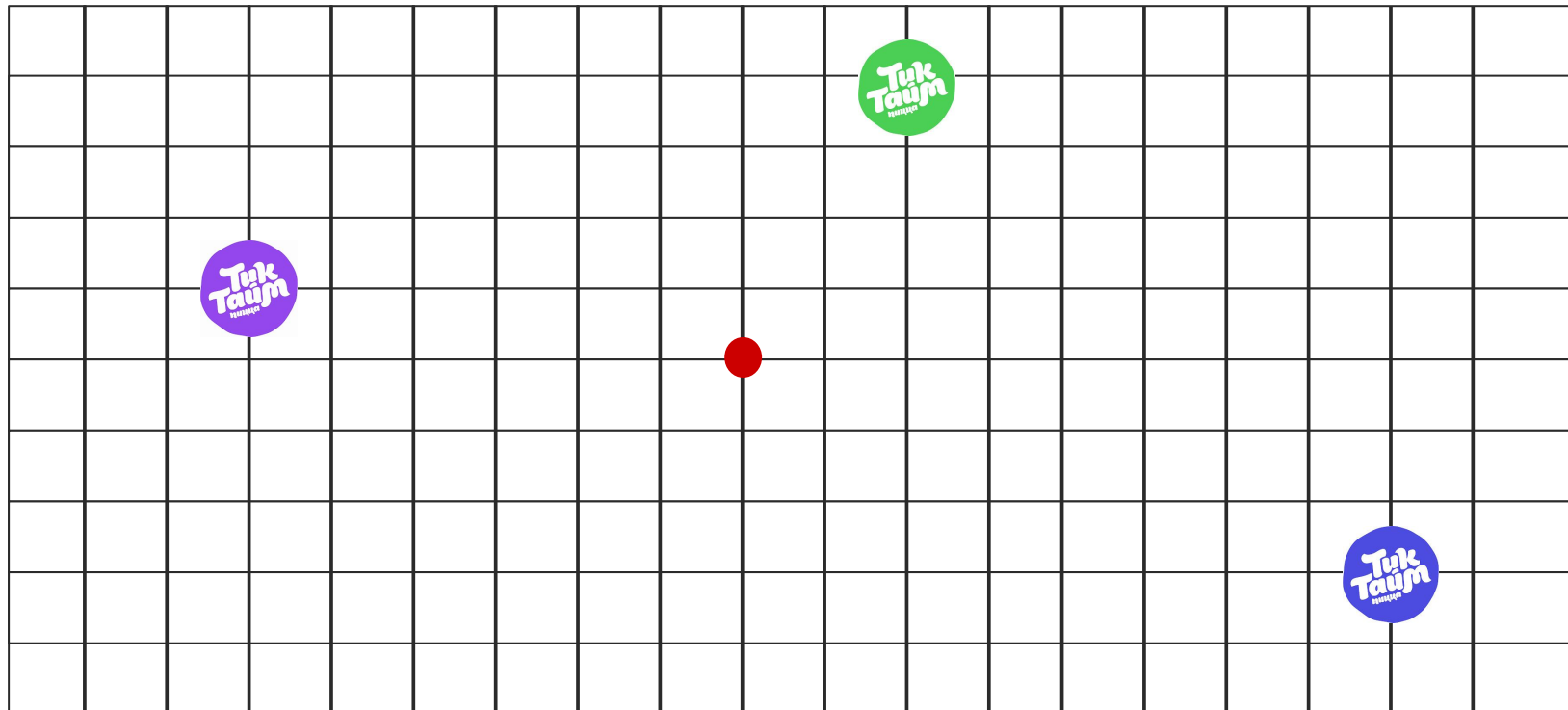
Это оскорбительно

Балабоба демонстрирует, как с помощью нейросетей семейства YaLM можно

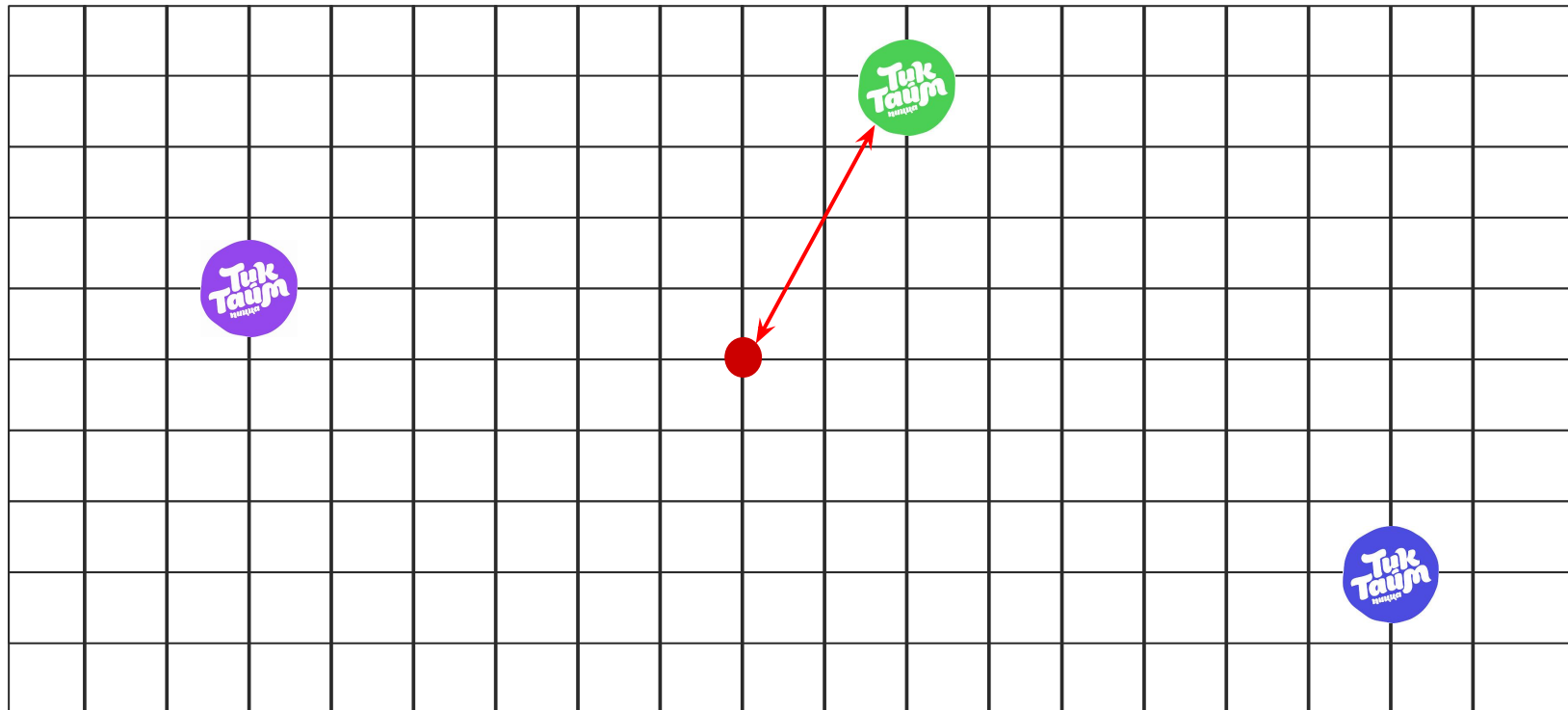
 OpenAI
ChatGPT



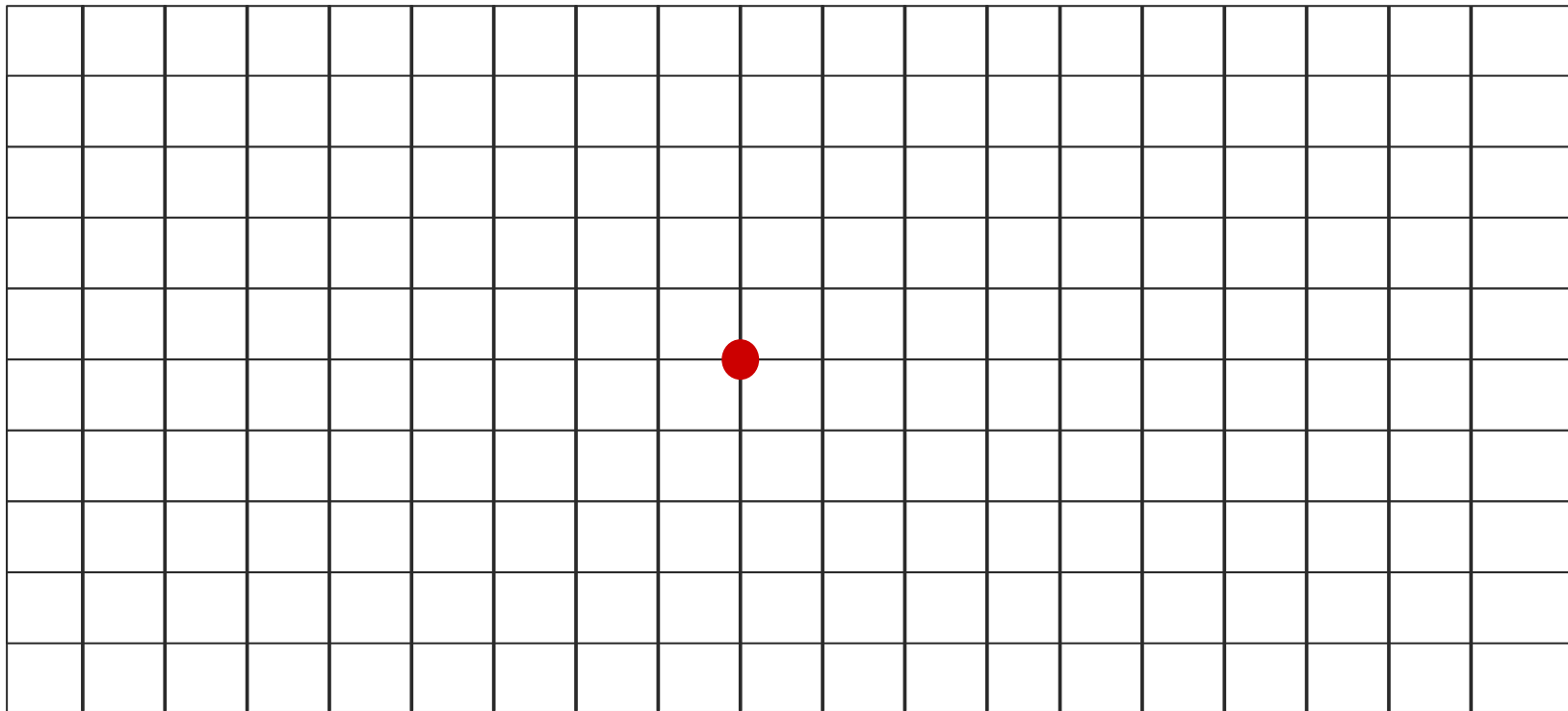
ПРИМЕР ЗАДАЧИ



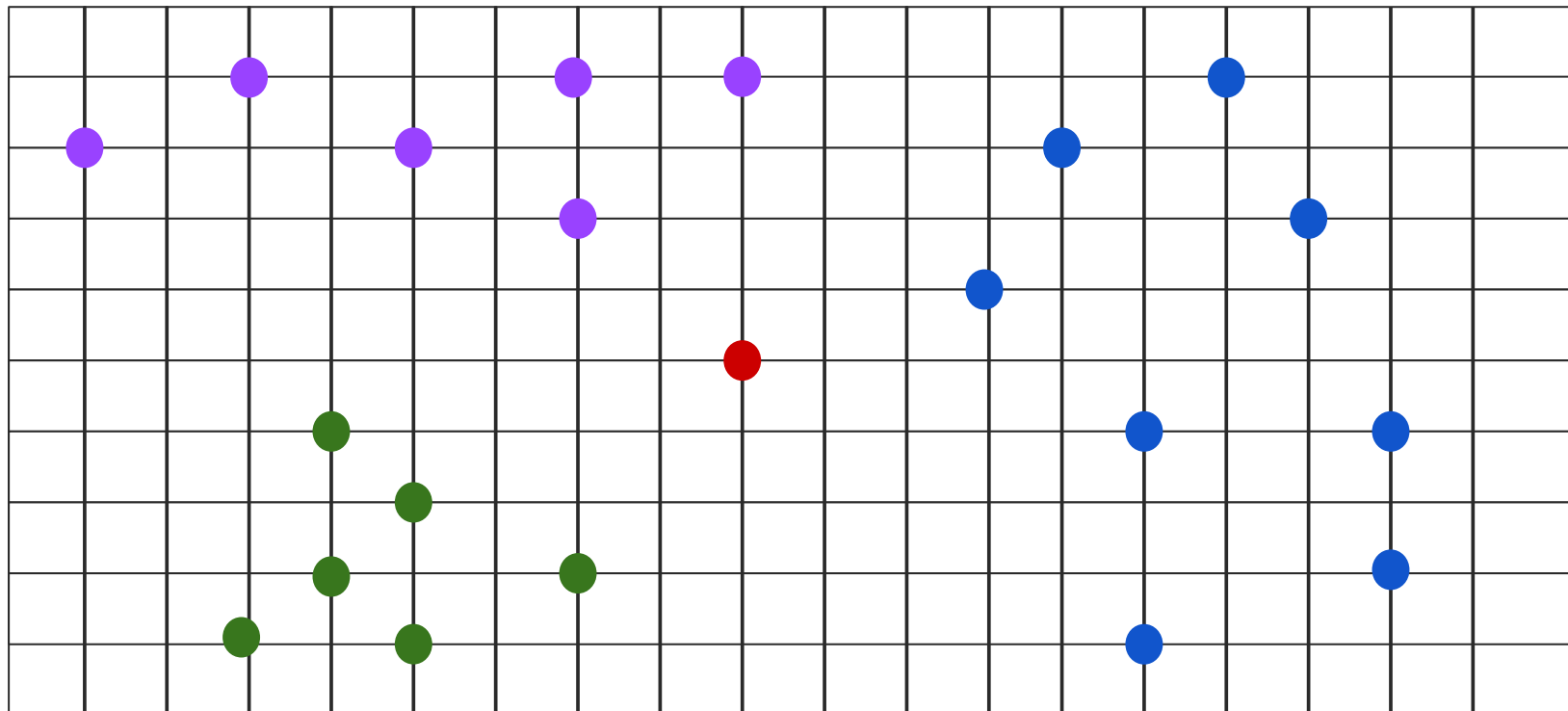
ПРИМЕР ЗАДАЧИ



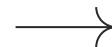
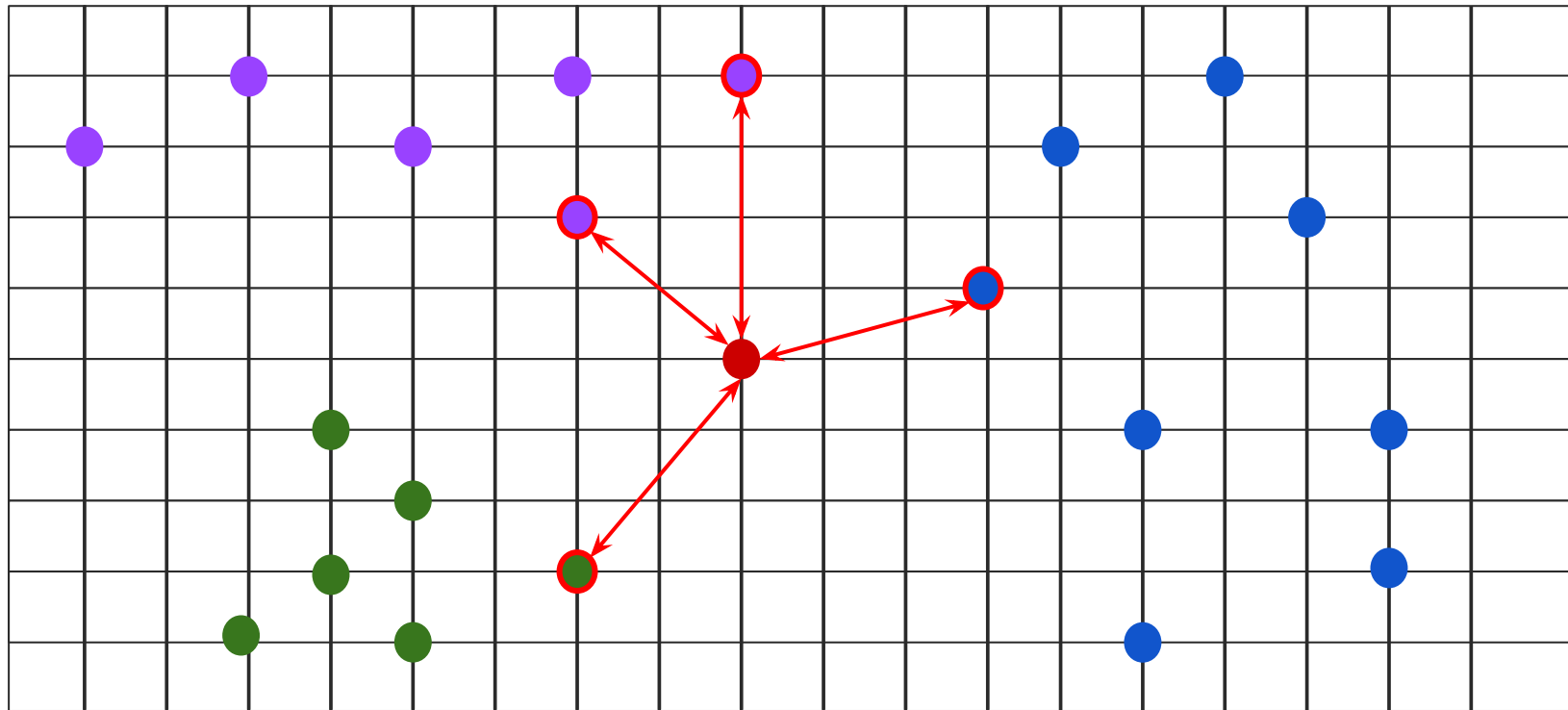
УСЛОЖНЕНИЕ ЗАДАЧИ



УСЛОЖНЕНИЕ ЗАДАЧИ



УСЛОЖНЕНИЕ ЗАДАЧИ



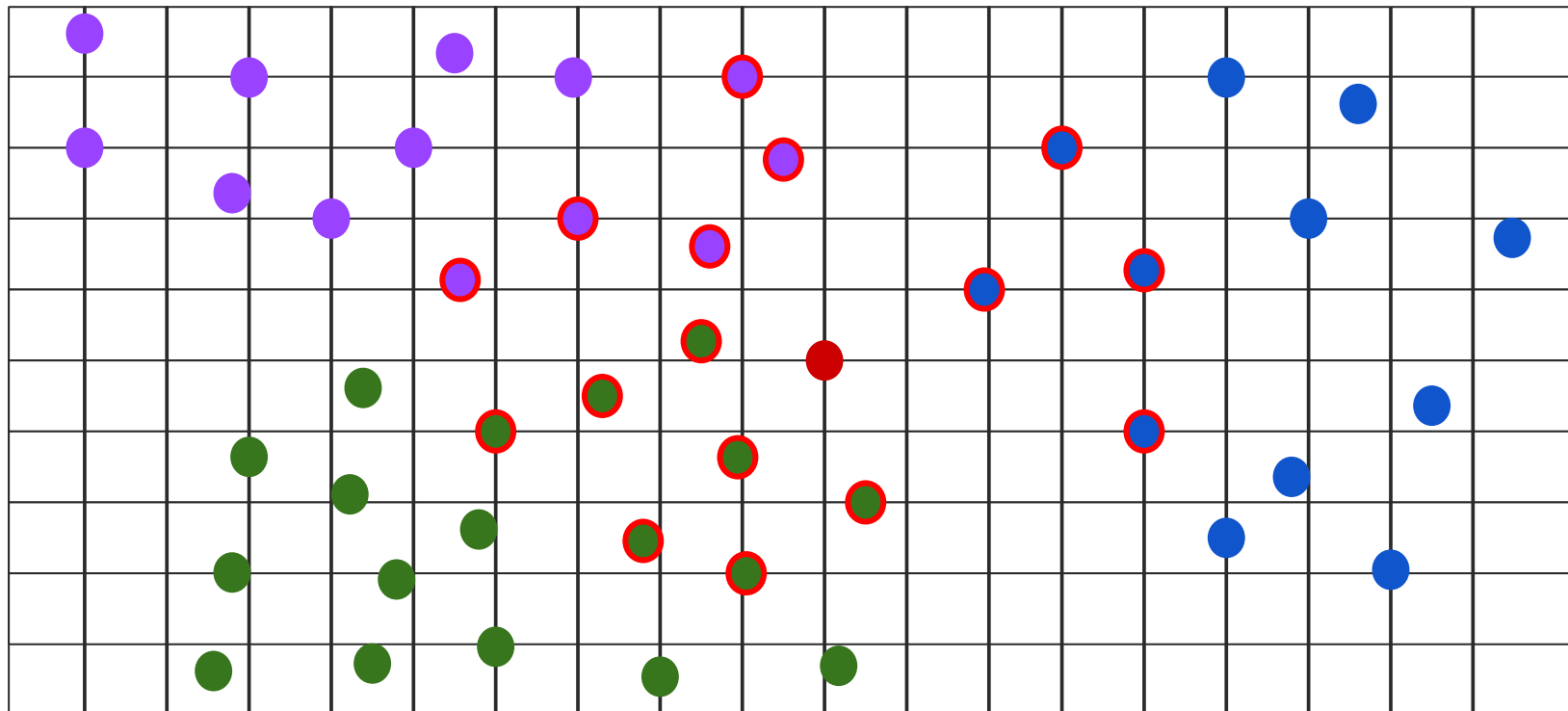
KNN (K-Nearest Neighbors)

1. Получаем точку **x**, в которой нужно сделать предсказание
2. Находим для нее **k** ближайших соседей
3. **Ответ:** наиболее часто встречающийся класс в соседях

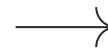
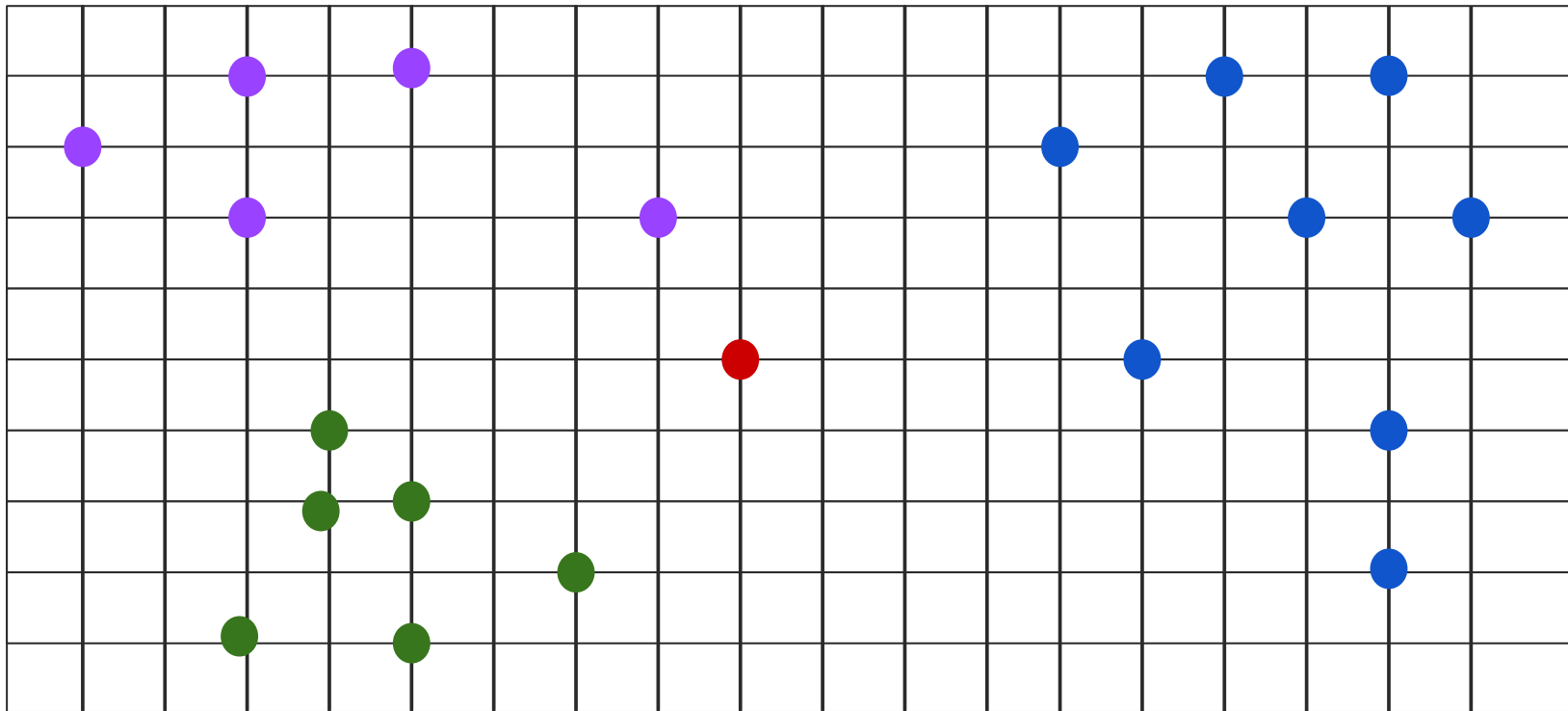
Класс - метка, группа (целочисленный ответ).

Классификация - предсказание класса для входных данных

КОЛИЧЕСТВО ДАННЫХ



ВЫБРОСЫ



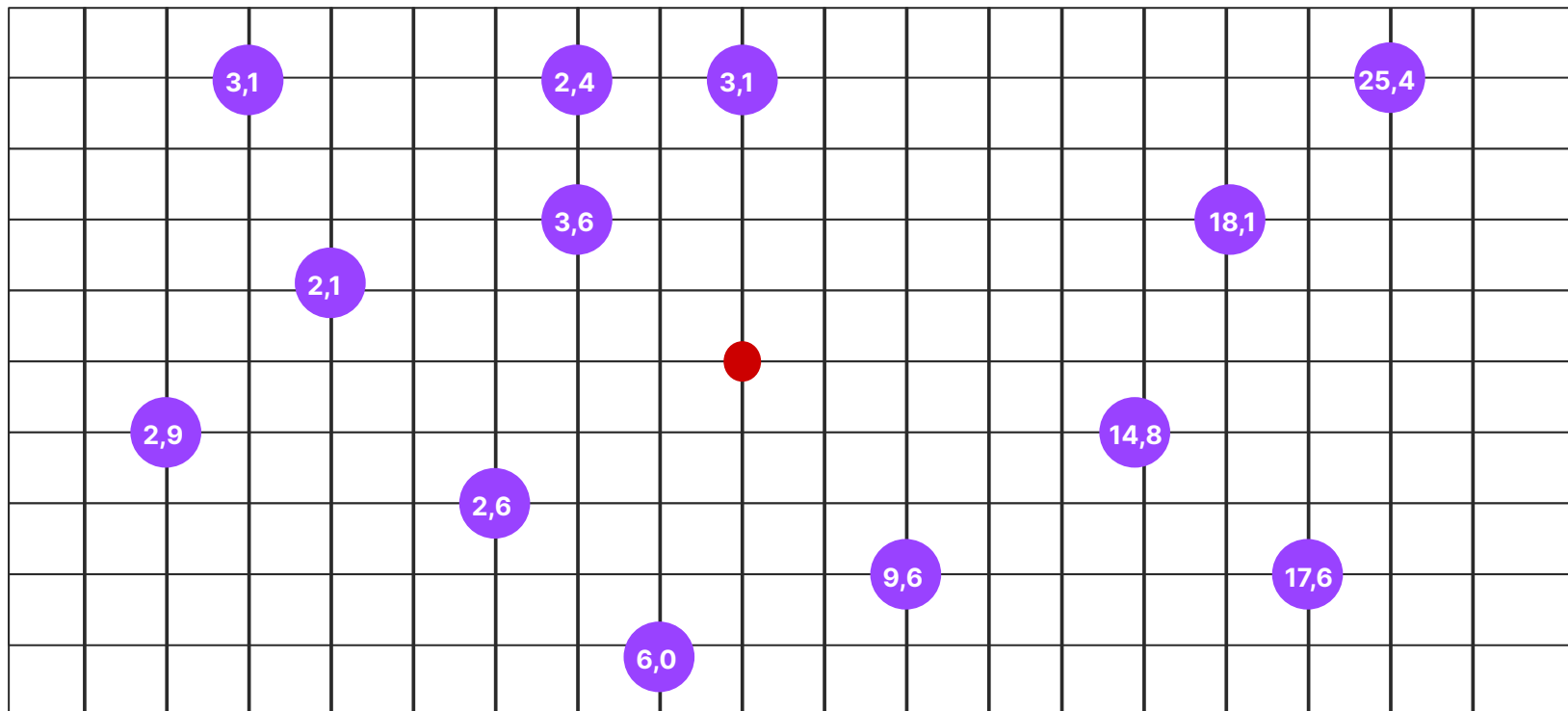
KNN (K-Nearest Neighbors)

Регрессия - задача предсказания вещественного числа по входным данным

Пример: Надо узнать расстояние до пиццерии из которой привезут пиццу.
Знаем расстояние до пиццерии из разных точек.

Решение: среднее значение из ближайших соседей

ПРИМЕР РЕГРЕССИИ



ДАННЫЕ

Датесет – набор данных

Датасет состоит из объектов с одинаковыми признаками.

Пример: задача - оценить стоимость квартиры.

Данные: разные квартиры.

Признаки: количество комнат, этаж, номер дома, район, расстояние до остановки, качество ремонта, и т.д.

ДАННЫЕ

Размерность данных — количество признаков.

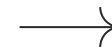
Размерность 3

- 1) Кол-во комнат
- 2) Площадь квартиры
- 3) Этаж

Размерность 5

- 1) Кол-во комнат
- 2) Площадь квартиры
- 3) Этаж
- 4) Расстояние до метро
- 5) Качество ремонта

Чем больше информации об объекте мы знаем, тем точнее наша модель



ТИПЫ ПРИЗНАКОВ

1

Числовые

1, 2, 3, 4, ...

2

Порядковые

отлично, хорошо, удовлетворительно

3

Категориальные

красный, синий, желтый

ДАННЫЕ

Представим данные в виде вектора чисел.

Пример: 1-комнатная квартира на 21 этаже в 13 доме.
(1, 21, 13)

3-комнатная квартира на 4 этаже в 9 доме.
(3, 4, 9)

4-комнатная квартира на 9 этаже в 9 доме.
(4, 9, 19)

ДАННЫЕ

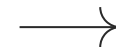
X - множество объектов (признаков/фичей/features).

Y - множество ответов (меток).

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

n – количество признаков

m – количество примеров



ДАННЫЕ

$$\begin{array}{l}
 \text{квартира 1} \\
 \text{квартира 2} \\
 X = \\
 \text{квартира } m
 \end{array}
 \begin{array}{c}
 \text{КОЛ-ВО} \\
 \text{КОМНАТ}
 \end{array}
 \begin{array}{c}
 \text{этаж}
 \end{array}
 \begin{array}{c}
 \text{дом}
 \end{array}
 \begin{array}{c}
 \text{СТОИМОСТЬ} \\
 \text{квартиры}
 \end{array}
 \begin{array}{c}
 Y =
 \end{array}
 \begin{array}{c}
 \begin{pmatrix}
 1 & 21 & \dots & 13 \\
 3 & 4 & \dots & 9 \\
 \vdots & \vdots & \ddots & \vdots \\
 4 & 9 & \dots & 9
 \end{pmatrix}
 \begin{pmatrix}
 6000000 \\
 12000000 \\
 \vdots \\
 7000000
 \end{pmatrix}
 \end{array}$$

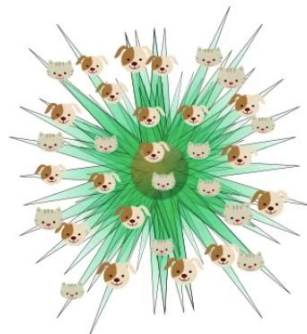
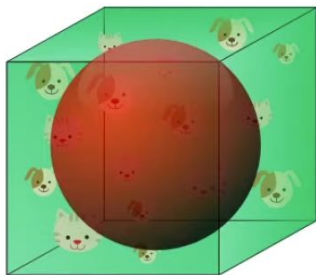
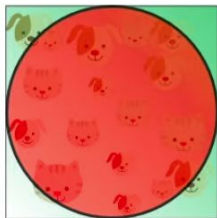
ПРОКЛЯТИЕ РАЗМЕРНОСТИ

Предположение: близкие точки будут иметь близкие ответы.

При большой размерности в близкую область попадает мало ответов

→ нужно больше данных

→ больше ресурсов на обработку



FEATURE SCALE

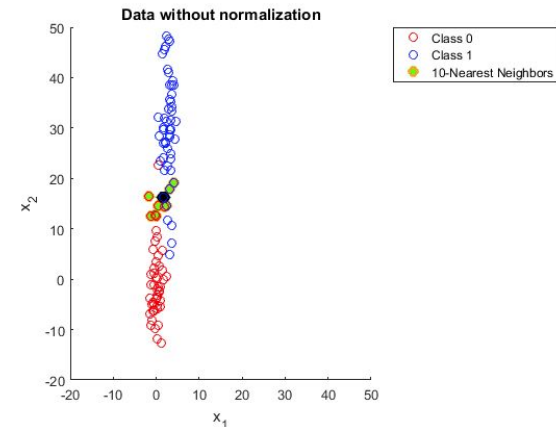
Если мы возьмем признаки с разным масштабом:

- Количество комнат от 1 до 5
- Расстояние до остановки от 10 до 500 м

То расстояние до метро будет вносить больший вклад в ответ (стоимость квартиры)

Решение: вычитаем из всех признаков среднее,
делим на стандартное отклонение.

$$X = \frac{X - E(X)}{\sigma(X)}$$



КЛАССИФИКАЦИЯ ЗАДАЧ



ОБУЧЕНИЕ С УЧИТЕЛЕМ

(supervised learning)

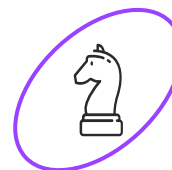
- КЛАССИФИКАЦИЯ
- РЕГРЕССИЯ
- РАНЖИРОВАНИЕ



ОБУЧЕНИЕ БЕЗ УЧИТЕЛЯ

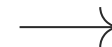
(unsupervised learning)

- ОЦЕНКА ПЛОТНОСТИ
- СНИЖЕНИЕ РАЗМЕРНОСТИ
- КЛАСТЕРИЗАЦИЯ



ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ

(reinforcement learning)



ЭТАПЫ РЕШЕНИЯ ЗАДАЧИ

1

Подготовка данных

2

Выбор модели
(метода)

3

Обучение

4

Тестирование
(подсчет метрик)

ИНСТРУМЕНТЫ

Python, Google Colab, kaggle,

Яндекс контекст, github

ПЛАН КУРСА

Lec 1



Введение

Sem 1



numpy

Sem 2



pandas

Lec 2



Постановка задачи,
Линейная регрессия

Sem 3



Постановка задачи,
Линейная регрессия

Lec 3



Классификация,
регрессия,
градиентный спуск

Sem 4



Градиентный спуск

Lec 4



Sklearn, (бустинги,
возможно)

