

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 (к-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

```
      0      1      2      3      4
0    5.1    3.5    1.4    0.2  Iris-setosa
1    4.9    3.0    1.4    0.2  Iris-setosa
2    4.7    3.2    1.3    0.2  Iris-setosa
3    4.6    3.1    1.5    0.2  Iris-setosa
4    5.0    3.6    1.4    0.2  Iris-setosa
..    ...    ...    ...    ...    ...
145   6.7    3.0    5.2    2.3  Iris-virginica
146   6.3    2.5    5.0    1.9  Iris-virginica
147   6.5    3.0    5.2    2.0  Iris-virginica
148   6.2    3.4    5.4    2.3  Iris-virginica
149   5.9    3.0    5.1    1.8  Iris-virginica

[150 rows x 5 columns]
```

Рис 1 – Загруженные данные

2. K-means

Провели кластеризацию методом k-means

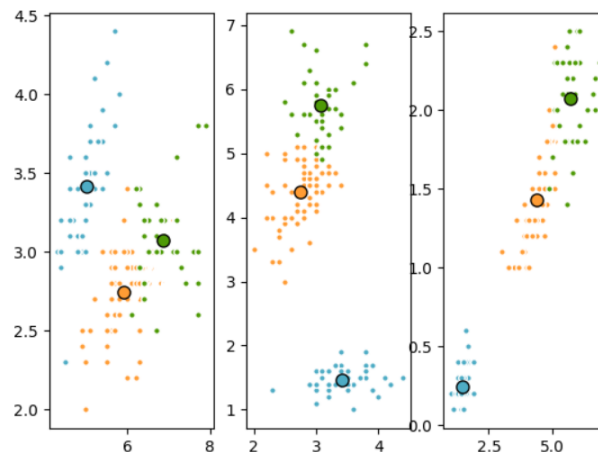


Рис 2 – Результат кластеризации

n_int – сколько раз перезапускать алгоритм.

Нарисовли карту для всей области значений

K-means clustering on the digits dataset (PCA-reduced data)
Centroids are marked with white cross

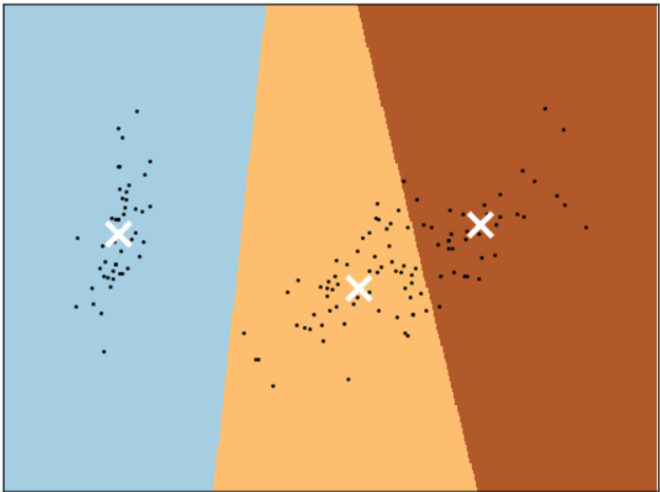


Рис 3 – Результат после PCA и k-means

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

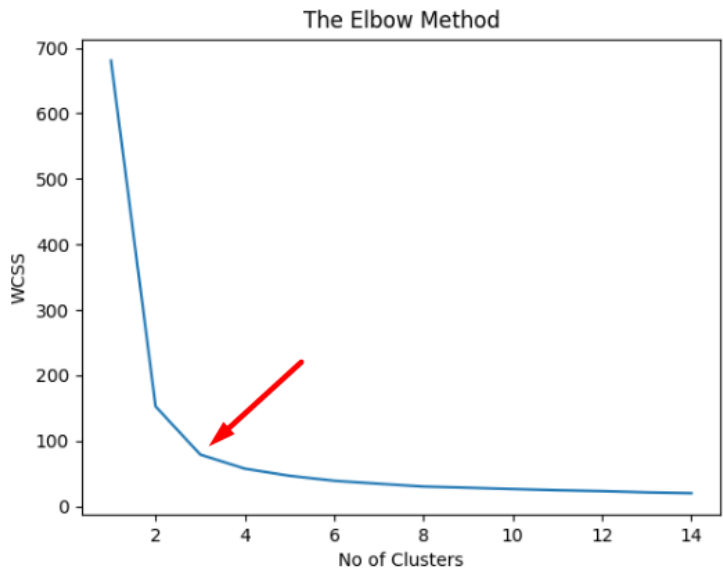


Рис 4 – Результат определения числа кластеров методом локтя



Рис 5 – Различие результатов работы методов k-means

3. Иерархическая кластеризация

Была проведена иерархическая кластеризация

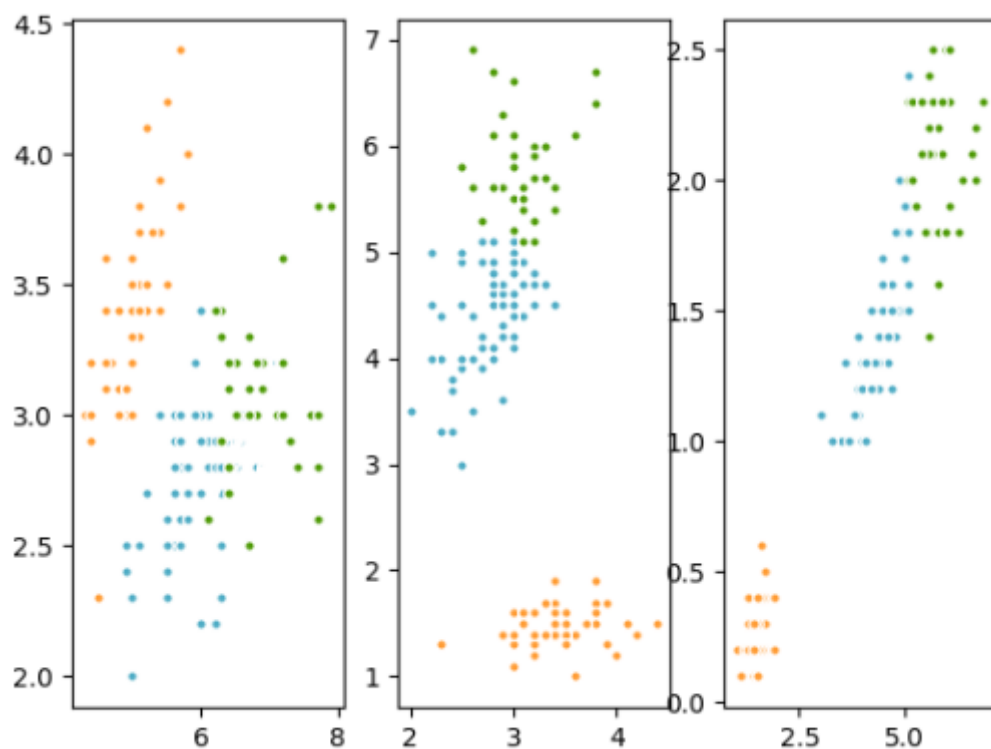


Рис 6 – Результат иерархической кластеризации

Выполнили кластеризацию для разного количества кластеров.

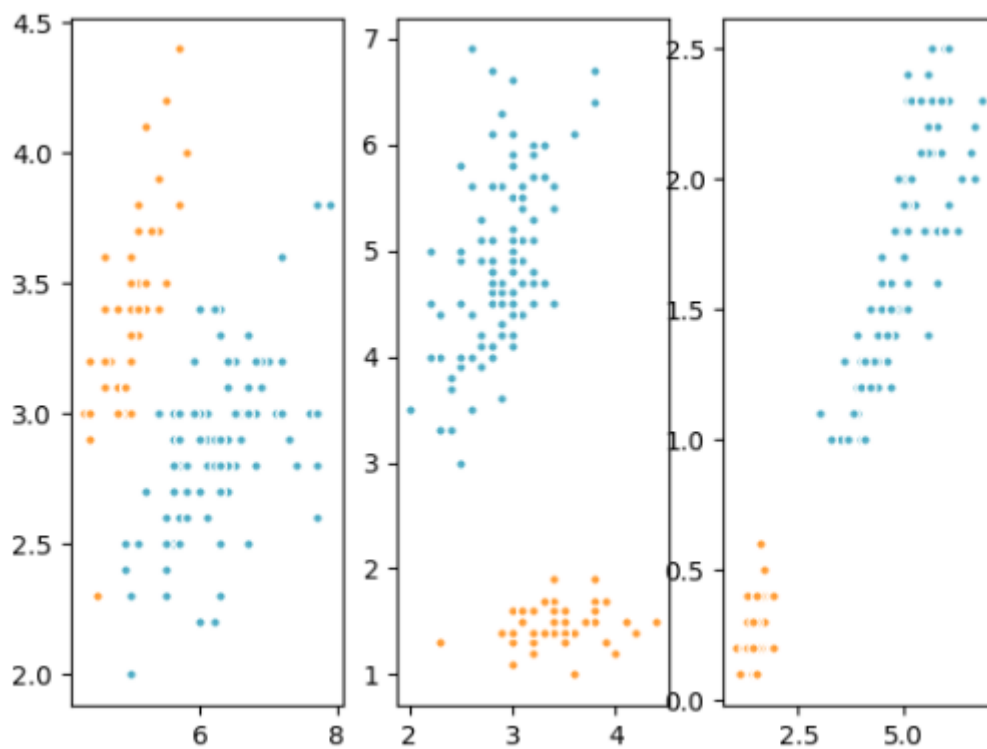


Рис 7 – 2 кластера

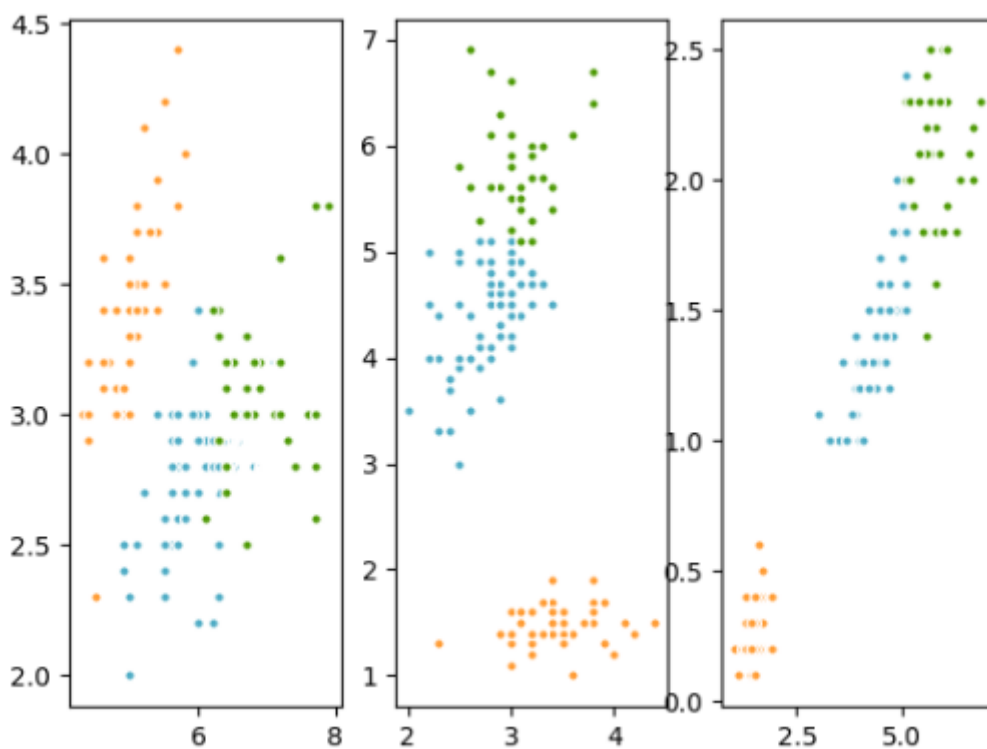


Рис 8 – 3 кластера

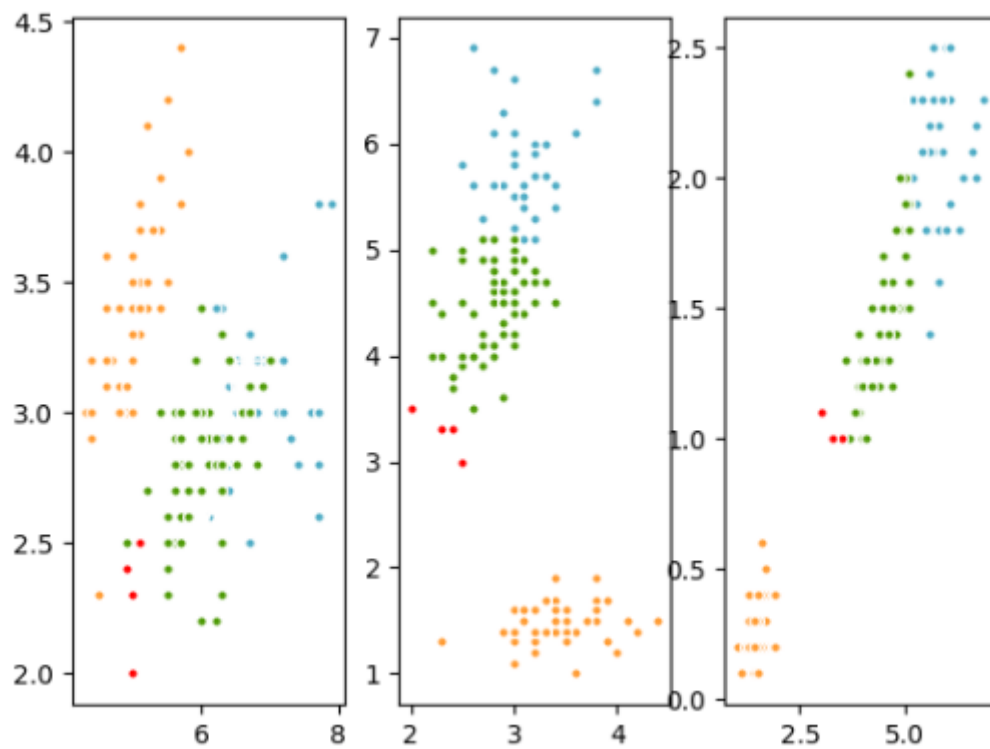


Рис 9 – 4 кластера

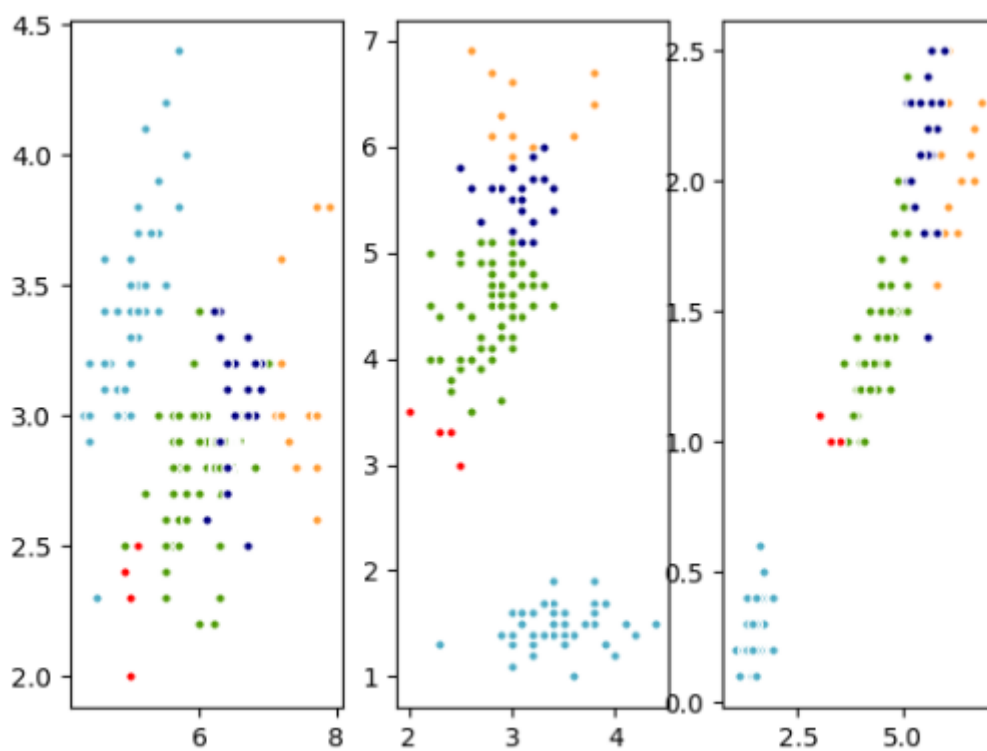


Рис 10 – 5 кластеров

Была нарисована дендограмма до 6-го уровня

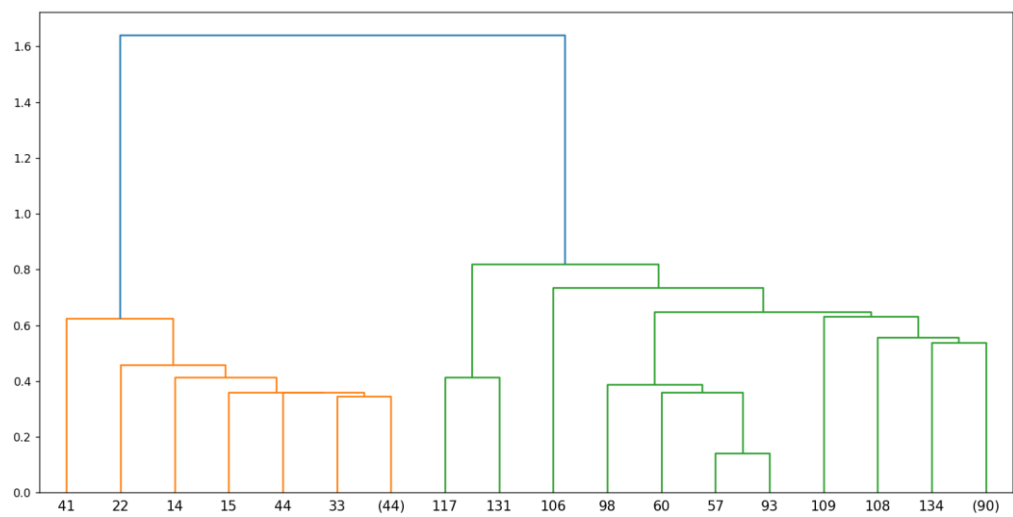


Рис 11 – Дендограмма

Исследовали кластеризацию при всех параметрах linkage

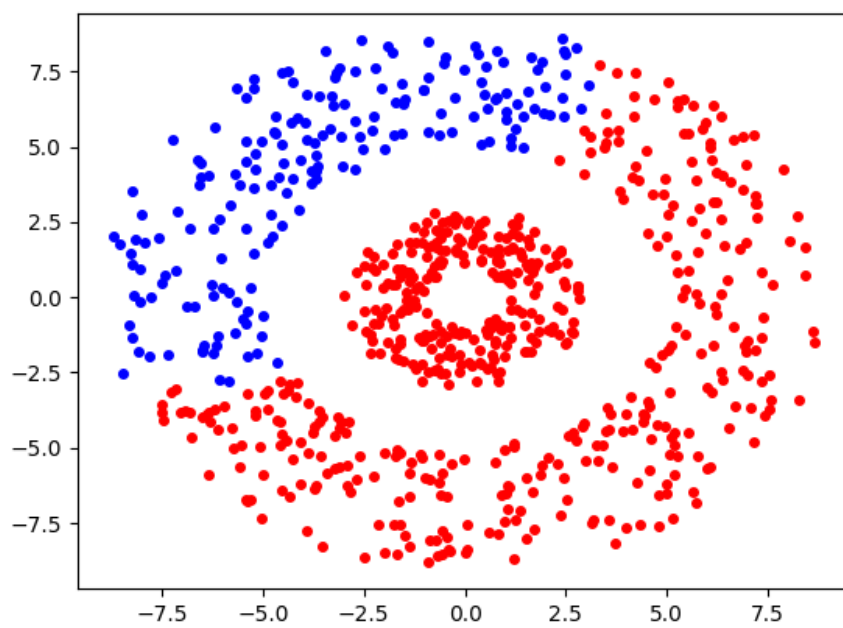


Рис 12 – ward

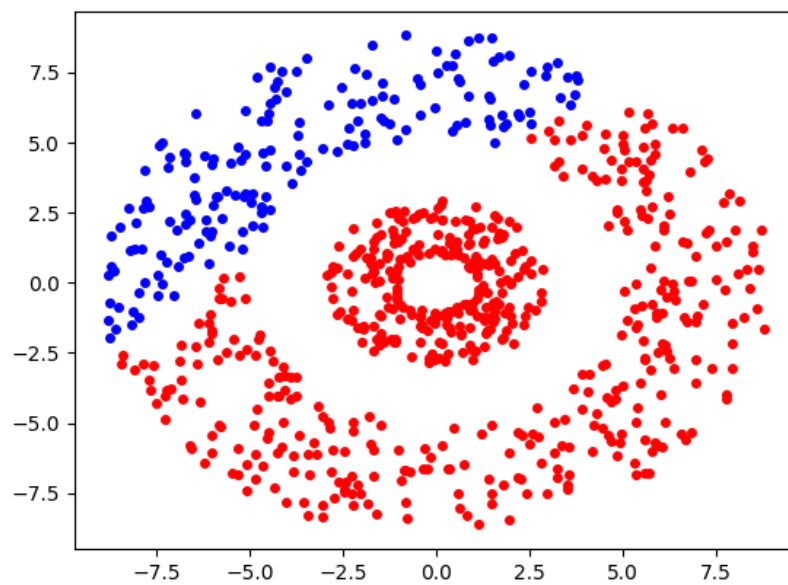


Рис 13 – average

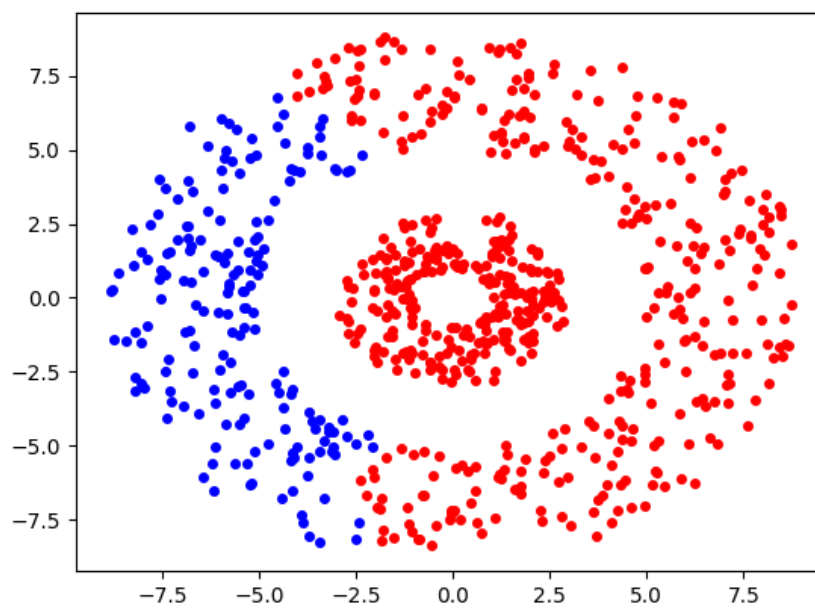


Рис 14 – complete

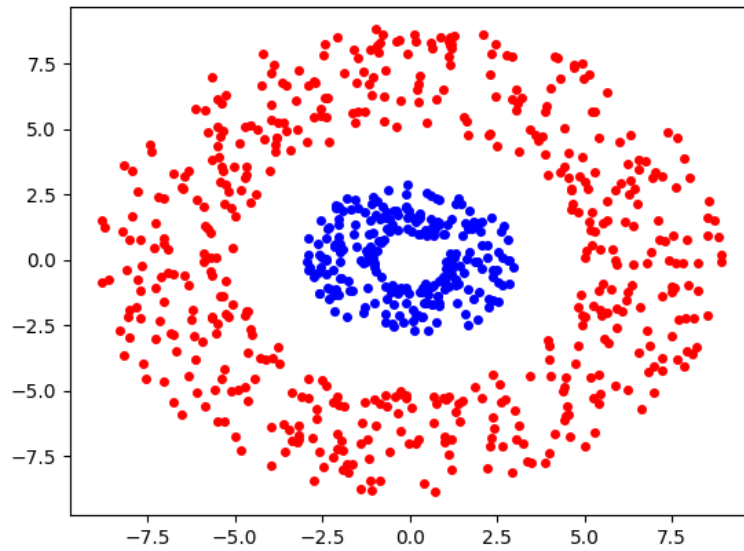


Рис 15 – single

Лучший результат кластеризации был достигнут при `linkage='single'`.

Вывод

Ознакомились с методами кластеризации модуля Sklearn.