ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ

ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО

ОБРАЗОВАНИЯ

«ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт Высоких Технологий и Пьезотехники



Кафедра «Информационных и измерительных технологий»

Направление:  
09.04.03 Прикладная информатика

**Лабораторная работа 3**

Тема: «Дискретная бинарная классификация»

По дисциплине: «Обработка больших данных»

Выполнил: студент 2М курса 6 группы

Аванесов А.А.

Принял:

проф. Щербань И.В.

Ростов-на-Дону   
2019

**Вариант 1**

Исходные данные: 64 образца, 16 features (столбцов, признаков), 17-й столбец - категориальная переменная: "0" = Класс А, "1" = Класс В. В отчете нумерация признаков идет от 0.

1. Анализ данных «вручную» по диаграммам рассеивания.

После построения диаграммы рассеивания наглядно видно три выброса, которые представлены на рисунках 1-4. Данные выбросов представлены на рисунке 5.

|  |  |
| --- | --- |
| C:\Users\Artem Avanesov\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Снимок.jpg | C:\Users\Artem Avanesov\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Снимок2.jpg |
| Рисунок 1. Диаграммы рассеивания с первым выбросом | Рисунок 2. Диаграммы рассеивания со вторым выбросом |

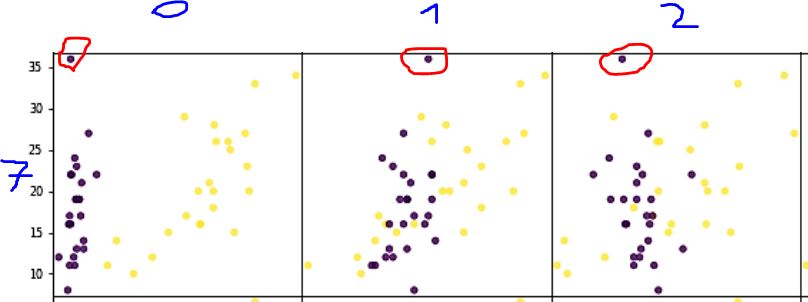


Рисунок 3. Диаграммы рассеивания по 7-му признаку с третьим выбросом

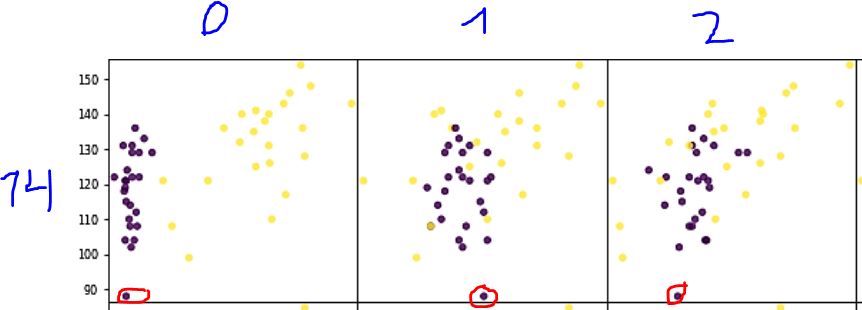


Рисунок 4. Диаграммы рассеивания по 14-му признаку с третьим выбросом

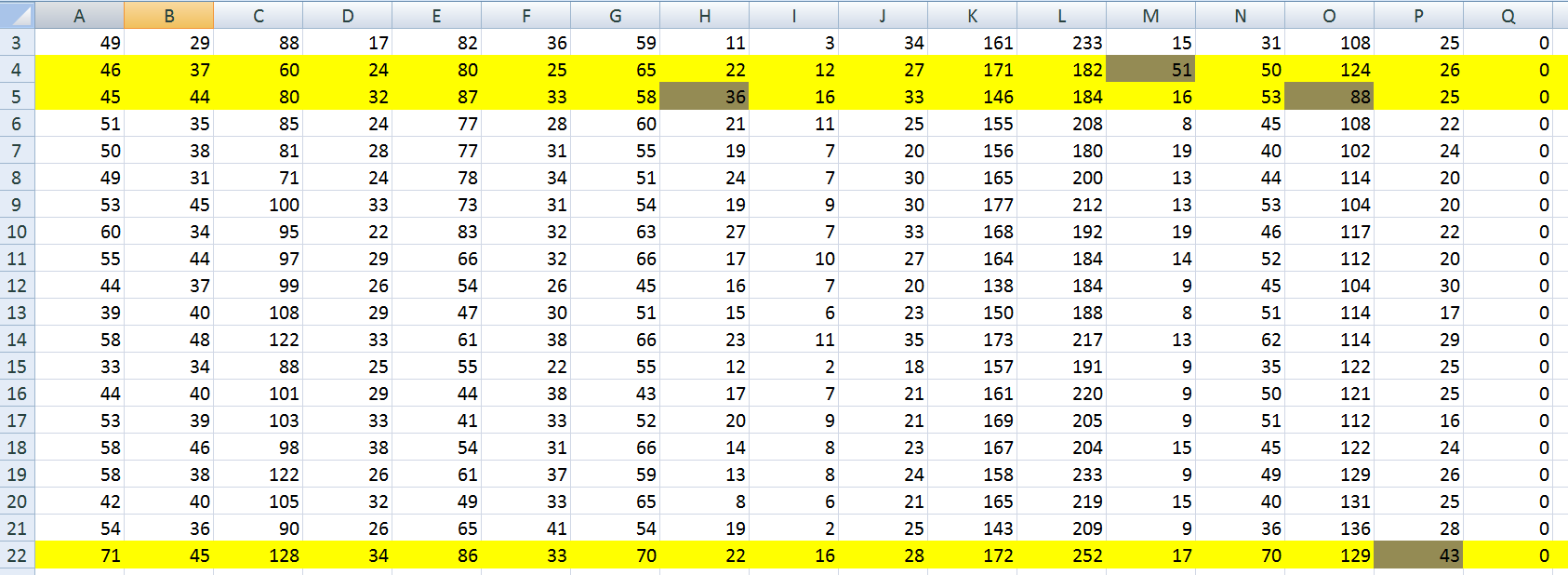


Рисунок 4. Данные выбросов

1. Сделать выводы о возможности разделения классов "вручную".

Проанализировав диаграммы рассеивания можно сделать вывод, что классы можно разделить вручную (рисунок 5). Однако в образовательных целях решим задачу и с помощью методов классификации.

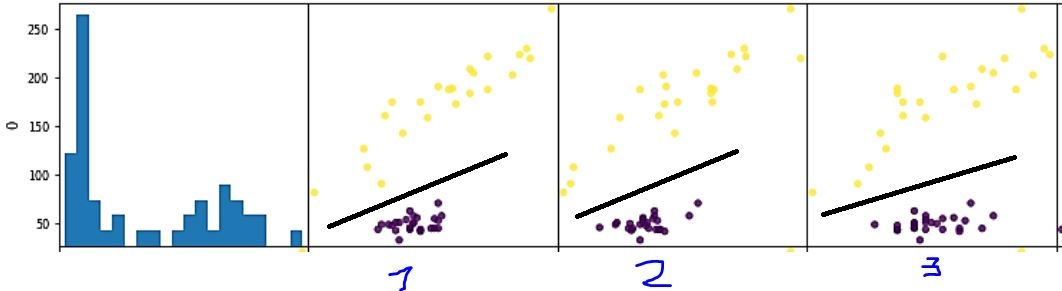


Рисунок 5. Разделение классов «вручную»

1. Провести анализ матриц диаграмм рассеивания по парам признаков.

Проведя анализ диаграмм рассеивания по парам признаков можно сделать вывод, что некоторые признаки не так важны при делении образцов на классы. Диаграммы рассеивания представлены на рисунке 6.

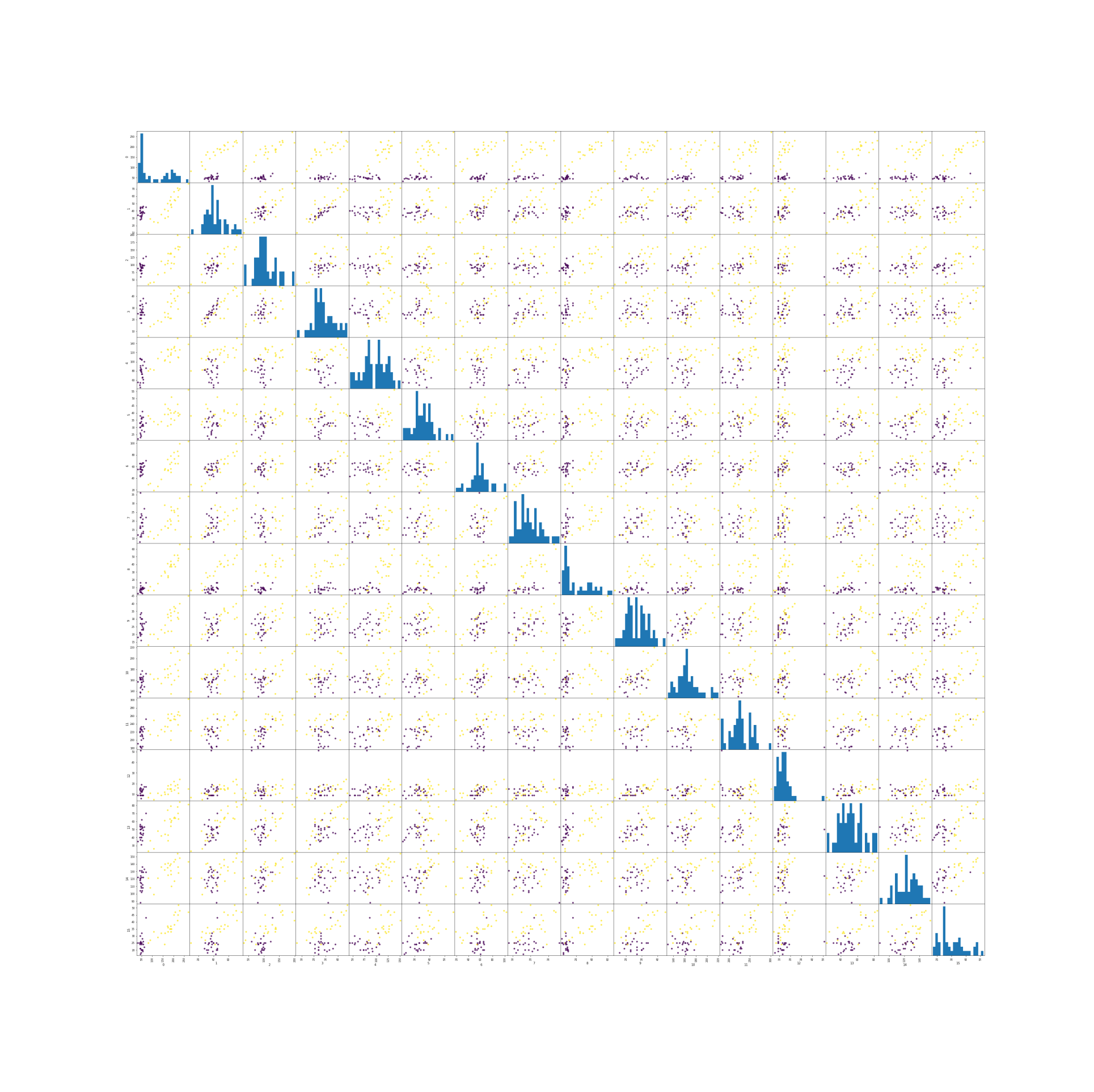


Рисунок 6. Диаграммы рассеивания по всем признакам

1. Выполнить дискретную бинарную классификацию

Полученные результаты обучения для логистической регрессии с использованием L1-регуляризации представлены на рисунке 7.

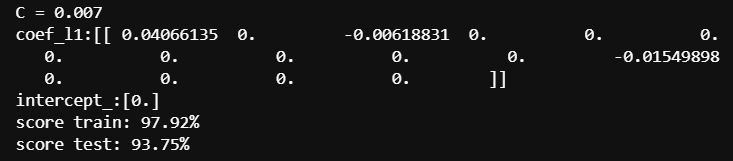


Рисунок 7. Результаты логистической регрессии с L1

Полученные результаты обучения для логистической регрессии с использованием L2-регуляризации представлены на рисунке 8.

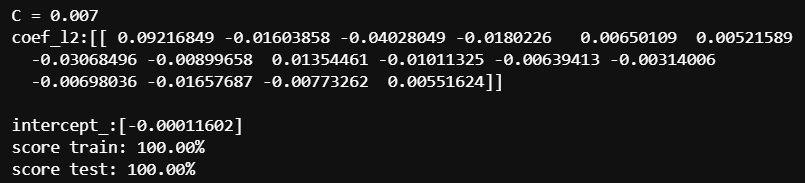


Рисунок 8. Результаты логистической регрессии с L2

На рисунке 9 представлены полученные весовые коэффициенты признаков.

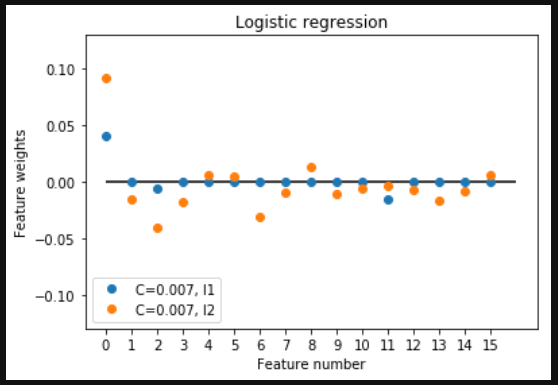


Рисунок 9. Весовые коэффициенты

1. Использовать все возможные классификаторы
2. Сделать выводы о качестве выборки
3. Сделать выводы о достоинствах и недостатках классификаторов
4. Выполнить анализ значимости признаков
5. Выполнить классификацию на наиболее дискриминативных признаках