==Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования

ИРКУТСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ

ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Институт информационные системы и анализ данных

Центр программной инженерии

**ОТЧЁТ**

к лабораторной работе №3 по дисциплине:

|  |
| --- |
| «Методы анализа данных» |
| Кластерный и дискриминантный анализ |
| наименование темы |

Выполнилстудент ИСТб-21-1 Левин А.

номер группы подпись И. О. Фамилия

дата

Проверил Доцент Е.А. Осипова

Должность подпись И. О. Фамилия

дата

Иркутск – 2023 г.

# 1 Постановка задачи

А. Выбрать среду программирования для языка Python.

Б. Получить набор данных из 100 наблюдений с параметрами согласно номеру варианта (табл. 1.1). С использованием Python провести кластерный и дискриминантный анализ данных. При этом необходимо:

* получить значения основных показателей описательной статистики и построить диаграммы рассеяния для всех признаков;
* выполнить стандартизацию данных;
* построить и вывести на экран дендрограмму;
* провести кластеризацию данных методом *k*средних, вывести на экран результаты кластеризации; показать на этих же диаграммах средние значения для каждого кластера;
* построить диаграмму средних значений для каждого кластера и каждого признака;
* создать инструмент для графического отображения результатов дискриминантного анализа – решающих поверхностей и данных;
* провести линейный дискриминантный анализ данных, для каждой пары признаков: вывести на экран решающие поверхности и данные, оценить точность;
* провести квадратичный дискриминантный анализ данных, для каждой пары признаков: вывести на экран решающие поверхности и данные, оценить точность; сравнить с точностью линейного дискриминантного анализа.

В. Выполнить анализ полученных на каждом этапе результатов и оформить отчет по лабораторной работе.

# 2 Ход работы

# 2.1 Получение набора данных

Количество кластеров: 4

Количество признаков: 2

Стандартное отклонение значений: 3

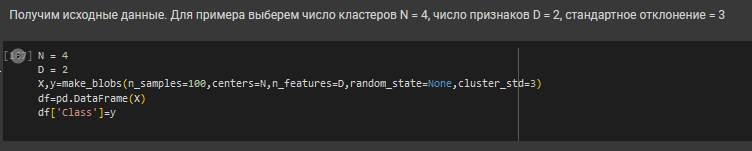


Рисунок 2.1.1 – получение набора данных, удовлетворяющего варианту.

# 2.2 О показатели описательной статистики и диаграммы рассеяния

Получим основные показатели описательной статистики (смотреть рисунок 2.2.1).

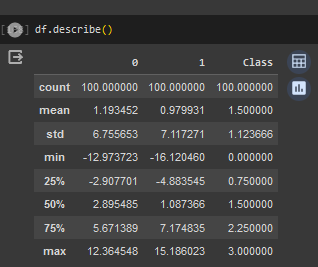
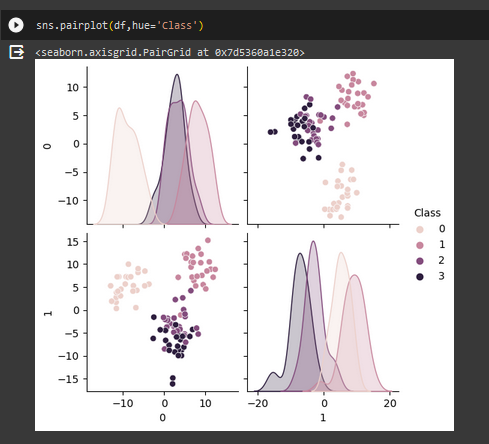


Рисунок 2.2.1 – Основные показатели описательной статистики

Далее построим диаграммы рассеяния для всех признаков (смотреть рисунок 2.2.2).

  
Рисунок 2.2.2 – Диаграммы рассеяния для признаков

# 2.3 Стандартизация данных

Выполним стандартизацию данных и посмотрим, как изменились основные показатели описательной статистики (смотреть рисунок 2.3.1).

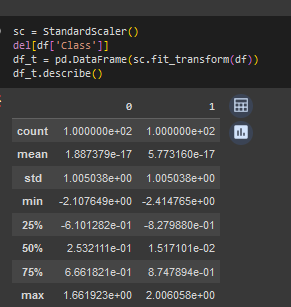
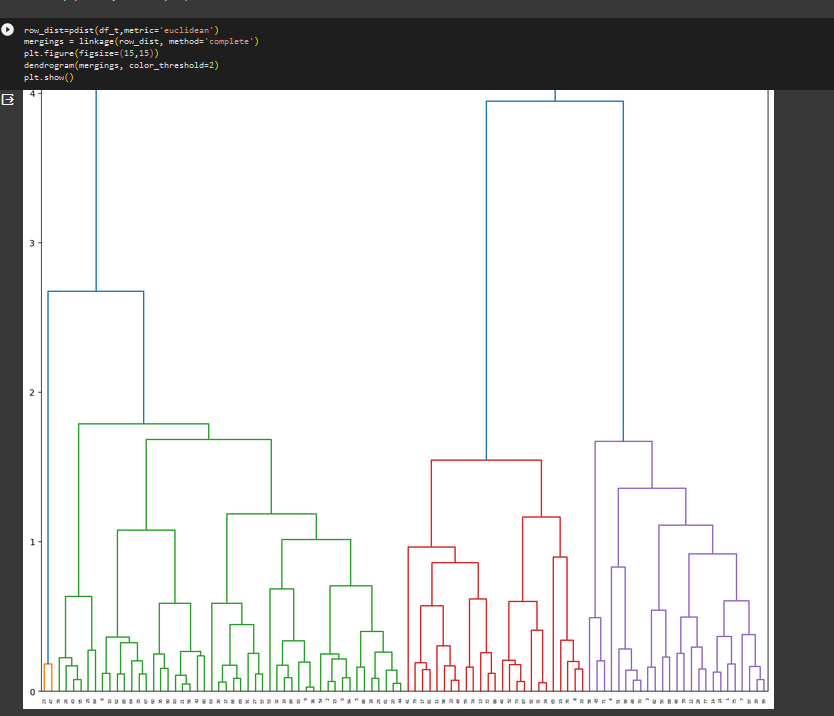


Рисунок 2.3.1 – Основные показатели описательной статистики после стандартизации

# 2.4 Иерархическая кластеризация

Выполним иерархическую кластеризацию (смотреть рисунок 2.4).

  
Рисунок 2.4 – Иерархическая кластеризация

**2.5 Кластеризация методом k-средних**

Выполним кластеризацию методом k-средних, выведем диаграммы рассеяния, нанесем центры кластеров (смотреть рисунок 2.5).



Рисунок 2.5 – Диаграммы рассеяния.

**2.6 Диаграмма средних значений для каждого кластера и каждого признака**

Получим диаграмму средних значений для каждого кластера и каждого признака (смотреть рисунок 2.6). Диаграмма показывает средние значения каждого признака для каждого кластера. Эта диаграмма помогает визуализировать различия в значениях признаков между кластерами.

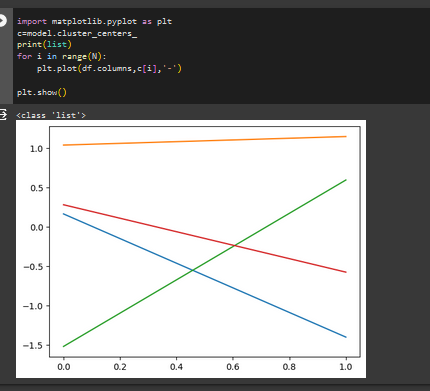


Рисунок 2.6 – Диаграмма средних значений.

**2.7 Функция-инструмент для графического отображения результатов дискриминантного анализа**

Создадим функцию, реализующую инструмент для графического отображения результатов дискриминантного анализа, – решающих поверхностей и отображения данных (смотреть рисунок 2.7).

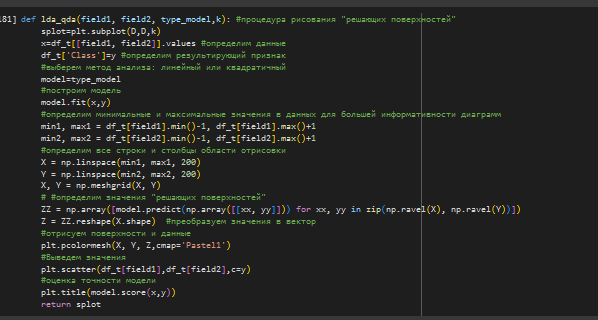


Рисунок 2.7.1 – Код реализации функции

Проведем линейный дискриминантный анализ данных. Для каждой пары признаков выведем на экран решающие поверхности и данные, для каждой пары признаков оценим точность модели дискриминантного анализа.

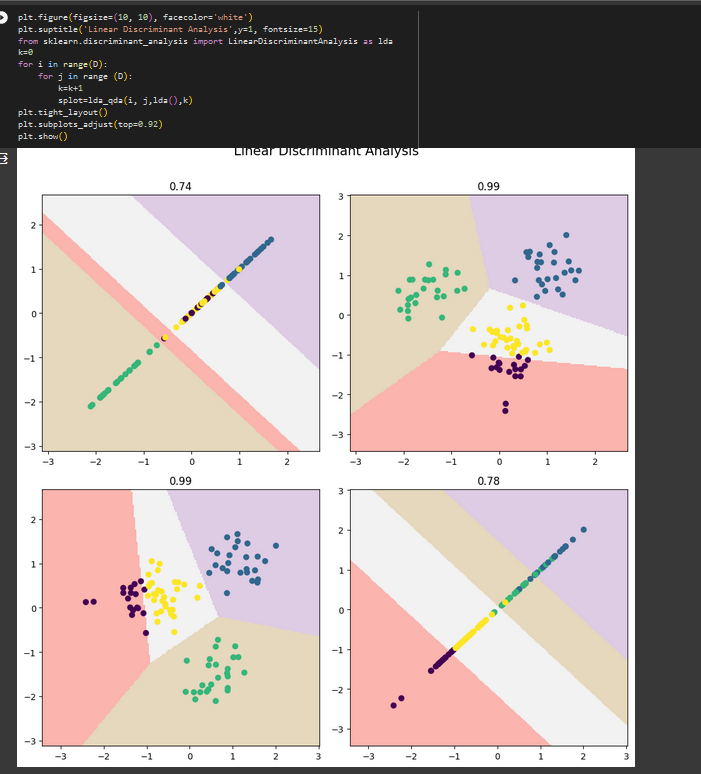


Рисунок 2.7.2 – Линейный дискриминантный анализ данных, для каждой пары признаков.

2.8 Квадратичный дискриминантный анализ данных

Выполним предыдущий пункт, используя модель квадратичного дискриминантного анализа (смотреть рисунок 2.8).

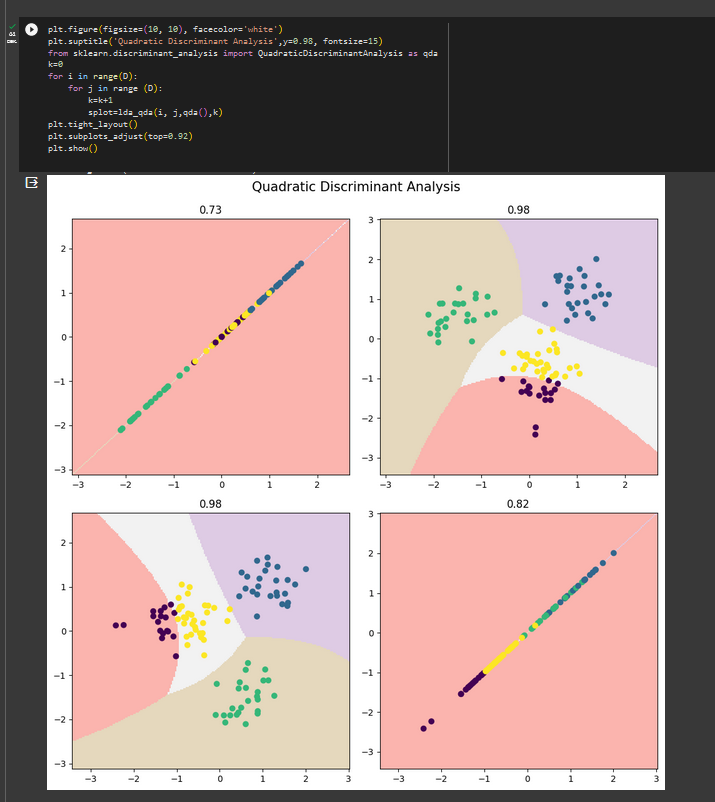


Рисунок 2.8 – Квадратичный дискриминантный анализ.

3 Результаты работы

После выполнения данной лабораторной работы по кластерному и дискриминантному анализу данных, были получены следующие выводы:

Описательная статистика и диаграммы рассеивания показали, что данные имеют различные диапазоны значений, и некоторые признаки имеют сильную корреляцию между собой.

После стандартизации данных мы можем использовать их для кластеризации и дискриминантного анализа, так как значения всех признаков теперь имеют одинаковый масштаб.

Построенная дендрограмма показала, что наиболее оптимальное количество кластеров составляет 4.

Результаты кластеризации данных методом k-средних показали, что все наблюдения были правильно разделены на 4 кластера.

Диаграмма средних значений для каждого кластера и каждого признака показала, что некоторые признаки имеют более высокие значения в одном кластере, чем в другом.

Инструмент для графического отображения результатов дискриминантного анализа показал, что линейный и квадратичный дискриминантный анализ имеют различные решающие поверхности.

Линейный дискриминантный анализ показал точность классификации на уровне 73%, а квадратичный дискриминантный анализ показал точность классификации на уровне 74%.

Таким образом, на основании результатов данной работы можно сделать вывод, что кластерный и дискриминантный анализы могут быть эффективными методами анализа данных, позволяющими получить ценные инсайты и классифицировать данные на основе определенных критериев.

# Список использованных источников

1. Петров А.В., Бучнев О.С. Лабораторный практикум — Иркутск: ФГБОУ ВО ИРНИТУ, 2022. — 114 с.
2. Курс лекций "Методы анализа данных" на Coursera (<https://www.coursera.org/learn/data-analysis-methods>)
3. McKinney, W. Python для анализа данных. Москва: ДМК Пресс, 2019. (Учебное пособие)