|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра прикладной математики (ПМ)**

**ОТЧЕТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ №8**

по дисциплине «Языки программирования для статистической обработки данных»

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Студент группы ИНБО-01-17 | *ИМБО-11-23, Журавлев Ф. А.* | (подпись) | |  |
| Преподаватель | *Трушин СМ* | (подпись) | |  |
|  |  | |  | |

Москва 2025 г.

## 1) ЦЕЛЬ И ЗАДАЧИ

**Цель практической работы:**

Научиться проводить кластеризацию данных с использованием методов K-means и иерархической кластеризации в Python, R, а также визуализировать результаты кластерного анализа.

**Задачи практической работы:**

1. Выполнить кластеризацию методом K-means:

• Python: использование библиотеки sklearn.

• R: функция kmeans.

2. Провести иерархическую кластеризацию:

• Python: библиотека scipy.cluster.hierarchy.

• R: функции hclust, dendrogram.

3. Проанализировать результаты кластеризации:

• Интерпретация кластеров (центроиды, количество объектов в каждом кластере).

• Сравнение кластеров, полученных разными методами.

4. Визуализировать результаты кластеризации:

• Python: графики кластеров с использованием matplotlib и seaborn. • R: графическое представление дендрограмм и кластеров.

5. Сравнить удобство выполнения кластерного анализа в Python, R.

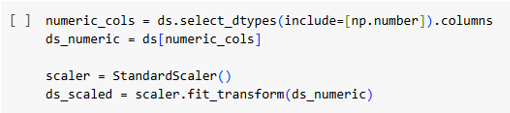
## 2) РЕЗУЛЬТАТЫ ПРАКТИКИ

**Шаг 1) Кластеризация в Python**

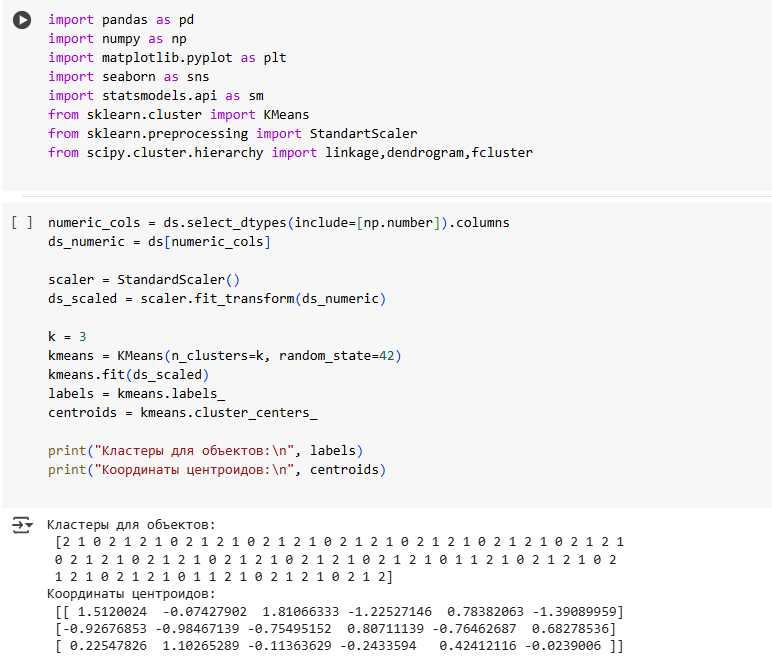
### 1.1) Иерархическая и K-means кластеризации в Python.

После загрузки исходной таблицы данных в формате .csv, следует провести кластеризацию двумя способами: K-means и иерархическую, но перед любой кластеризацией первым делом необходимо масштабировать исходный набор данных. Ниже приведен коротенький код:

*Рисунок 1.1 — Мастштабирование данных.*

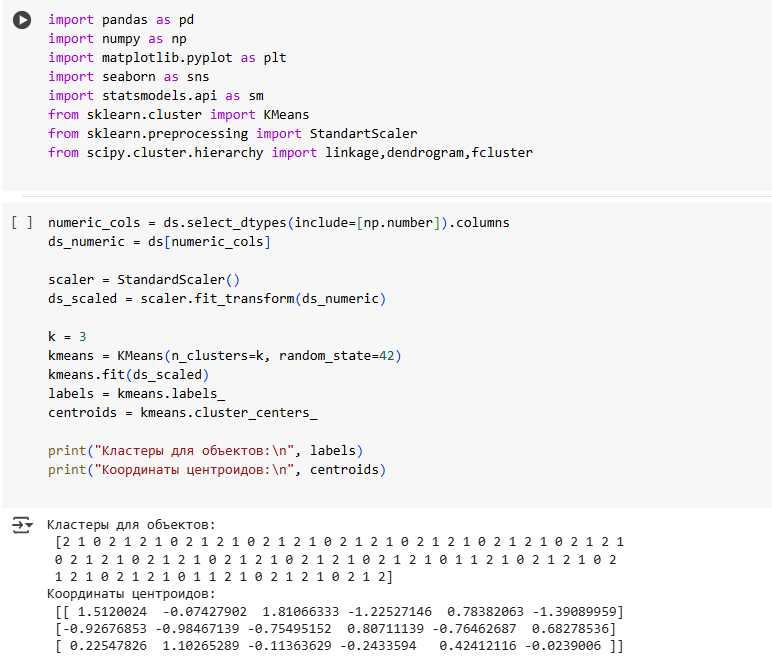


Далее напишем код, который реализует кластеризацию K-means и иерархическую, перед этим убедившись, что были добавлены все необходимые библиотеки:

*Рисунок 1.2 — K-means кластеризация.*

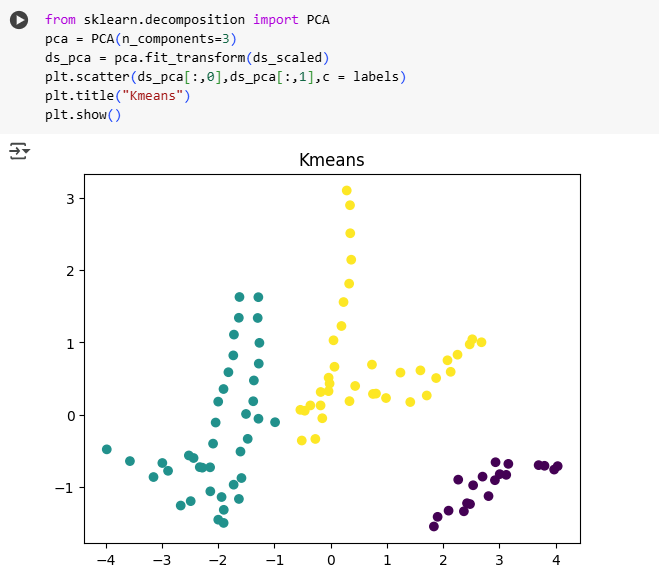
Вот какой результат выводит этот код:

*Рисунок 1.3 — Результат K-means.*



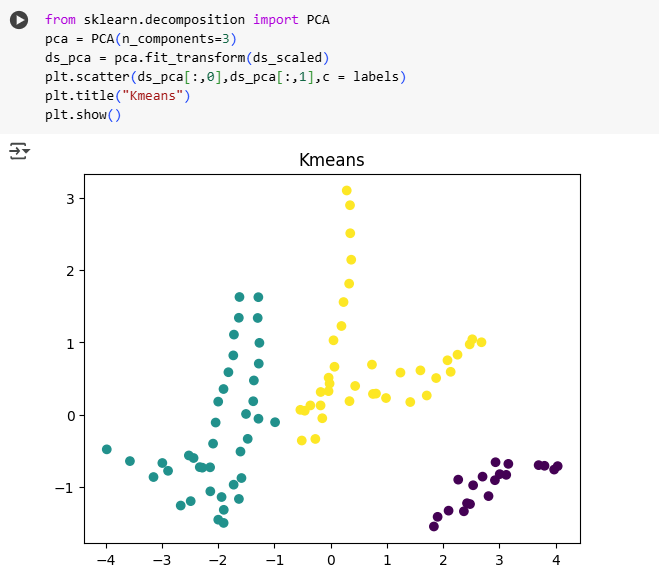
Все результаты и итоги подведем в параграфе 3 «Сравнение результатов», а далее напишем код иерархической кластеризации.

*Рисунок 1.4 — Код графика кластеризации.*

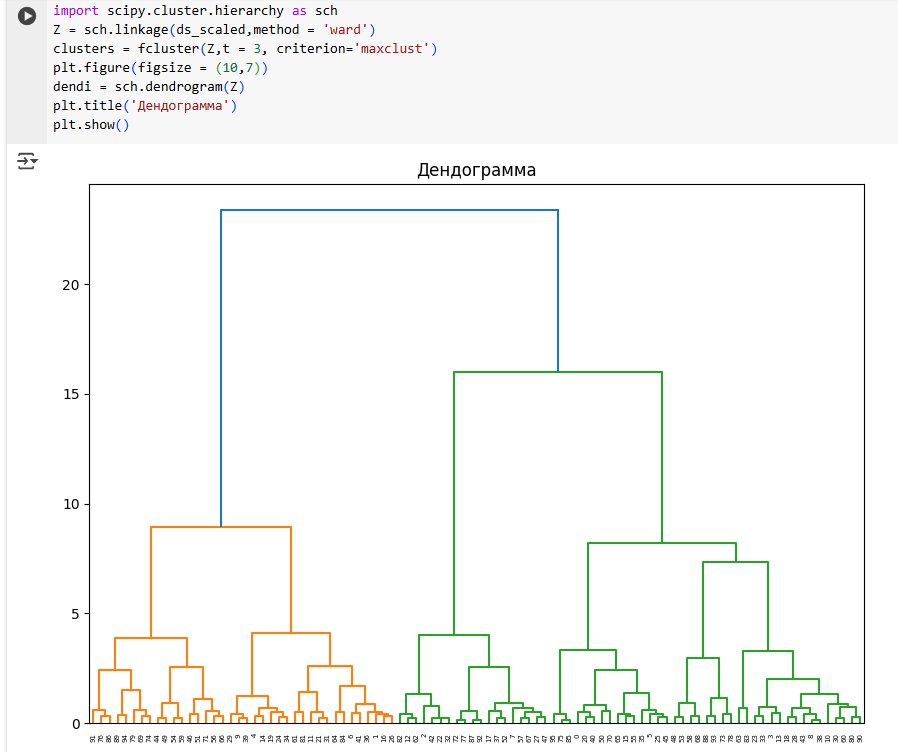


Далее рассмотрим график, который получился в результате K-means кластеризации указав, что количество кластеров равняется трем. Затем, можно построить будет Дендограмму, для более наглядной визуализации.

*Рисунок 1.5 — Код иерархической кластеризации.*

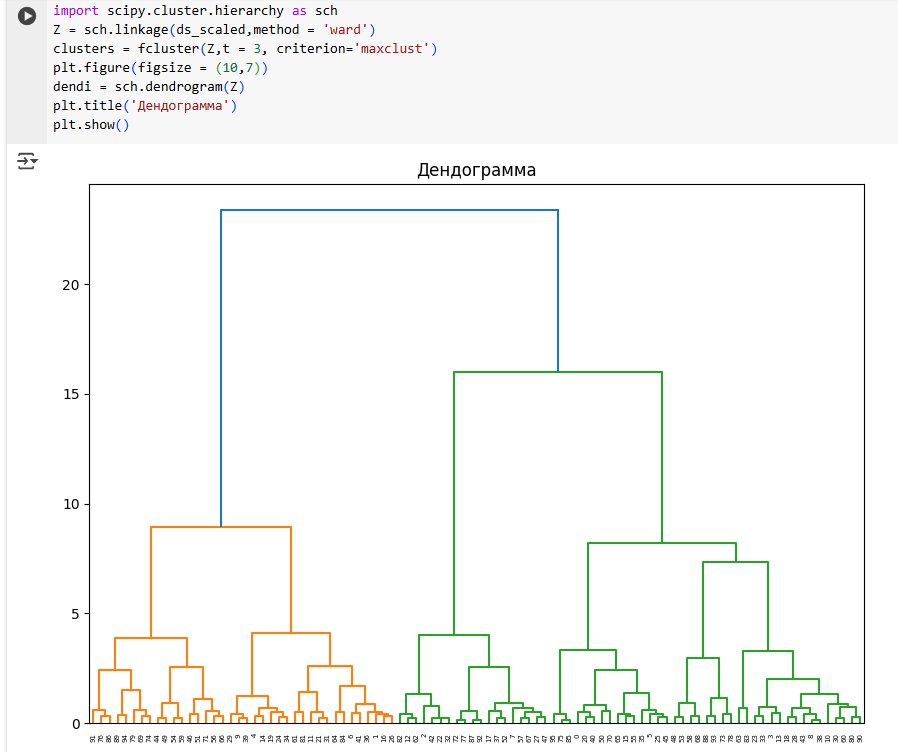


Далее построим дендограмму:

*Рисунок 1.6— Код для построения дендограммы.*

А вот график, который получился в результате написания вот этого кода:

*Рисунок 1.7— Дендограмма(иерархическая кластеризация).*

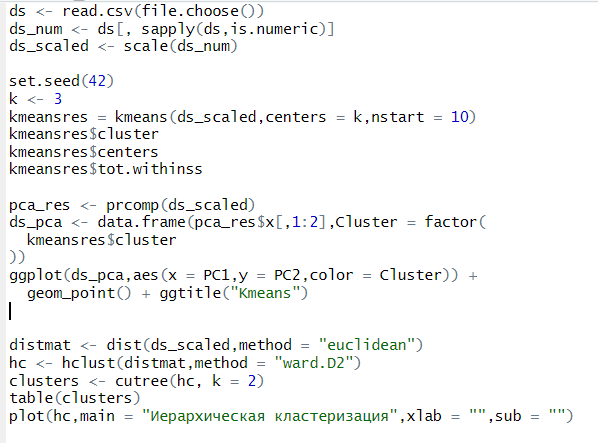


**Шаг 2) Кластеризация в R.**

**2.1) Иерархическая и K-means кластеризация в R.**

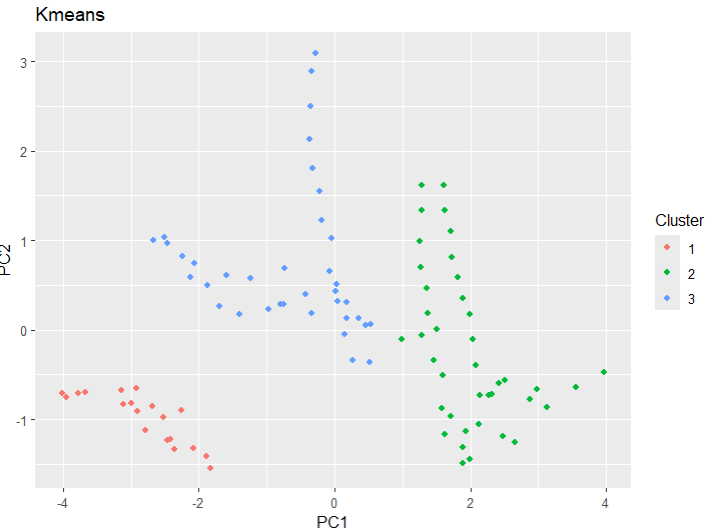
Далее проделаем все тоже самое, но уже с помощью языка программирования R, убедившись, что все необходимые пакеты были успешно установлены.

*Рисунок 2.1 кластеризация в R.*



На этом скриншоте была произведена сначала операция масштабирования, потом написан код для K-means кластеризации, а потом для иерархической. Давайте посмотрим, какие графики нам выводит R.

*Рисунок 2.2 График K-means кластеризации в R.*



Далее построим дендограмму:

*Рисунок 2.3 Иерархическая кластеризация в R.*

**3 СРАВНЕНИЕ РЕЗУЛЬТАТОВ**

И там, и там результаты получились одинаковые. Понятное дело, .Работа в R куда удобнее чем в Python, по причине того, что в принципе код куда более компактный, и понятный, единственное, что графики в Python выглядят более красиво, нежели в R.

Давайте подведем небольшие итоги по кластеризации.

PCA и K-means выявили структуру данных, но требуется дополнительный анализ для оптимизации и интерпретации результатов. Чёткие группы точек указывают на наличие естественных кластеров в данных.

Высокие "ветви" (расстояния между кластерами) указывают на хорошую разделимость групп..Если некоторые объекты объединяются на малых высотах — это может сигнализировать о шуме или перекрытии признаков.

Метод подтверждает наличие 3 устойчивых групп, но требует содержательной интерпретации в контексте данных.

**4 ВЫВОДЫ**

В 8 практической работе исследована работа со способами реализации K-means и иерархической кластеризации. Разобрались, в какой среле разработки более удобно работать, а также подвели общие итоги относительно кластеризации.