

Task 12

Макуха Илья

August 2021

1 Постановка задачи

Нужно реализовать метод кластеризации KMeans, а также научиться выбирать оптимальное число кластеров.

2 Теоретический минимум

Алгоритм:

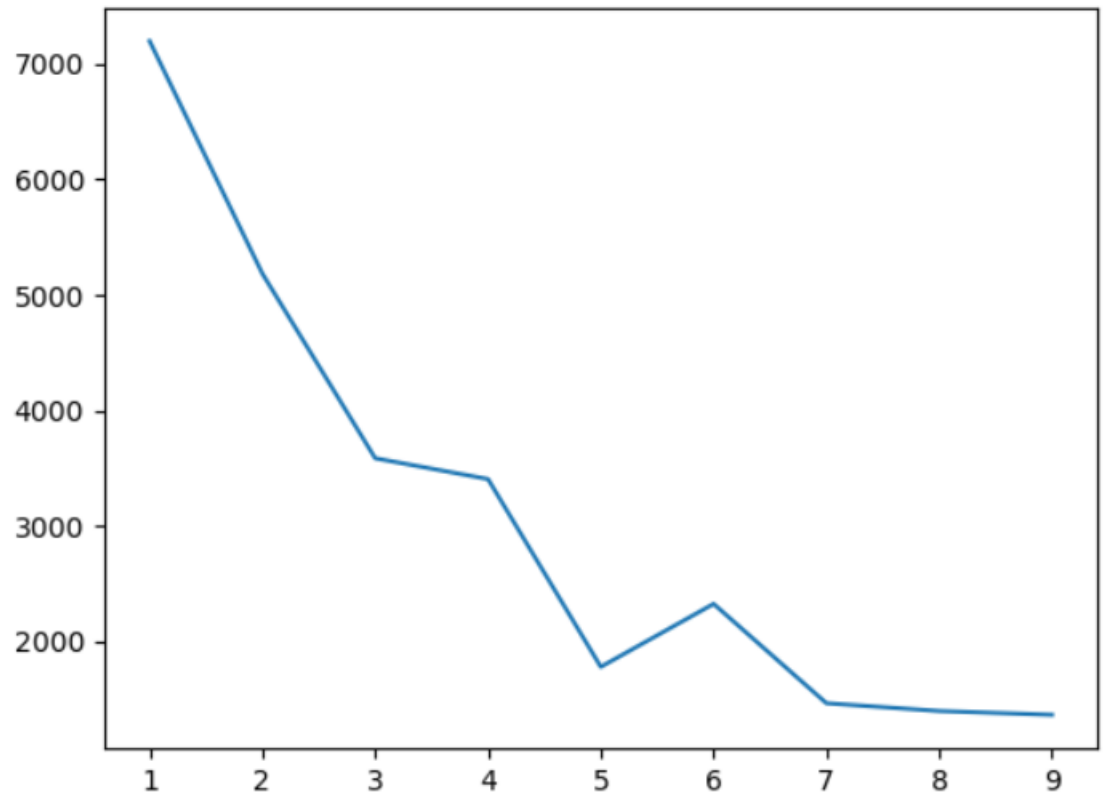
- 1) Выбрать начальные центроиды
- 2) Посчитать расстояния до центроид
- 3) Присвоить метки с минимальным расстоянием
- 4) Усреднить координаты по меткам.
- 5) Заменить старые центроиды на координаты из п.4
- 6) Продолжать до сходимости

Выбор оптимального k:

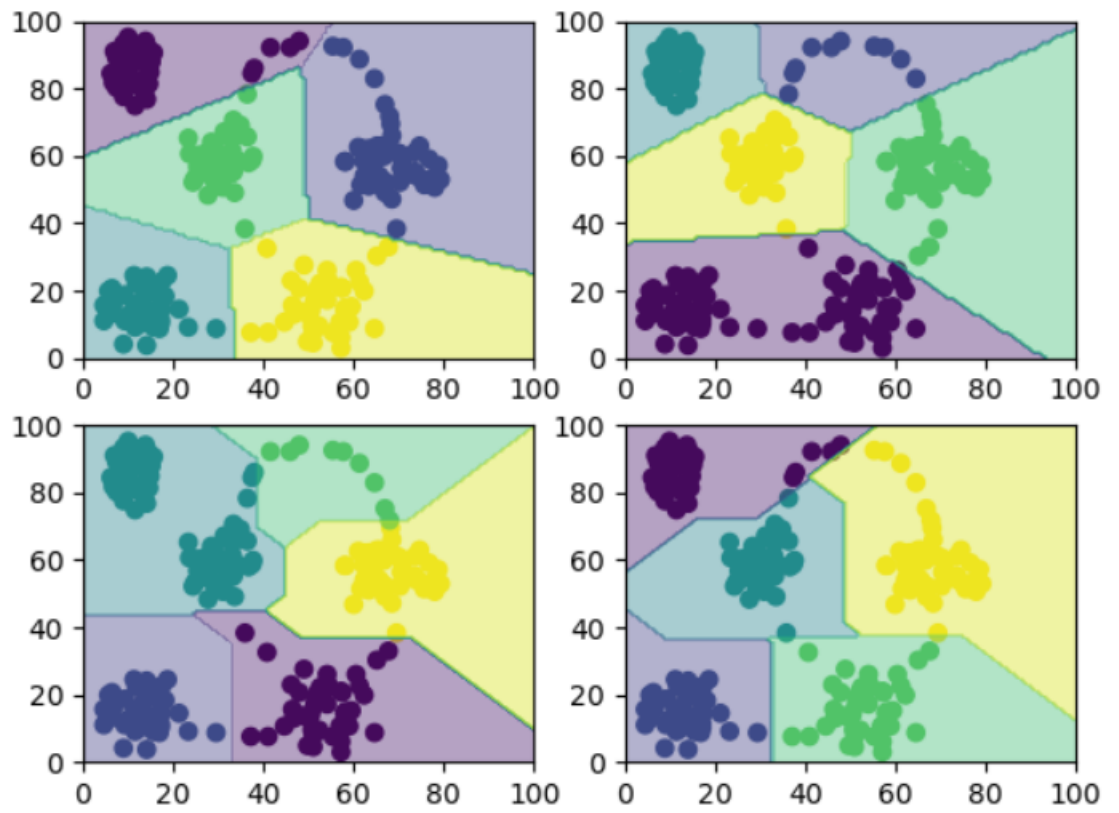
Прогнать алгоритм для разных k. Посчитать inertia. Выбрать локоть.

3 Тесты

Начальные инициализации: нормальная и максимум или минимум(случайно) по признакам. Расстояния: минковского $p=2$ и $p=\inf$ (в коде реализован еще gbf, но там наблюдается чередование двух наборов центроид). Оптимальное $k = 5$.



1-ая строка - евклидово расстояние
2-ая строка - $\max(\text{abs})$
1-ый столбец - нормальная инициализация
2-ой столбец - граничная инициализация



4 Github

https://github.com/MakuhIlyukh/mak_cm