

---

---

# knn

*k Nearest Neighbor*  
*к Ближайших Соседей*

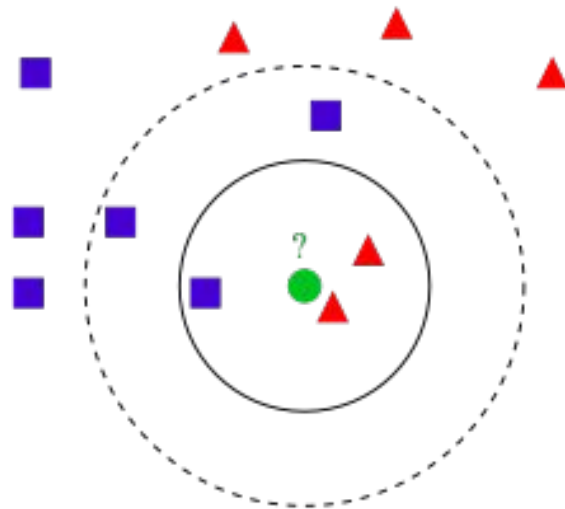
---

---

# KNN

KNN - одна из самых простых и интуитивно понятных моделей классификации

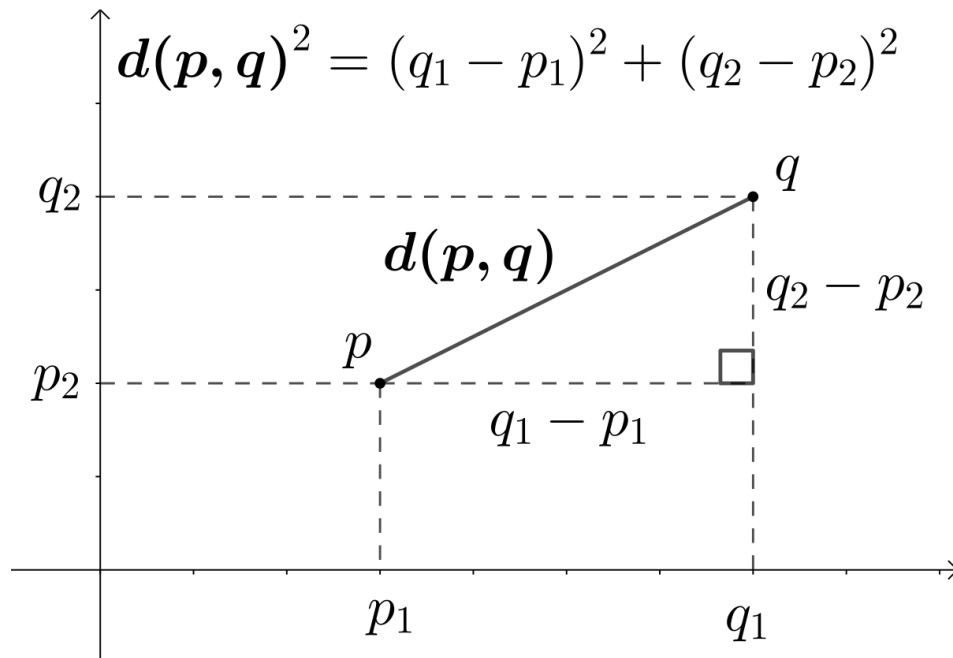
- Модель содержит все объекты из обучающей выборки
- Чтобы классифицировать новый объект нужно найти K ближайших объектов из обучающей выборки. Класс нового объекта - самый частый класс соседей



# Параметры

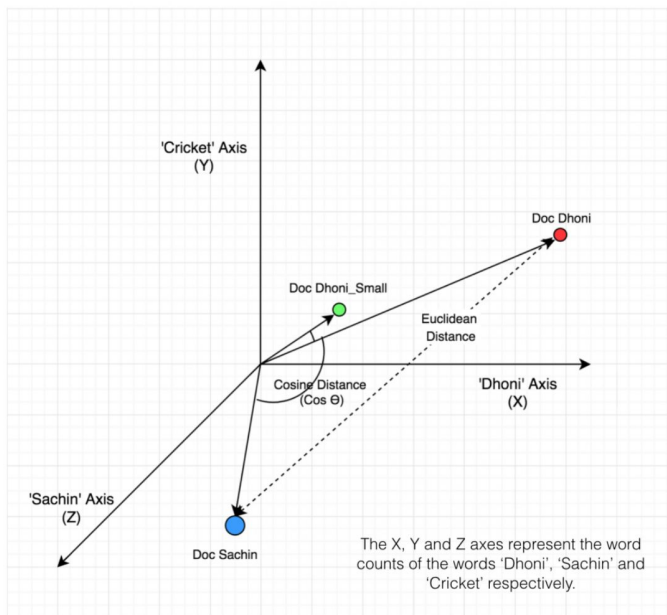
- К - гиперпараметр
- Метрика (расстояние)
  - Евклидово расстояние
  - Косинусное расстояние
  - Манхэттенское расстояние

# Евклидово расстояние



# Косинусное расстояние

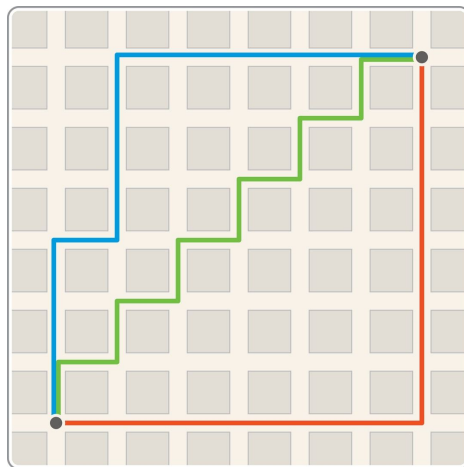
Чаще используется в разреженном пространстве



$$\text{similarity}(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}}$$

# Манхэттенское расстояние

$$d(x, y) = \sum_{i=1}^m |x_i - y_i|$$



# Нормализация

Для KNN данные необходимо нормализовывать для снижения зависимостей от атрибутов с большим диапазоном значений.

Разные атрибуты - разный диапазон значений.

## MinMax нормализация

$$x' = (x - \min[X]) / (\max[X] - \min[X])$$

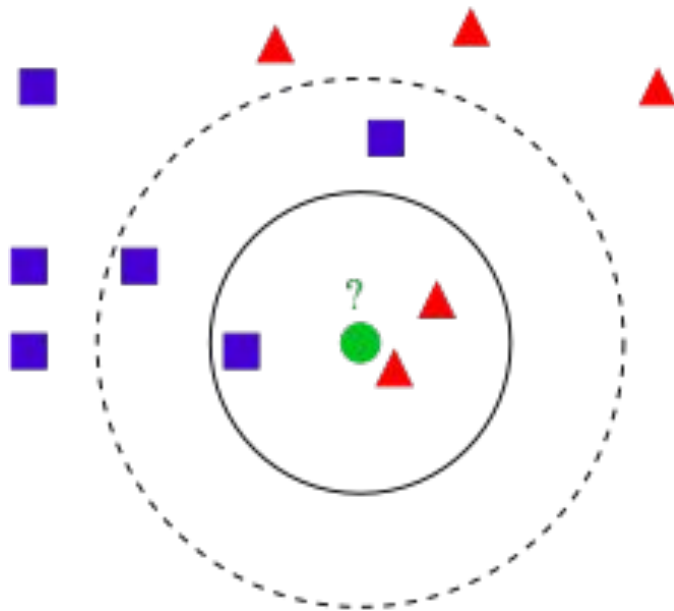
## Z нормализация

$$x' = (x - M[X]) / \sigma[X]$$

# Применение

## Классификация

Чтобы классифицировать новый объект нужно найти  $K$  ближайших объектов из обучающей выборки. Класс нового объекта - самый частый класс соседей



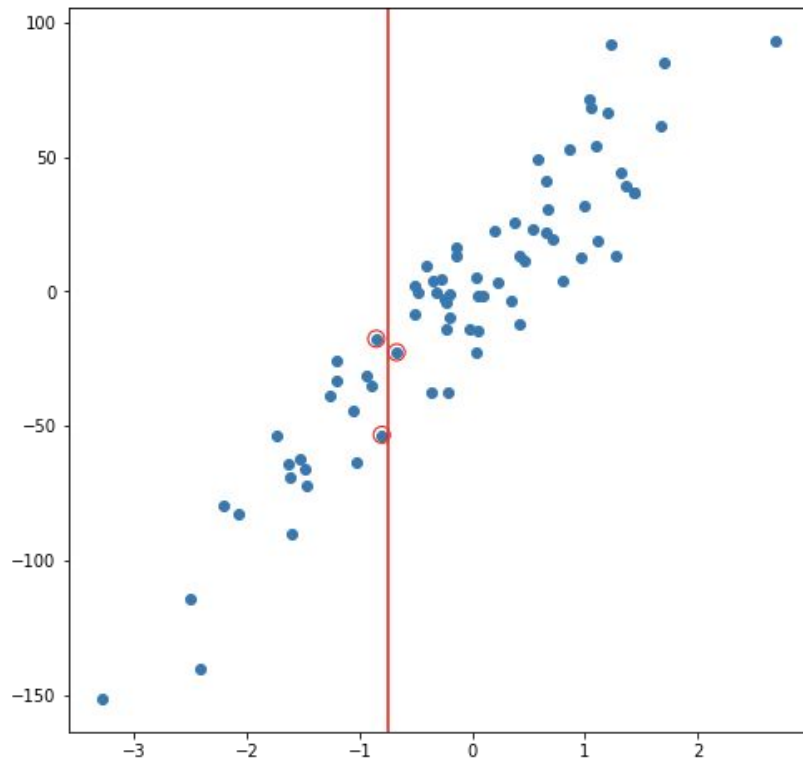


# Применение

## Регрессия

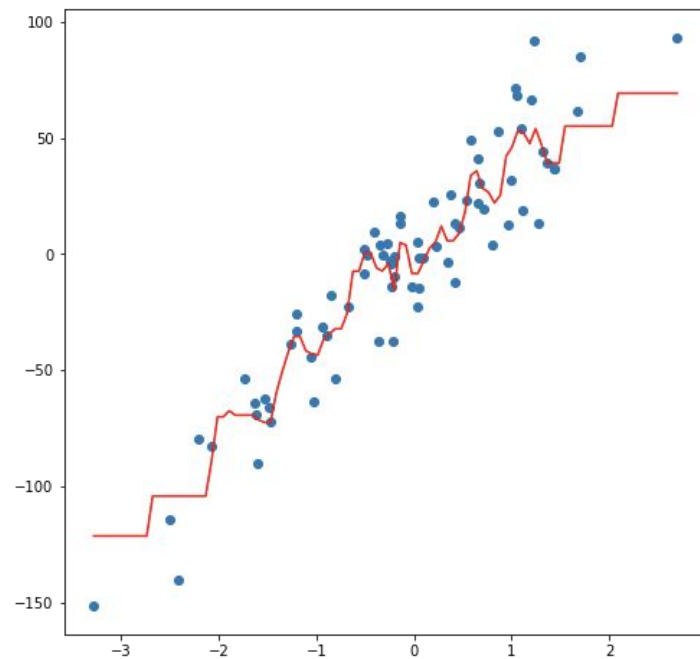
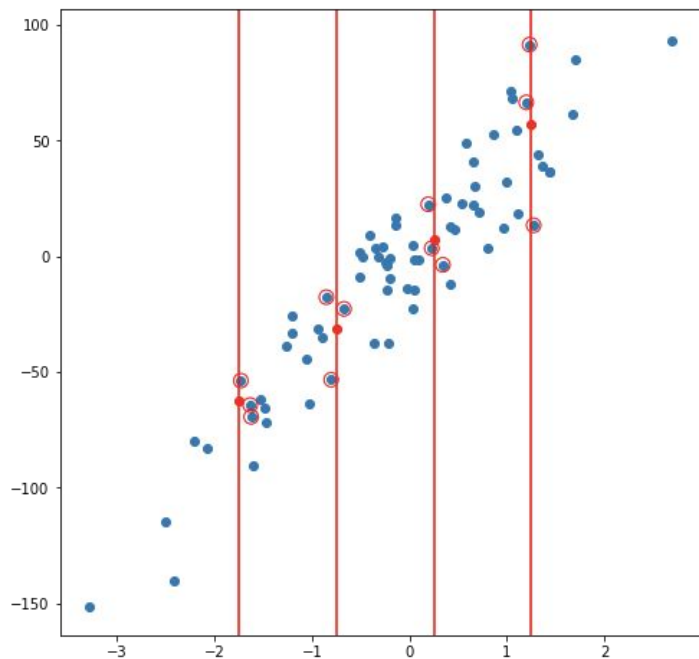
Нахождение точек, наиболее близких с выбранному значению.

Итоговое значение - среднее между ближайшими.



# Применение

## Регрессия



# Преимущества и недостатки

## Преимущества

- Легок в реализации
- Не чувствителен к выбросам
- Нет необходимости строить модель, настраивать несколько параметров
- Алгоритм универсален. Его можно использовать для обоих типов задач: классификации и регрессии.

# Преимущества и недостатки

## Недостатки

- Возрастание вычислительных затрат при увеличении выборки
- Сложно интерпретировать
- Всегда нужно определять оптимальное значение **k**