

Санкт-Петербургский государственный университет
Математико-механический факультет
Информационно-аналитические системы

Ким Юния Александровна
18.Б07-мм

Вычислительный практикум

Отчёт по заданию №12

Преподаватель:
Евдокимова Т.О.

Санкт-Петербург
2021

Содержание

1. Ссылка на код	3
2. Постановка задачи	3
3. Теоретическая часть	3
4. Численный эксперимент	3
4.1. Описание	3
4.2. Результаты	3

1. Ссылка на код

<https://github.com/yuniyakim/MethodsOfComputation/pull/20>

2. Постановка задачи

Задача – реализация метода k-средних решения задачи кластеризации.

3. Теоретическая часть

Выбираем начальные центры кластеров. В наших тестах будем использовать два способа выбора начальных центров: случайный выбор и выбор центров, равных минимуму/максимуму 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 тестах будем использовать две функции расстояния: евклидово расстояние и расстояние городских кварталов (манхэттенское расстояние).

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

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

4. Численный эксперимент

4.1. Описание

Для численного эксперимента генерировался массив из 200 случайных точек с дробными координатами от 0 до 200.

4.2. Результаты

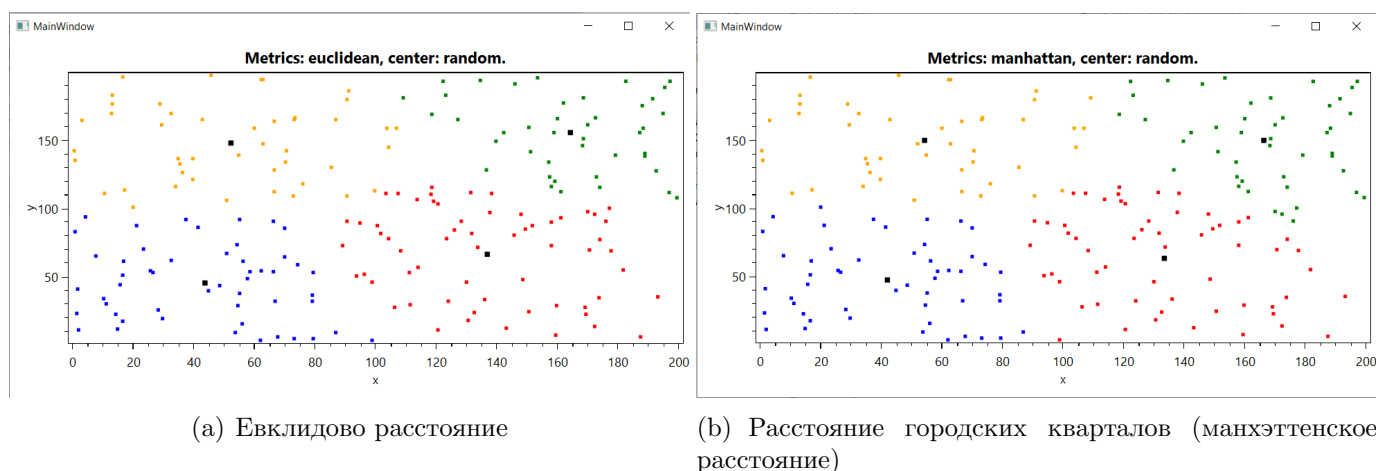
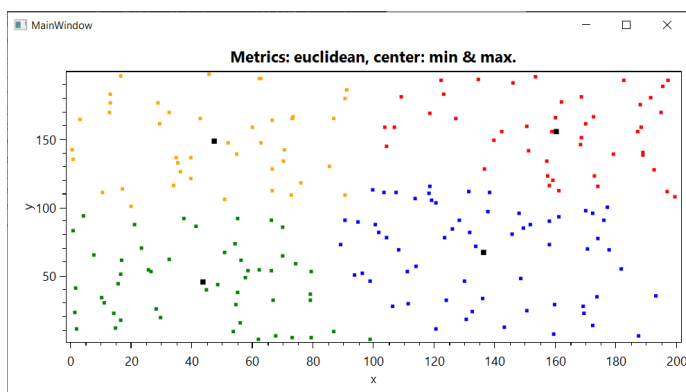
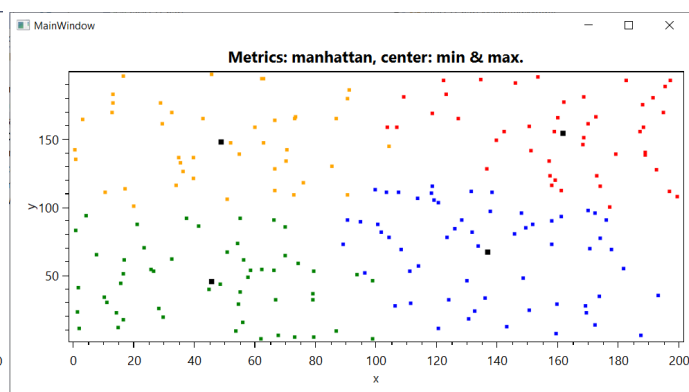


Рисунок 4.1. Результаты кластеризации при случайном выборе центров



(a) Евклидово расстояние



(b) Расстояние городских кварталов (манхэттенское расстояние)

Рисунок 4.2. Результаты кластеризации при выборе центров, равных минимуму/максимуму по координатам