**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра МОЭВМ**

отчет

**по лабораторной работе №5**

**по дисциплине «Машинное обучение»**

Тема: **Кластеризация (k-средних, иерархическая)**

|  |  |  |
| --- | --- | --- |
| Студент гр. 8304 |  | Холковский К.В |
| Преподаватель |  | Жангиров Т. Р. |

Санкт-Петербург

2021

**Цель работы**

Ознакомиться с методами кластеризации модуля Sklearn.

**Ход работы**

1. **Загрузка данных**

Были загружены данные

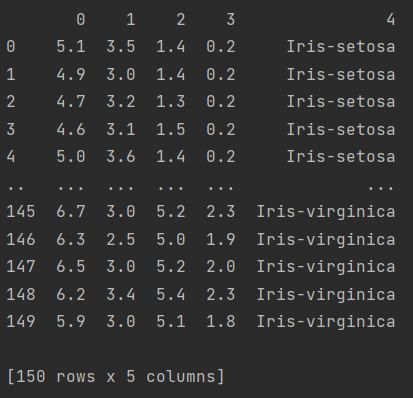


Рис 1 – Загруженные данные

1. **K-means**

Провели кластеризацию методом k-means

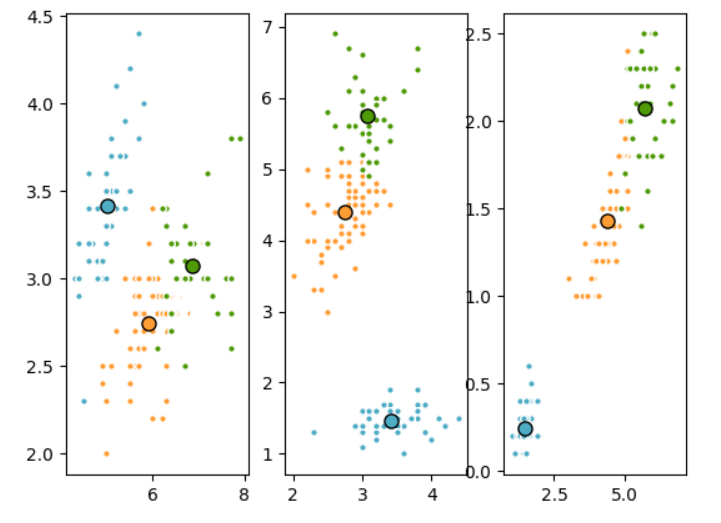


Рис 2 – Результат класстеризации

n\_int – сколько раз перезапускать алгоритм.

Нарисовли карту для всей области значений

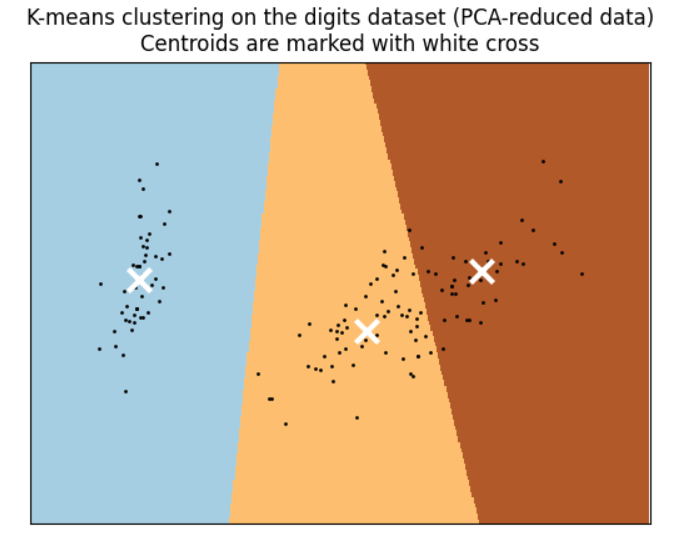


Рис 3 – Результат после PCA и k-means

|  |  |
| --- | --- |
| Методы init | Суть работы |
| k-means++ | Выбирает первый центр случайно. Каждый последующий выбирается так, чтобы вероятность выбора точки была пропорциональна вычисленному для неё квадрату расстояния |
| random | Случайные центры |
| массив точек | Заданные центры |

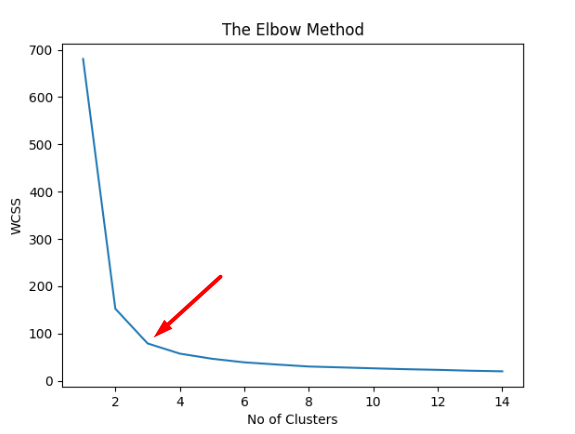


Рис 4 – Результат определения числа кластеров методом локтя

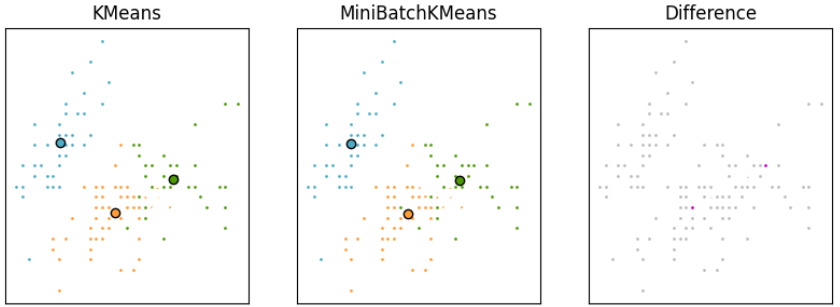


Рис 5 – Рзличие результатов работы методов k-means

1. **Иерархическая кластеризация**

Была проведена иерархическая кластеризация

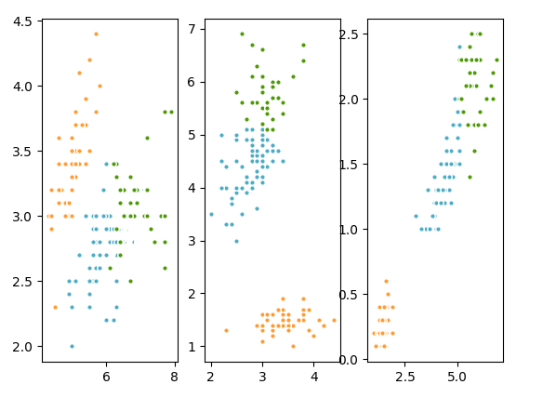


Рис 6 – Результат иерархической кластеризации

Выполнили кластеризацию для разного количества кластеров.

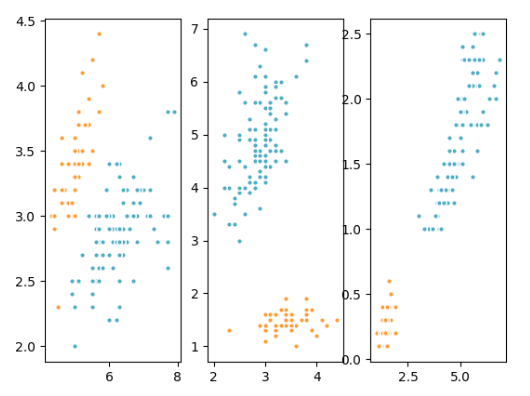


Рис 7 – 2 кластера

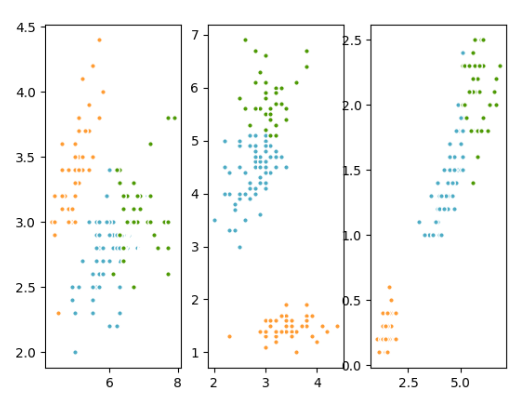


Рис 8 – 3 кластера

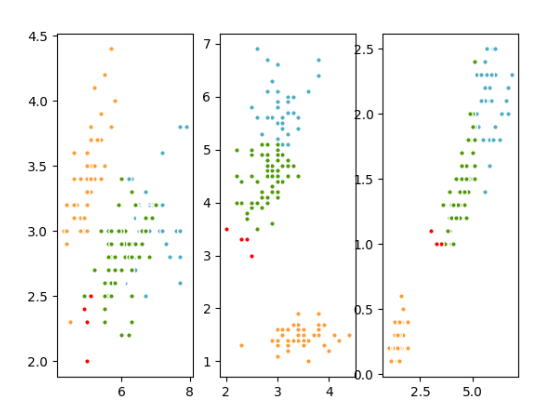


Рис 9 – 4 кластера

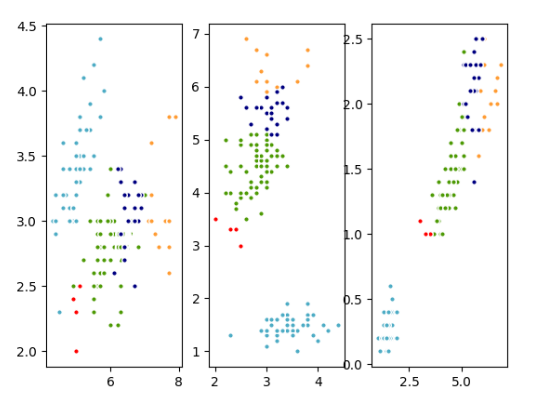


Рис 10 – 5 кластеров

Была нарисована дендограмма до 6-го уровня

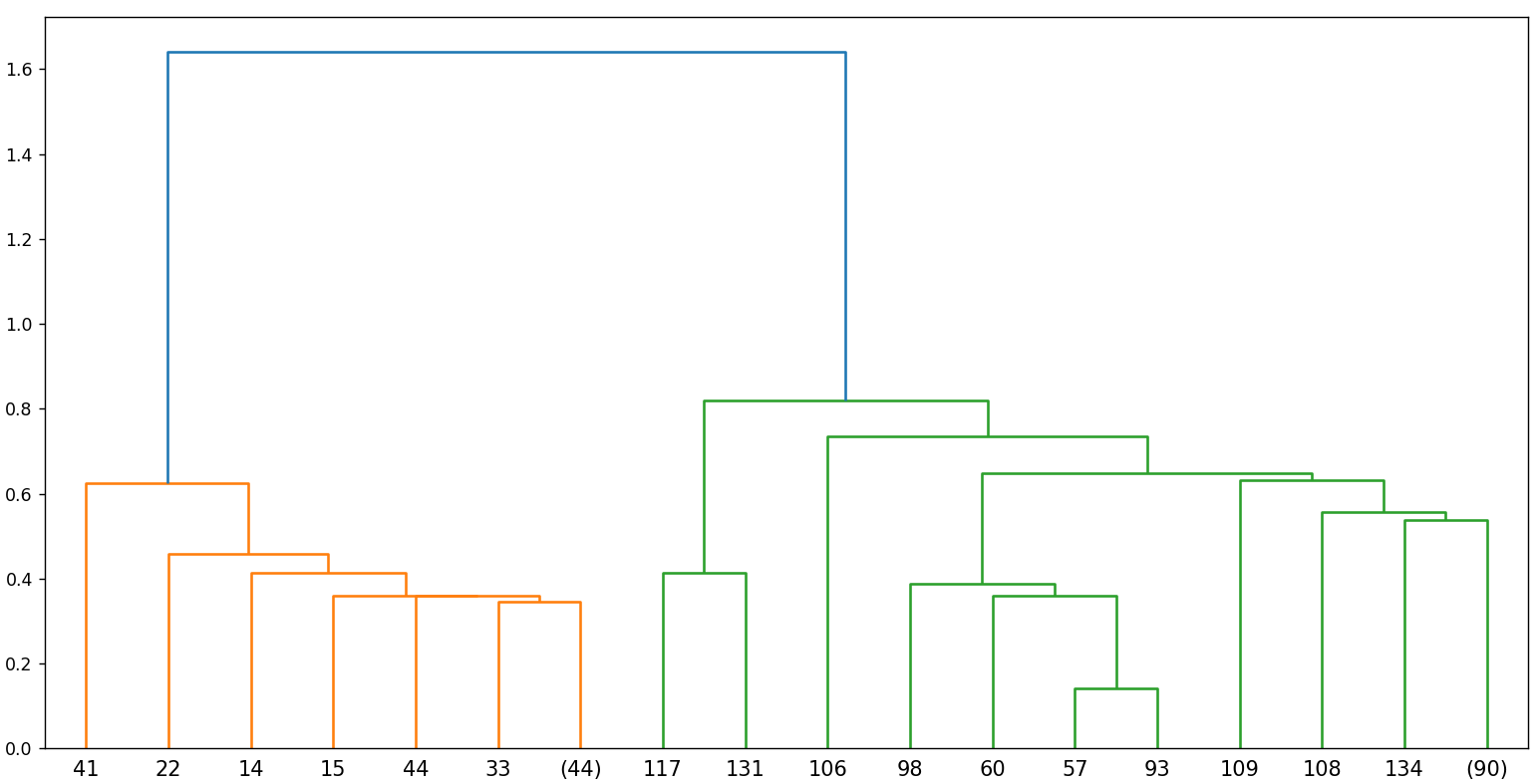


Рис 11 – Дендограмма

Исследовали кластеризацию при всех параметрах linkage

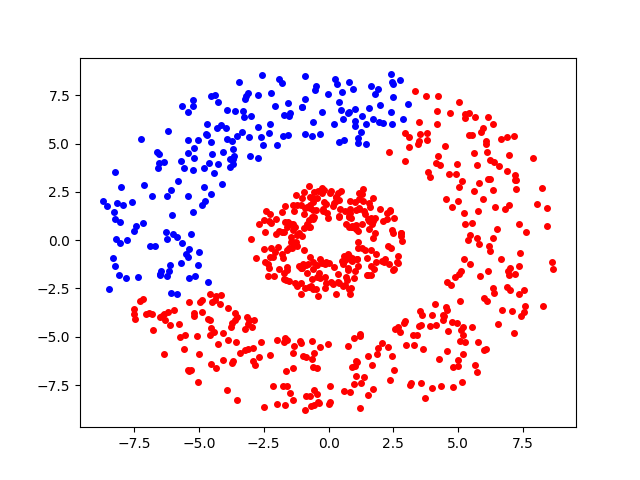


Рис 12 – ward

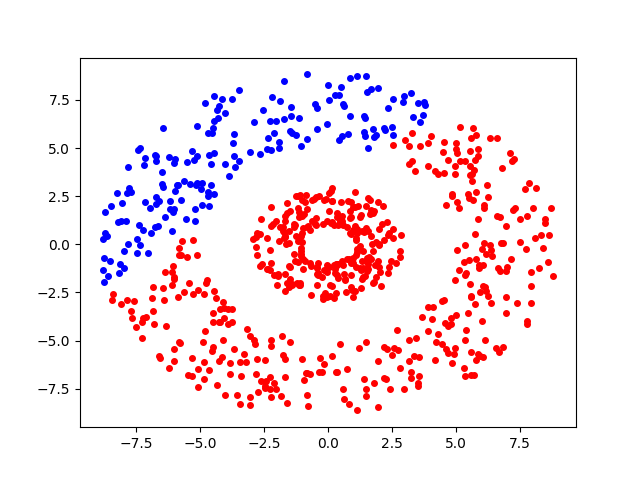


Рис 13 – average

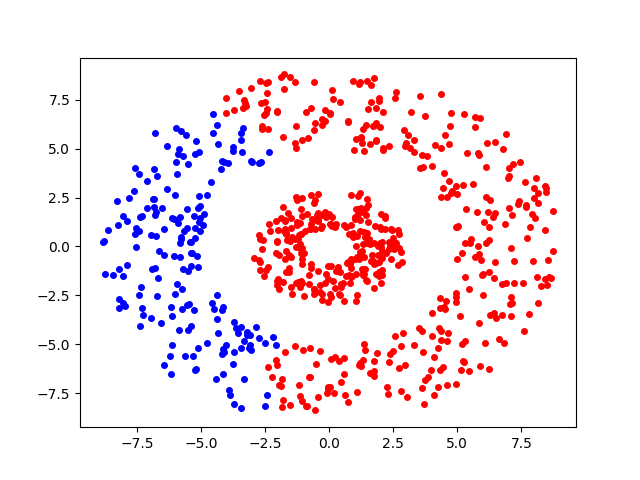


Рис 14 – complete

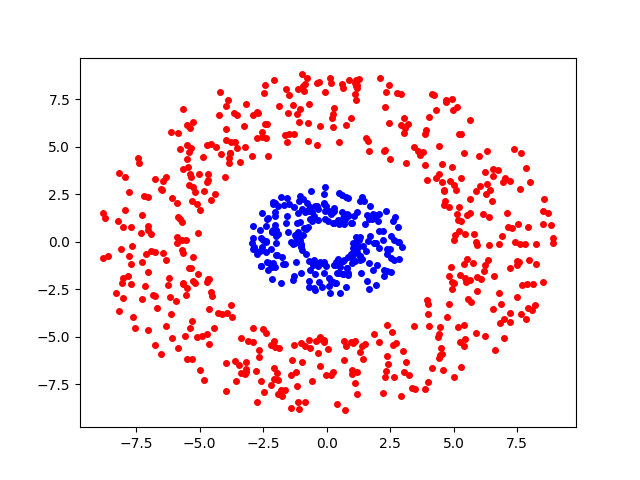


Рис 15 – single

Лучший результат кластеризации был достигнут при linkage=’single’.

**Вывод**

Ознакомились с методами кластеризации модуля Sklearn.