Санкт-Петербургский политехнический университет Петра Великого

Институт Компьютерных

наук и кибербезопасности

**Разработка графических приложений**

Отчет по Лабораторной работе 1 на тему:

«**Классификация данных на плоскости**»

Направление 09.04.01\_15 «Технологии проектирования системного и прикладного программного обеспечения»

Студент группы 5140901/31502 \_\_\_\_\_\_\_\_\_\_\_\_\_\_Алешковский А.А.

Преподаватель \_\_\_\_\_\_\_\_\_\_\_\_\_\_ Никитин К.В.

Санкт-Петербург

2023

1. **Исследование ML-классификатора**
   1. **Задание параметров**

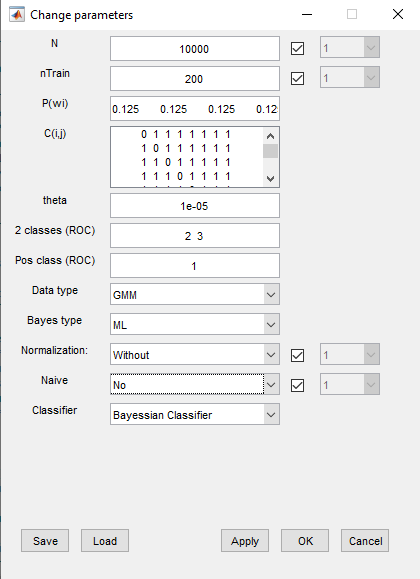
****

Рисунок 1.1 – Исходные параметры

* 1. **Построение графиков**

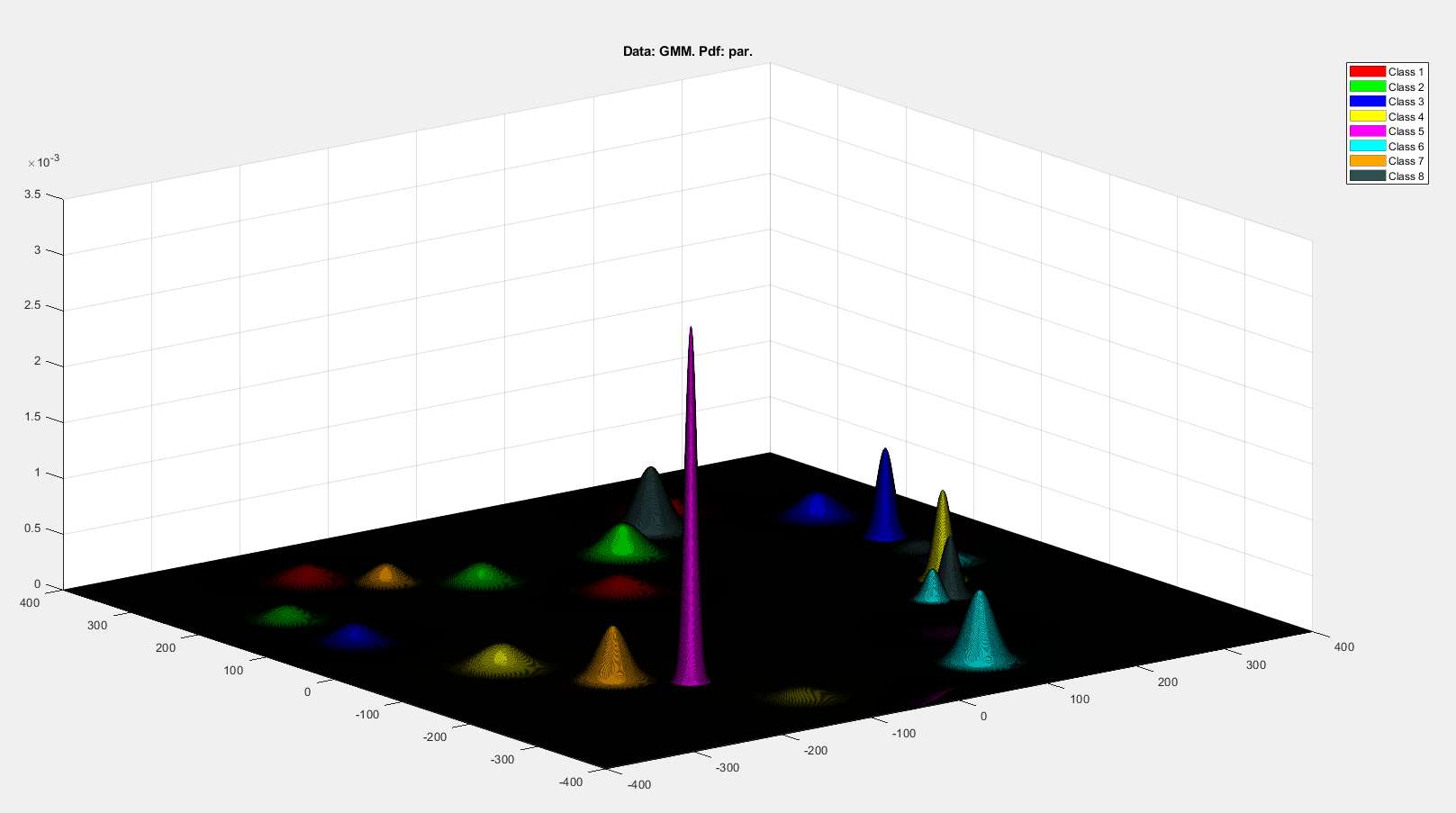


Рисунок 1.2.1 – Истинная плотность

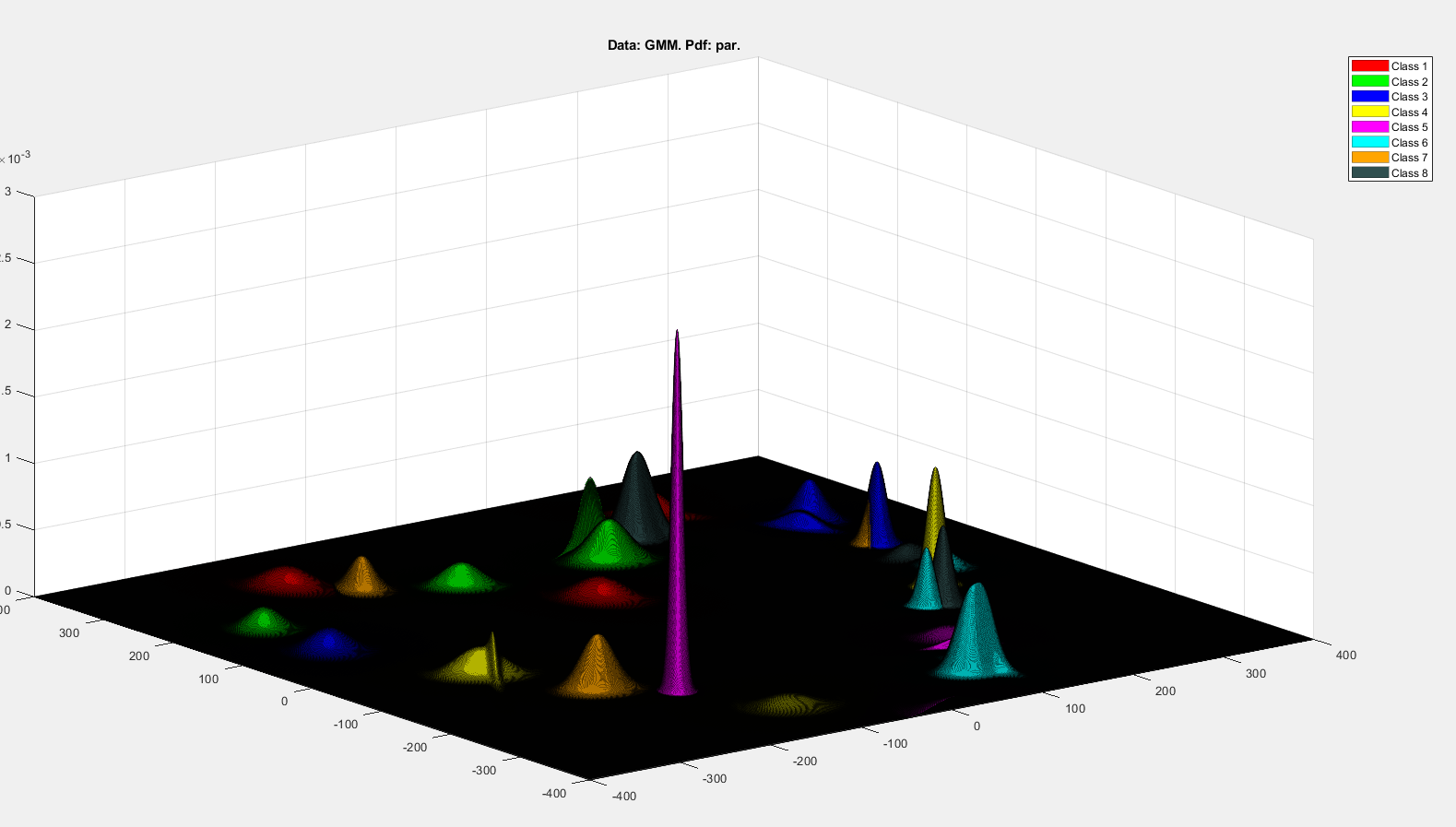


Рисунок 1.2.2 – Оцениваемые плотности

* 1. **Тестовые примеры**

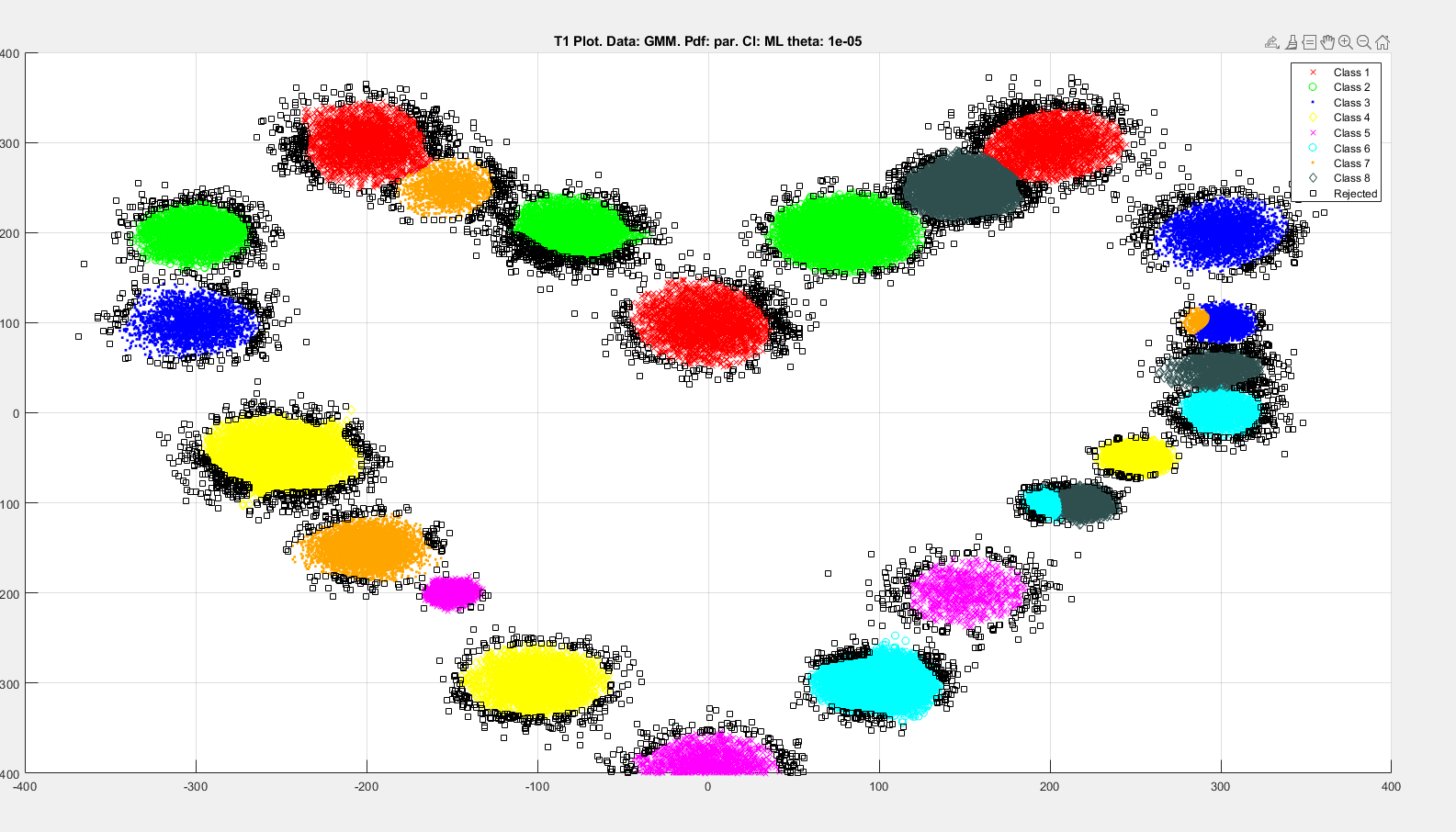
****

Рисунок 1.3.1 – Множество тестовых примеров

* 1. **Матрица неточности**

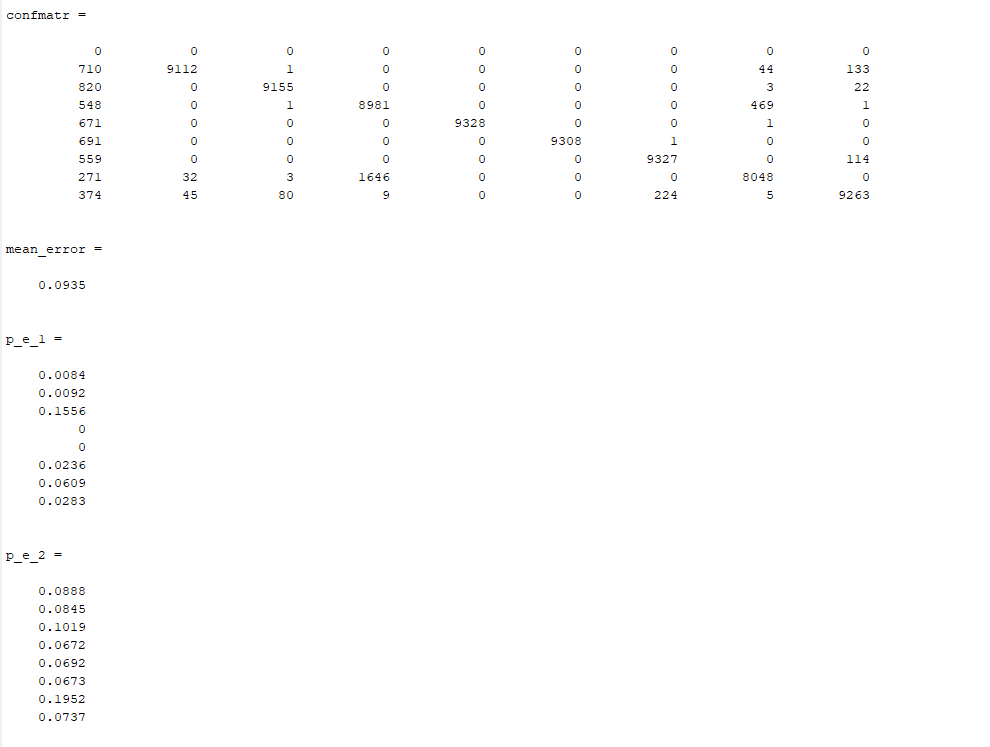
****

Рисунок 1.4.1 – Матрица неточности

* 1. **Множество примеров в прямоугольной области**

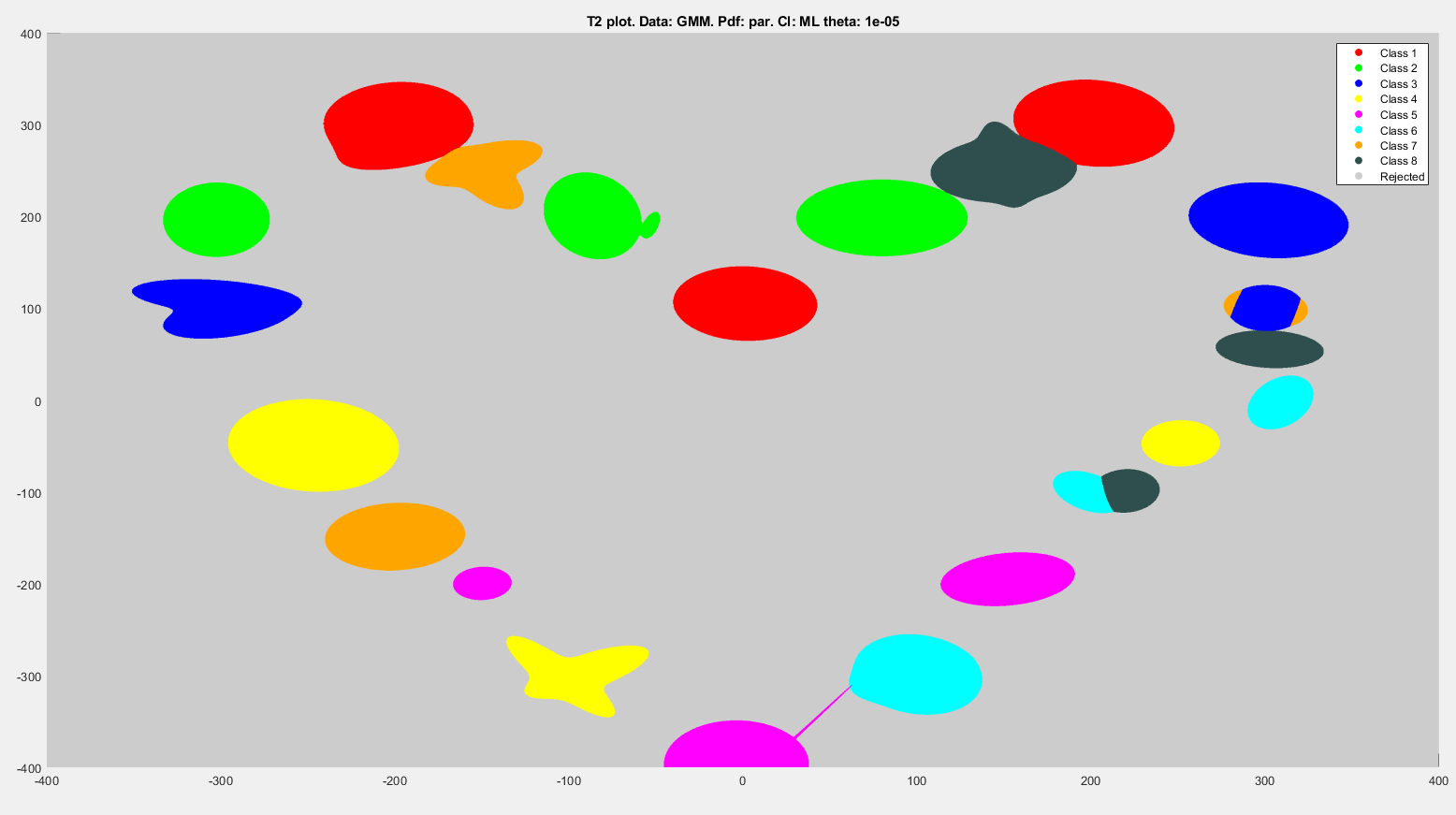
****

Рисунок 1.5.1 – График примеров

* 1. **Исследования влияния порога срабатывания**

**1.7.2 Выбор нескольких значений**

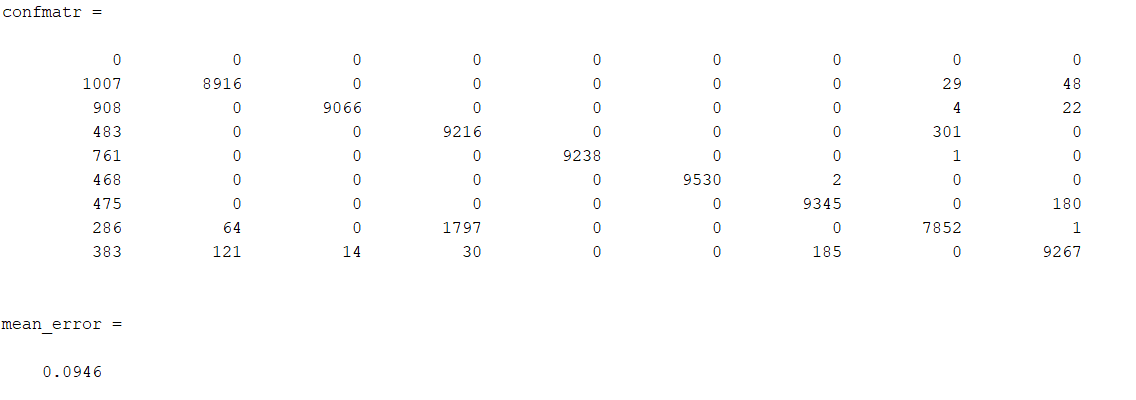
****

Рисунок 1.7.2 – Матрица неточности при оптимальном пороге

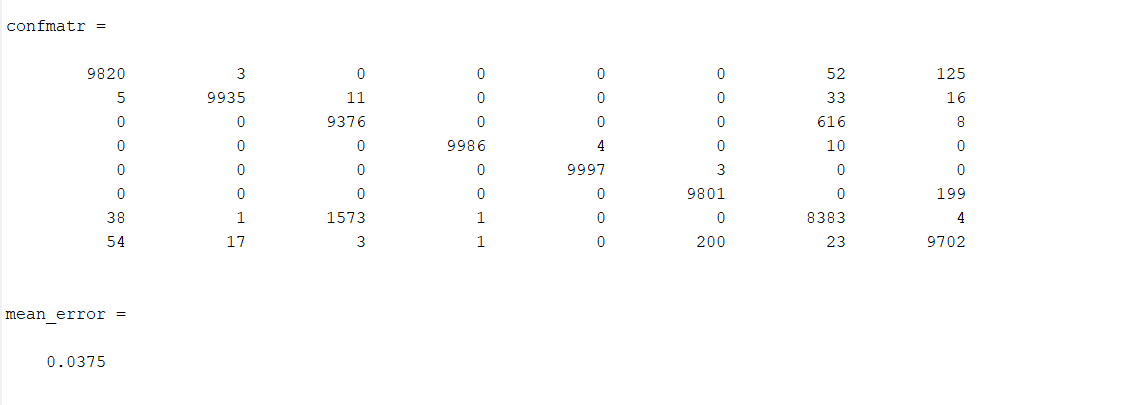


Рисунок 1.7.3 – Матрица неточности при нулевом пороге

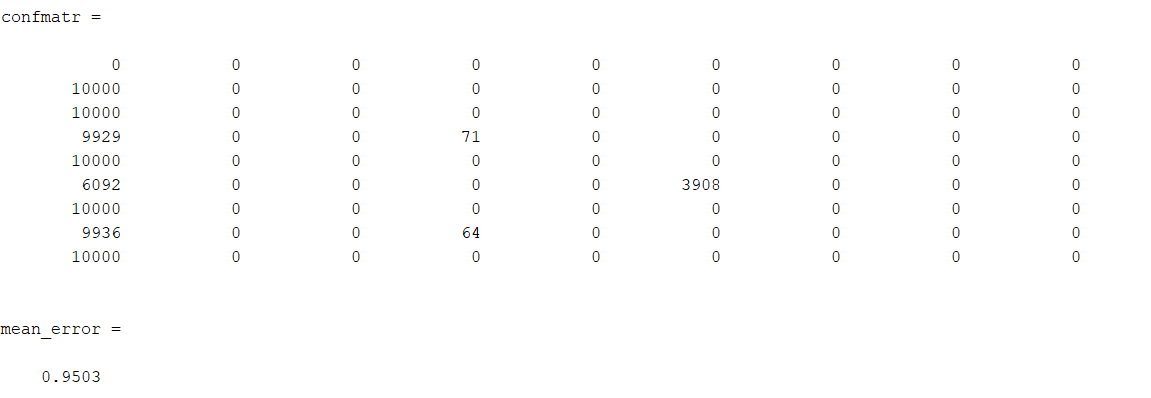
****

Рисунок 1.7.4 – Матрица неточности при излишне большом пороге

Можно сделать вывод о том, что при оптимальном значении порога ошибка является наименьшей.

* 1. **Построение ROC-кривой**

**1.8.2 Результаты выполнения программы**

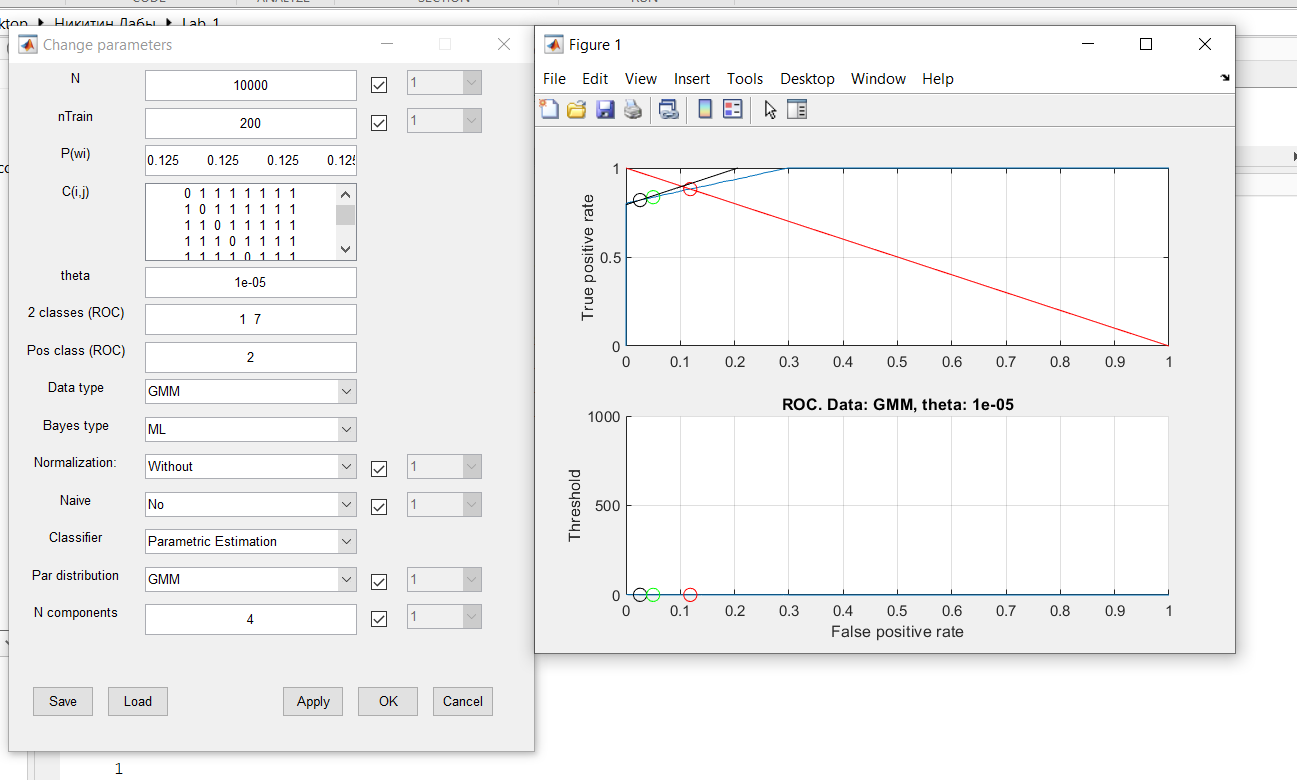
****

Рисунок 1.7.4 – Матрица неточности при излишне большом пороге

Рисунок 1.8.2 – Результат выполнения программы

1. **MAP-классификатор**

В результате проделанных изменений в программе, ошибка стала больше.

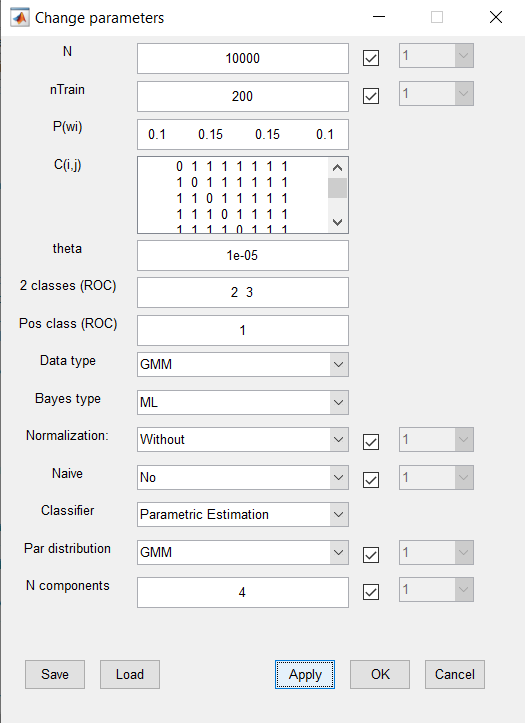


Рисунок 2.1 – задание различных значений априорных распределений

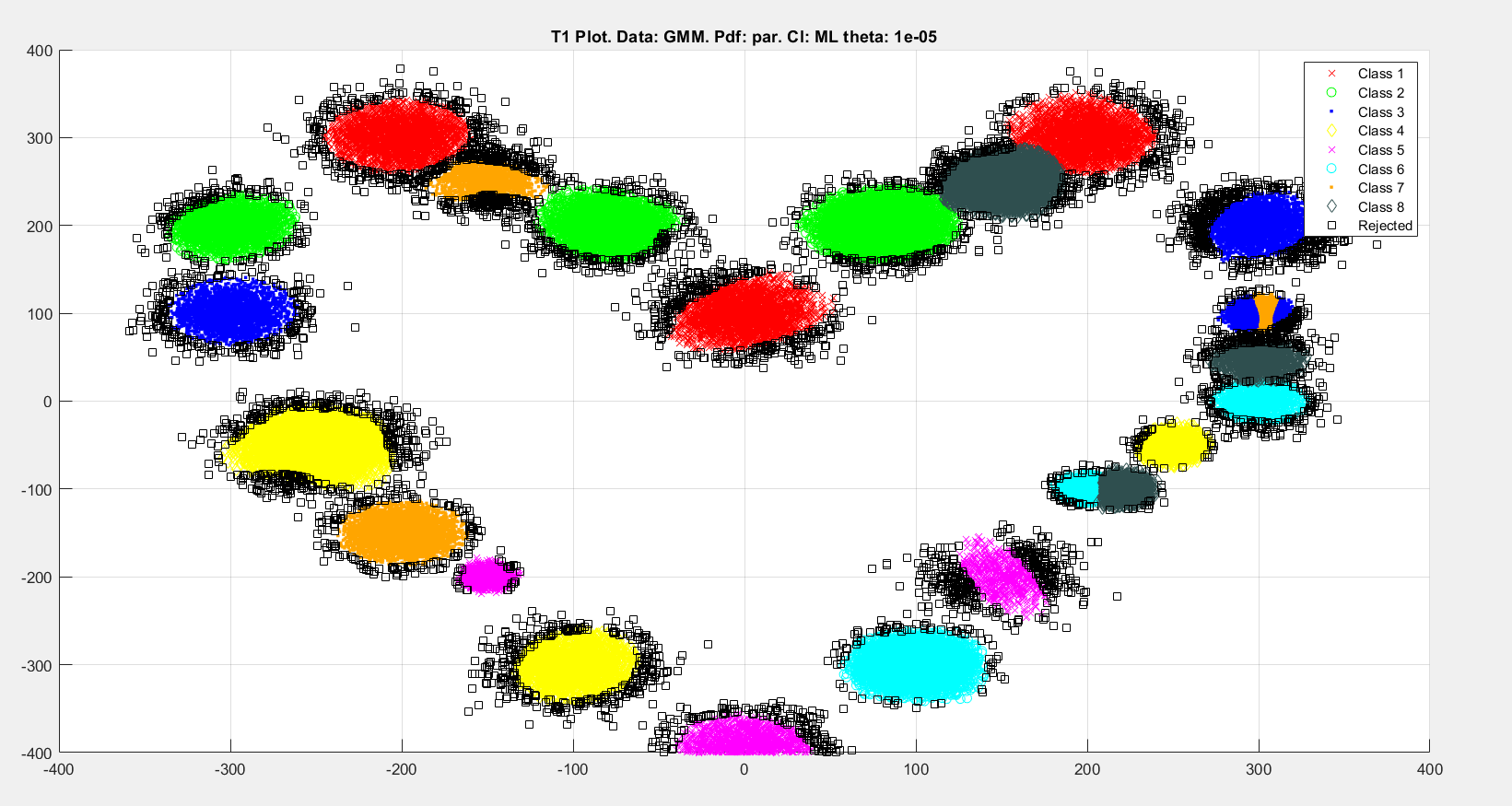


Рисунок 2.2 – график T1

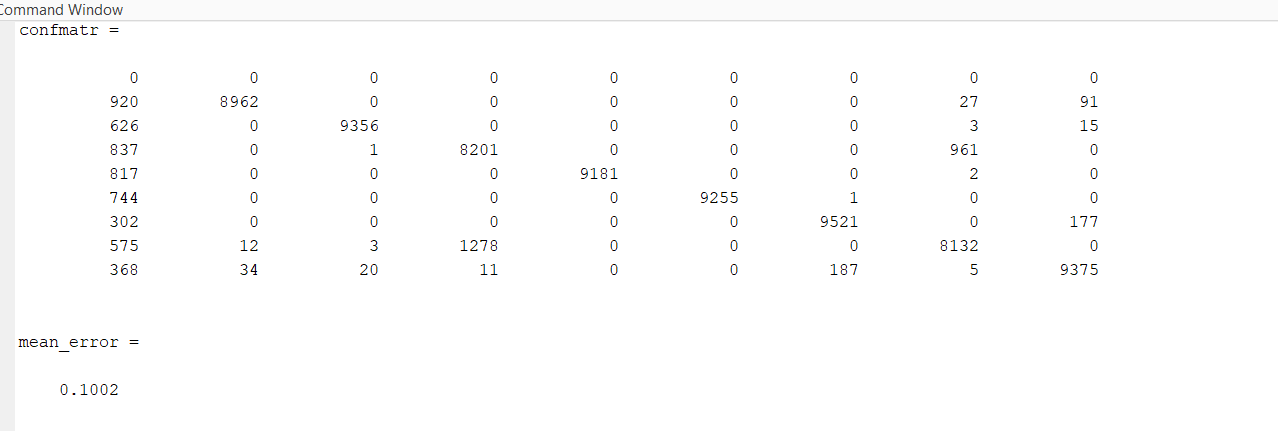


Рисунок 2.3 – результаты работы программы

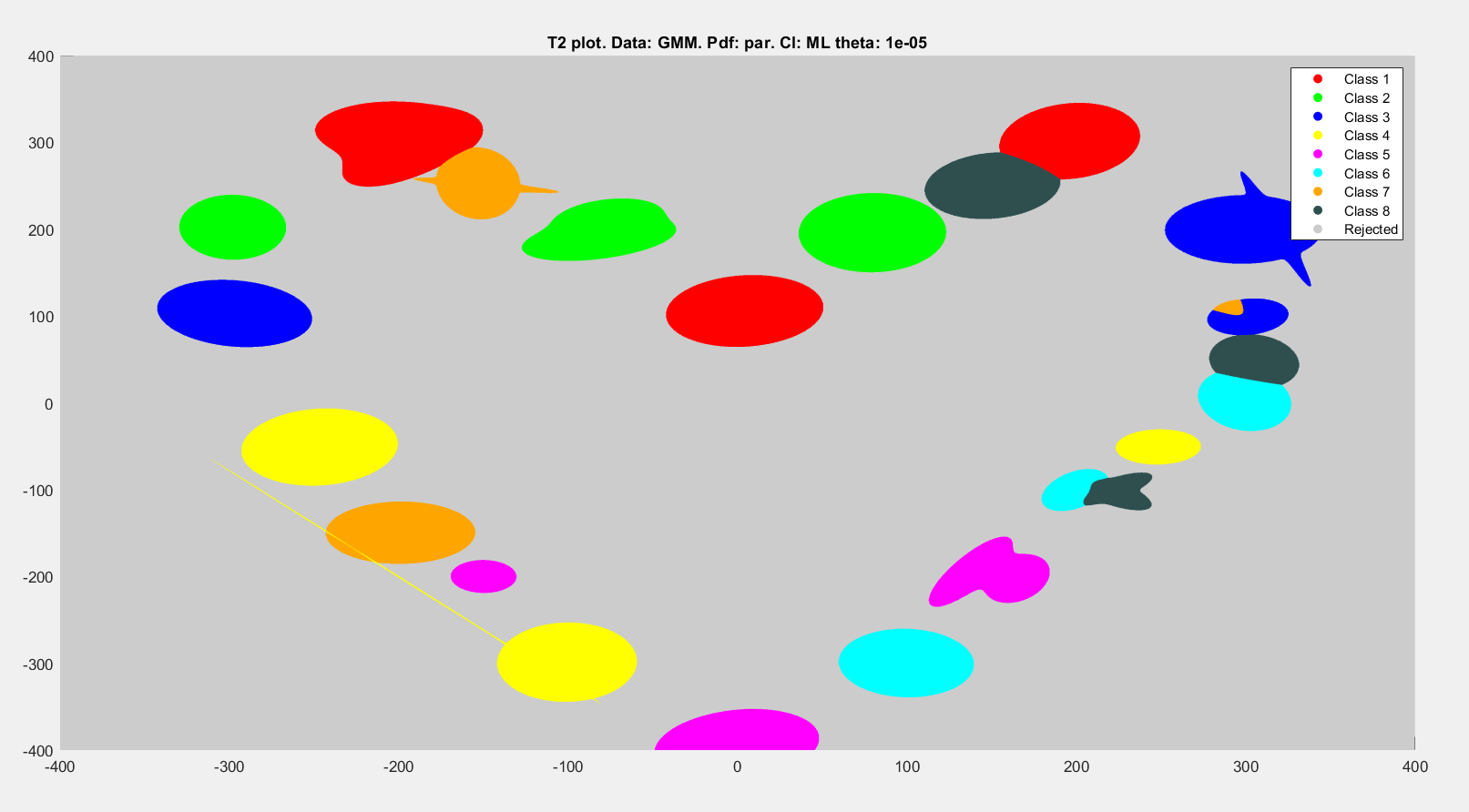


Рисунок 2.4 – T2

1. **Исследование обобщенного Байесовского классификатора**

Исходя из предоставленной матрицы ошибок, видно, что определенные ошибки классификации происходят чаще других. Для определения значений в матрице стоимости ошибок мы можем проанализировать, где ошибки встречаются наиболее часто. Вот какие ошибки чаще всего происходят:

Класс 1 часто ошибочно определяется как класс 5.

Класс 2 часто ошибочно относится к классам 1 и 5.

Класс 3 часто ошибочно относится к классам 1, 2 и 6.

Класс 5 часто ошибочно классифицируется как класс 2.

Класс 6 часто ошибочно классифицируется как класс 4.

Класс 7 часто ошибочно классифицируется как классы 5 и 9.

Класс 8 часто ошибочно классифицируется как класс 6.

Для выделения этих частых ошибок можно увеличить стоимость соответствующих ячеек в матрице стоимости ошибок. Для иллюстрации, мы можем установить увеличенные значения в 10.

0 10 1 10 1 10 1 1 1

10 0 10 1 10 1 1 1 1

10 10 0 1 1 10 1 1 1

1 1 1 0 1 1 1 10 1

1 10 1 1 0 1 10 10 10

10 1 1 10 1 0 1 1 1

1 1 1 1 10 10 0 10 1

1 1 1 10 10 1 10 0 10

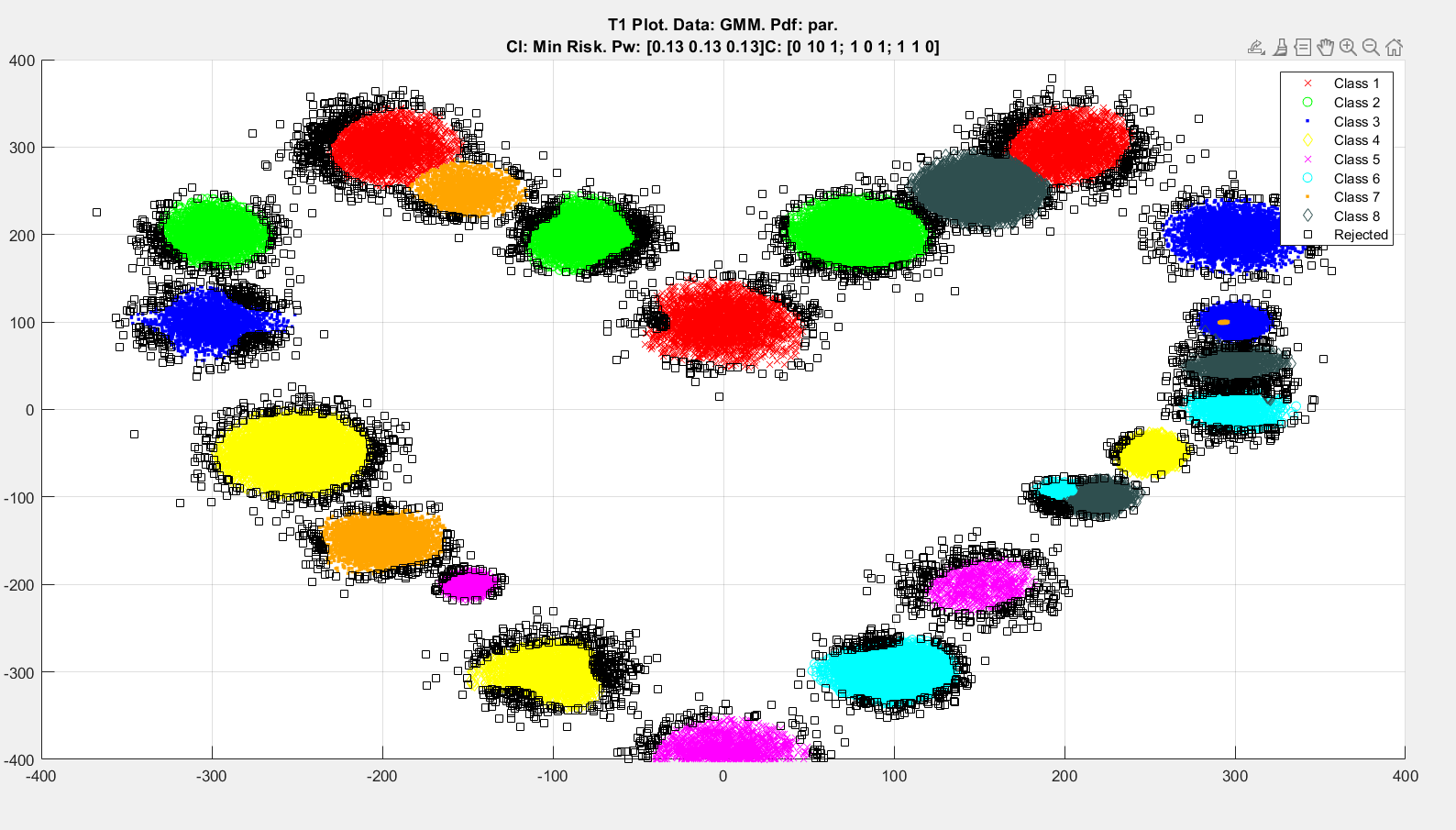


Рисунок 3.1 – T1

Ошибка возросла.

confmatr =

0 0 0 0 0 0 0 0 0

967 8838 0 0 0 0 0 60 135

841 0 9147 0 0 0 0 3 9

521 0 0 9428 0 0 0 49 2

790 0 0 0 9207 0 0 3 0

730 0 0 0 0 9270 0 0 0

583 0 0 0 2 0 9221 0 194

309 32 0 1922 0 0 0 7736 1

294 23 5 19 0 0 137 0 9522

mean\_error =

0.0954

p\_e\_1 =

0.0062

0.0005

0.1707

0.0002

0

0.0146

0.0146

0.0346

p\_e\_2 =

0.1162

0.0853

0.0572

0.0793

0.0730

0.0779

0.2264

0.0478

4.1 Различия между различными методами классификации:

MAP (Максимальное апостериорное правдоподобие) учитывает и данные, и априорное распределение параметров.

ML (Максимальное правдоподобие) максимизирует вероятность наблюдаемых данных, не учитывая априорное распределение.

Обобщенный Байесов классификатор позволяет более гибко настроить классификатор с учетом матрицы стоимости ошибок и других параметров.

Критерий Неймана-Пирсона фокусируется на максимизации мощности теста при заданном уровне значимости.

Минимакс стремится минимизировать максимально возможный риск или потери.

Когда использовать каждый метод:

ML проще в реализации, но может не учитывать важные аспекты данных и предполагает равнозначность всех классов.

MAP полезен, когда доступна информация об априорном распределении.

Обобщенный Байесов классификатор полезен, если существуют разные стоимости ошибок для различных типов ошибок.

Критерий Неймана-Пирсона и минимакс применяются в специфических сценариях, где важно контролировать ошибки первого и второго рода или минимизировать максимальный риск.

4.2 Установка порога для отклонения гипотез:

Порог отклонения гипотезы необходим, когда нужно контролировать частоту ошибок первого или второго рода или установить определенный уровень чувствительности или специфичности.

Сложности в установке порога:

Определение оптимального порога сложно, так как требуется балансировка между разными типами ошибок.

Порог может зависеть от контекста применения и может потребовать дополнительного анализа и тестирования.

4.3 Ненулевые значения на диагонали матрицы стоимости:

Ненулевые значения на диагонали могут быть полезны, когда есть стоимость, связанная с верным классифицированием объектов, например, стоимость проверки или валидации результатов.

4.4 Возможности ROC-кривой:

ROC-кривая отображает зависимость между чувствительностью и специфичностью для различных пороговых значений.

ROC-кривая позволяет анализировать производительность классификатора независимо от порога.

Площадь под ROC-кривой (AUC) предоставляет количественную оценку производительности классификатора.