

Санкт-Петербургский государственный университет

Математико-механический факультет

Бадмаев Чингис Юрьевич

# Кластеризация (k-means)

Практическая работа

Санкт-Петербург  
2021

# Оглавление

1. Постановка задачи	3
2. Теорминимум	4
3. Тесты	5
4. Ссылка на код	6

# 1. Постановка задачи

В данном задании речь идет о решении задачи кластеризации методом k-средних.

## 2. Теорминимум

Выбираем начальные центры кластеров. В наших тестах будем использовать два способа выбора начальных центров: случайный выбор и выбор центров, равных максимуму/минимуму по координатам.

На каждой итерации:

- Определяем кластер, к которому относится точка

$$l_j = \arg \min_{i=1, \dots, k} \rho(x_j, c_i),$$

где  $l_j$  — метка кластера,  $c_i$  — центр кластера,  $\rho(x_j, c_i)$  — функция расстояния. В наших тестах будем использовать две функции расстояния: евклидово расстояние и расстояние городских кварталов.

- Пересчитываем координаты нового центра каждого из кластеров, используя среднее арифметическое.

Продолжаем процесс до тех пор, пока составы кластеров не перестанут меняться.

### 3. Тесты

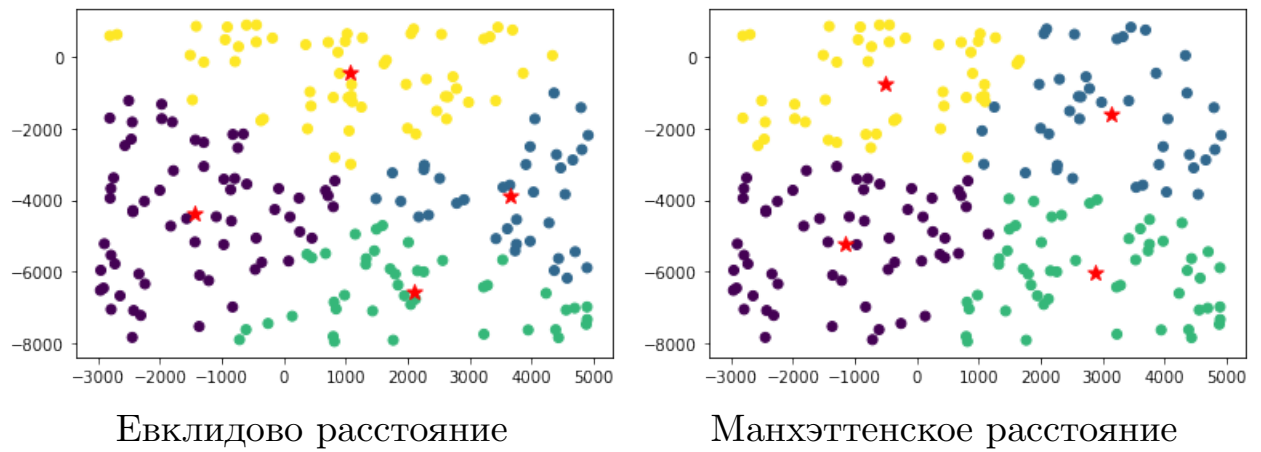


Рис. 1: Результаты кластеризации при случайном выборе начальных центров.

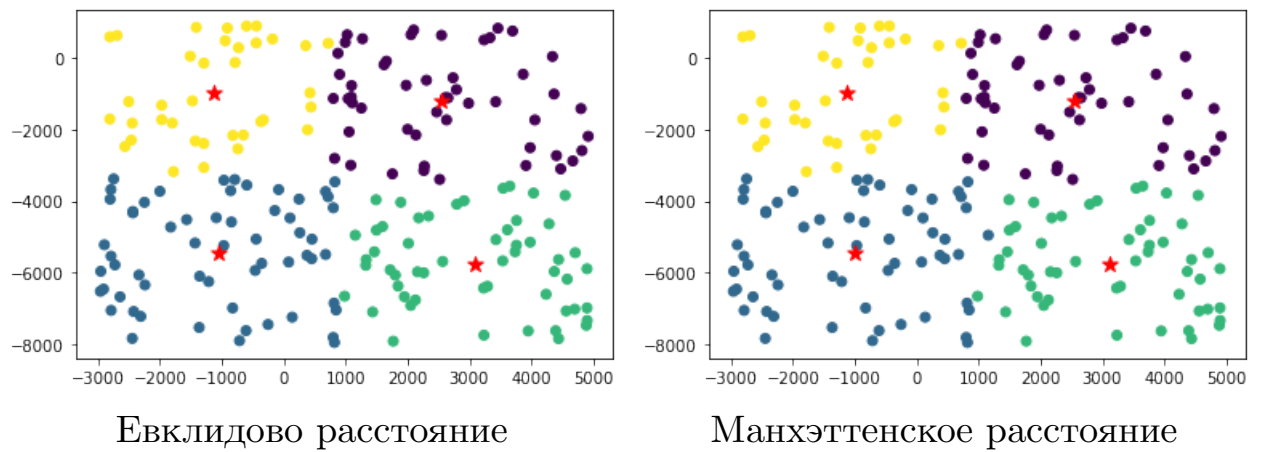


Рис. 2: Результаты кластеризации при выборе начальных центров, равных максимуму/минимуму по координатам.

## 4. Ссылка на код

[Ссылка](#)