Федеральное государственное автономное

образовательное учреждение

высшего образования

«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт космических и информационных технологий

Кафедра «Информатика»

**РАСЧЕТНО-ГРАФИЧЕСКОЕ ЗАДАНИЕ №1**

Классификация в распознавании образов

Преподаватель \_\_\_\_\_\_\_\_\_ А. С. Михалев

Студент КИ16-17б, №031620650 \_\_\_\_\_\_\_\_\_ Е. Г. Щепин

Красноярск 2019

**1 Введение**

Цель настоящей работы заключается в исследовании решающего правила классификации для случая двух независимых информативных признаков, двух классов при заданных условных плотностях вероятности.

Для ее достижения реализован классификатор, проведено исследование его свойств от параметров путем перебора данных в некотором диапазоне с заданным шагом и накоплением статистической информации.

Основная информация, содержащаяся в отчете по работе, содержит: описание хода выполнения работы, примеры работы программы с включенной в него таблицей с результатами запуска классификатора для различных входных параметров, исследование зависимости свойств классификатора от его параметров.

# 2 Постановка задачи

Имеется два независимых непрерывных информативных признака , два класса . Известны априорные вероятности классов , условные плотности вероятности (при условии истинности того или иного класса) для информативных признаков с точностью до параметров :

(1)

Одномерные условные плотности вероятности – законы нормального распределения с параметрами :

(2)

Известна обучающая выборка объема :

* когда истинным является первый класс;
* когда истинным является второй класс.

Необходимо разработать графическое приложение с использованием любого языка программирования, позволяющее по заданным параметрам условных плотностей вероятностей, априорным вероятностей классов на основе обучающей выборки установленного объема, полученной посредством применения центральной предельной теоремы (ЦПТ) (число слагаемых определяется пользователем) или метода полярных координат, построить байесовское решающее правило. В качестве выходной информации должны выступать: значения оценок дисперсий, математических ожиданий, априорных вероятностей классов, вероятностей ошибок классификации, количество элементов выборки, относящихся к первому, второму классу.

Также приложение должно предусматривать возможность построения таких графиков, как:

* зависимость вычисленной оценки ошибки классификации от объема выборки;
* зависимость от ;
* зависимость от при одинаковых математических ожиданиях;
* зависимость от при разных математических ожиданиях;
* зависимость от априорной вероятности первого класса при одинаковых математических ожиданиях;
* зависимость от при разных математических ожиданиях.

**3 Ход выполнения работы**

Реализация классификатора выполнена с помощью языка программирования R, библиотеки, реализующей графический интерфейс, “shiny” и среды разработки R Studio.

**3.1 Разработка логики**

Первоначально созданы 4 функции – centralLimTheor, polarCoordinatMeth, generatePoints, sampleGeneration. Их назначение – генерация обучающей выборки. Далее разработаны 3 основные функции – func, decision, classification. Данные функции служат для построения Баесовского классификатора.

**3.2 Разработка пользовательского интерфейса**

Написан метод ui – представляющий собой графический интерфейс, используя интегрированные средства js и html. Данный метод напрямую связан с функцией server, который описывает логику обработки информации при взаимодействии пользователя с интерфейсом (рисунок 1).

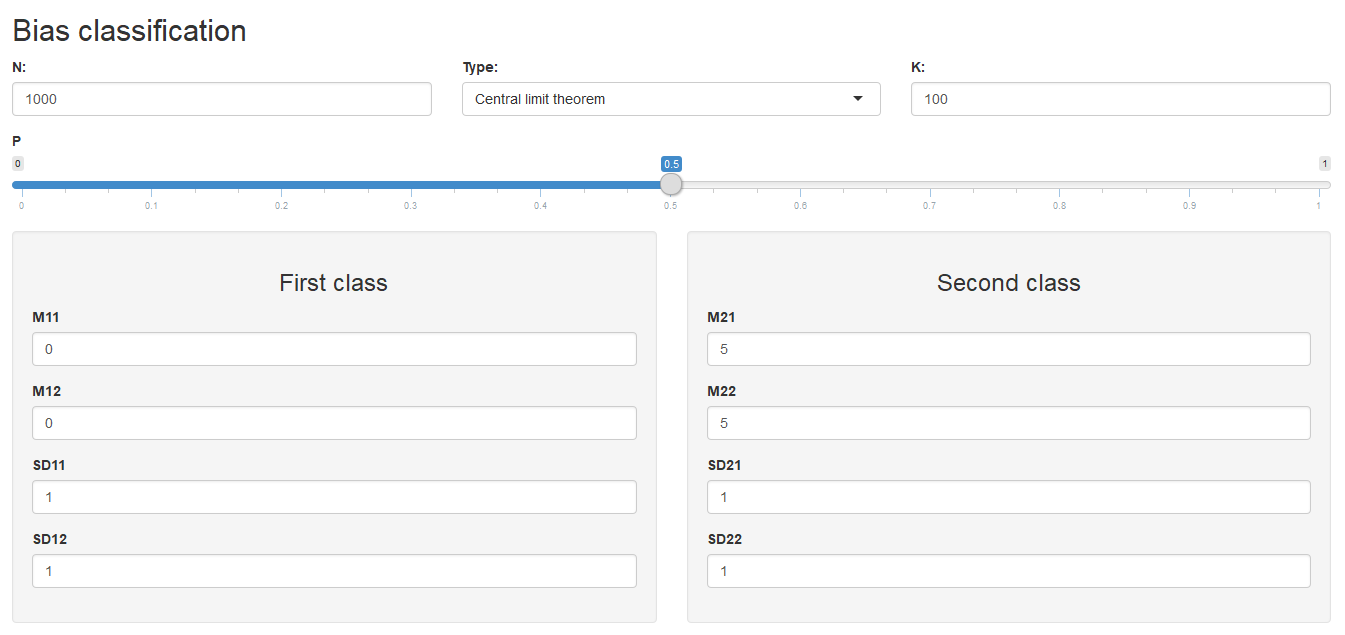


Рисунок 1 – Пользовательский интерфейс

**4 Примеры работы программы**

Произведена серия запусков программы. Было запущено 3 запуска с различными входными параметрами. Результаты работы представлены на рисунках 2-7 при m11 = 0, m12 = 0, m21 = 3, m22 = 3, m11 = 1, m21 1, m21 = 2, m22 = 2, m11 = -5, m12 = 0, m21 = 0, m22 = 0 соответственно. Стандартное отклонение для всех запусков одинаково sd11 = 1, sd12 = 1, sd21 = 1, sd22 = 1.

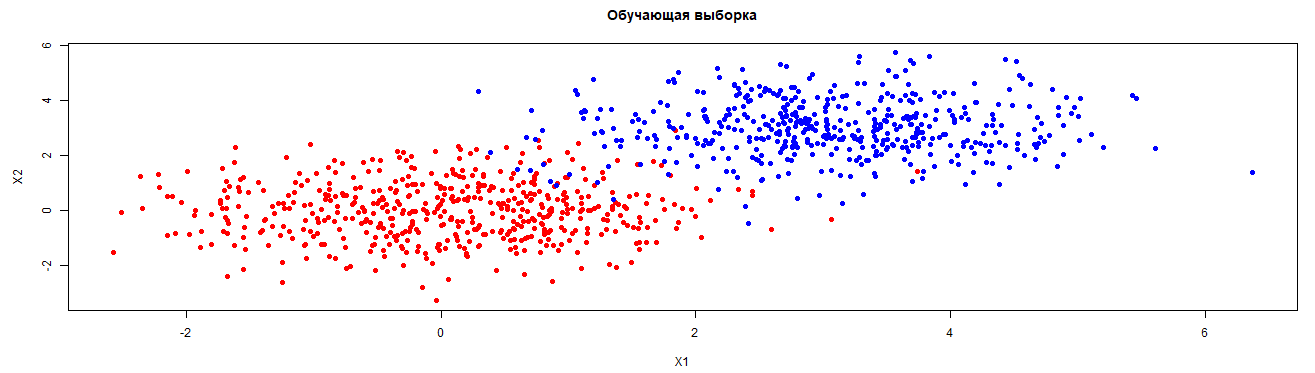


Рисунок 2 – Обучающая выборка 1

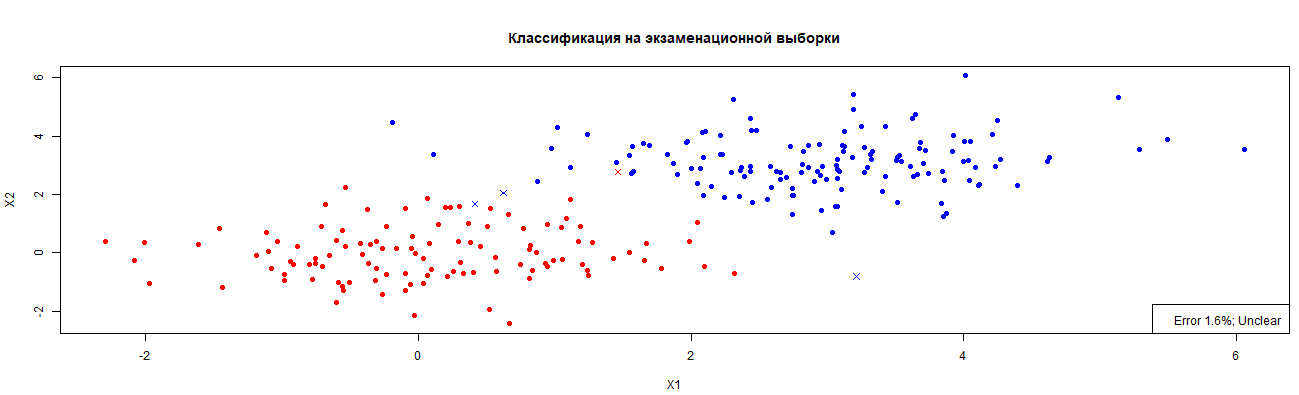


Рисунок 3 – Классификация на экзаменационной выборке 1

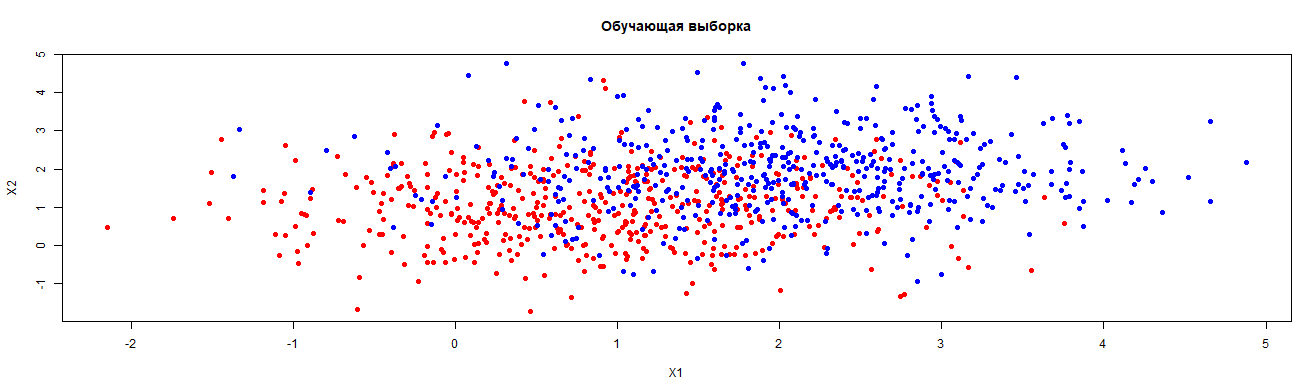


Рисунок 4 – Обучающая выборка 2

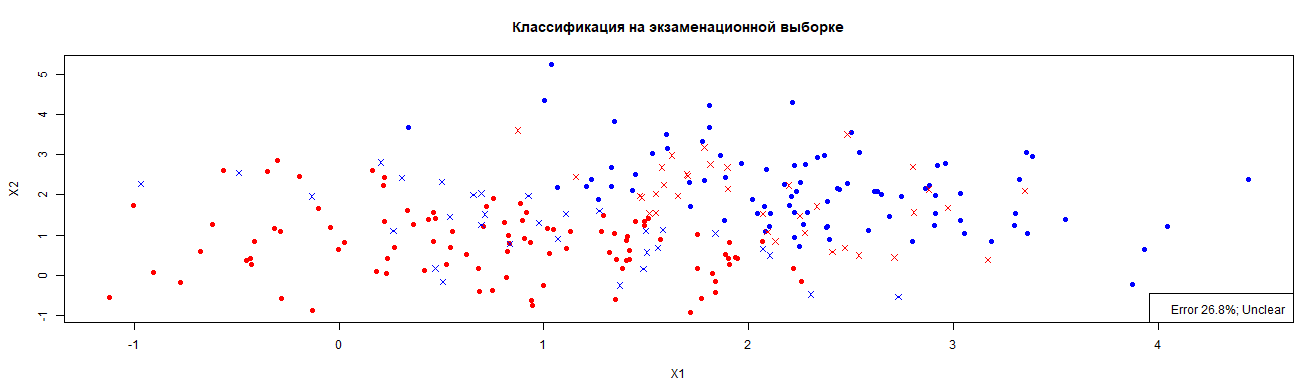


Рисунок 5 – Классификация на экзаменационной выборке 2

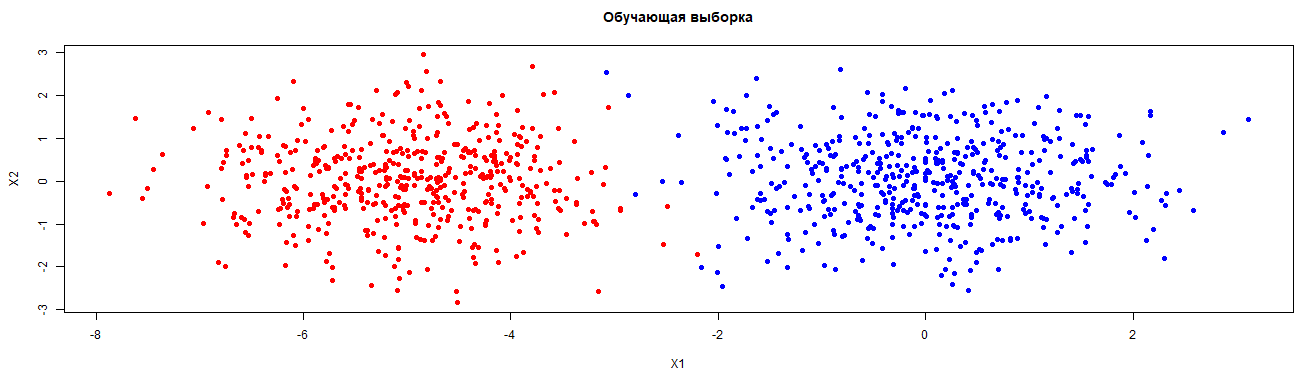


Рисунок 6 – Обучающая выборка 3

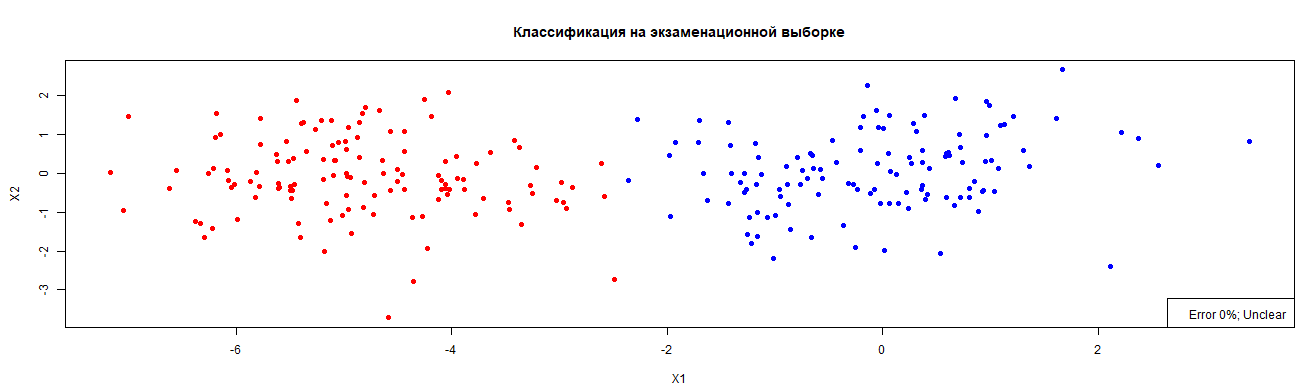


Рисунок 7 – Классификация на экзаменационной выборке 3

На рисунке 8 представлены оценочные значения математических ожиданий и дисперсий на 3 примере.

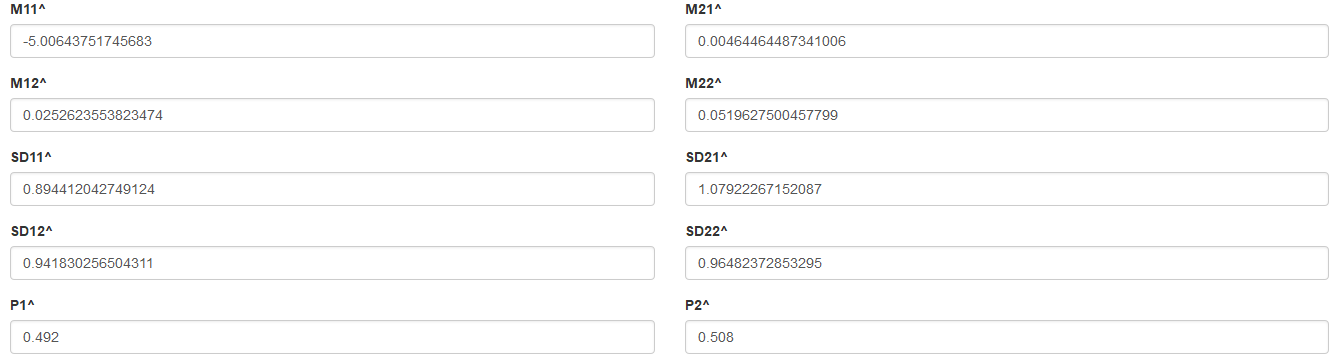


Рисунок 8 – Оценочные значения

**5 Исследование свойств классификатора**

Выполнено исследование зависимости вероятности ошибки классификации от различных входных параметров классификатора (рисунки 9-14).

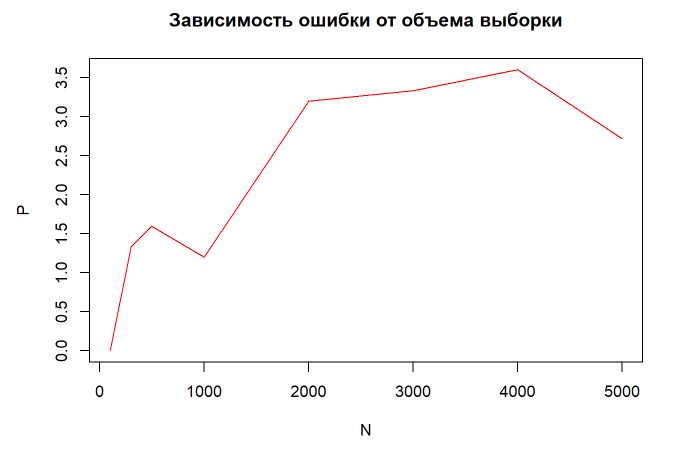


Рисунок 9 – Зависимость ошибки от объёма выборки

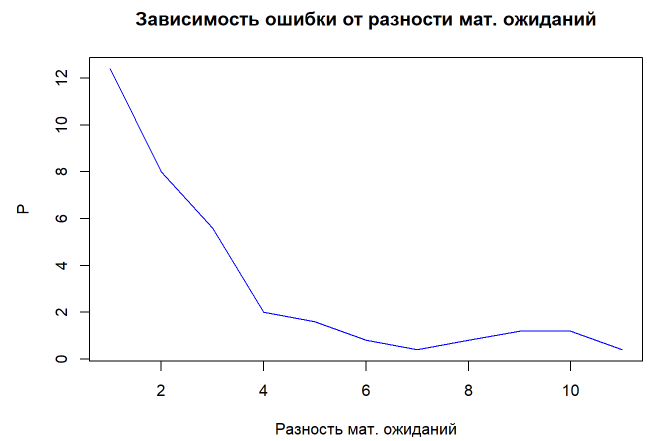


Рисунок 10 – Зависимость ошибки от разности мат.ожиданий

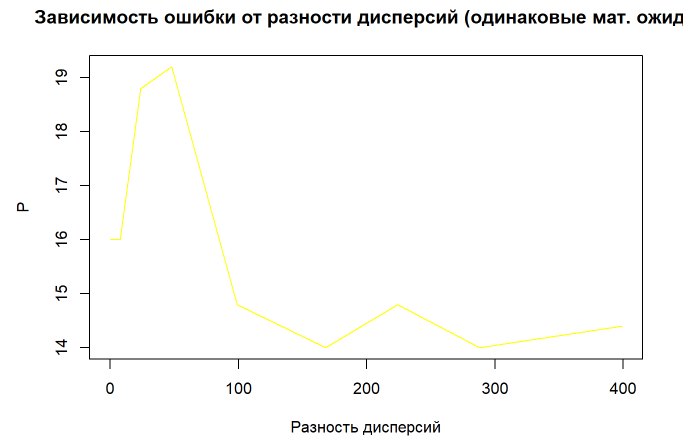


Рисунок 11 – Зависимость ошибки от разности дисперсий (один. мат. ожид)

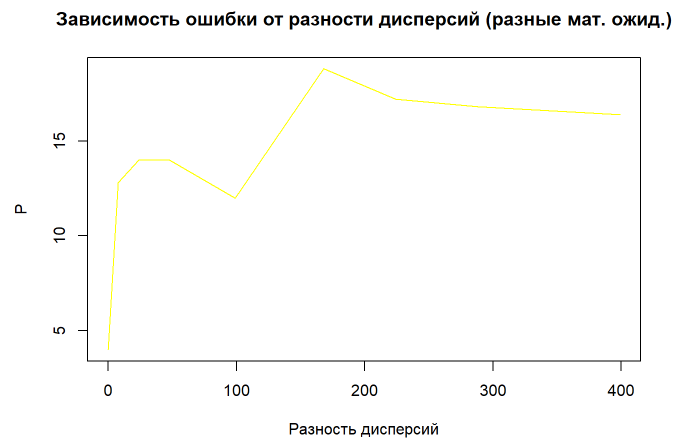


Рисунок 12 – Зависимость ошибки от разности дисперсий (разн. мат. ожид)

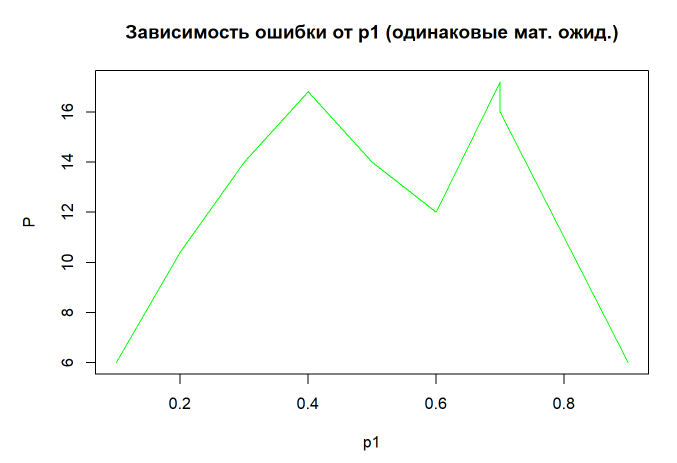


Рисунок 13 – Зависимость ошибки от p1 (один. мат. ожид)

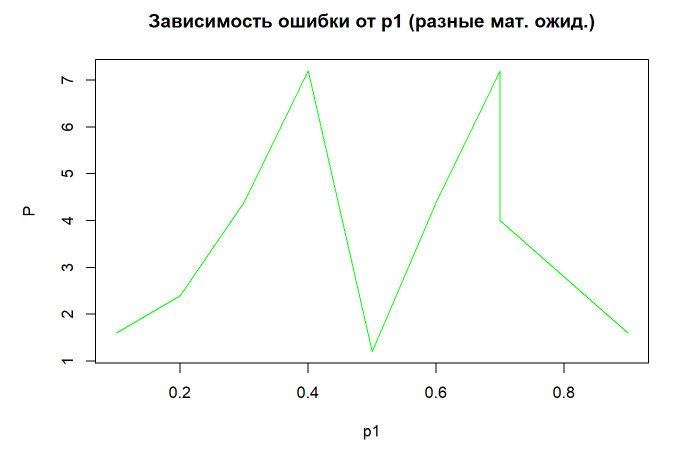


Рисунок 14 – Зависимость ошибки от p1 (разн. мат. ожид)

# 6 Вывод

В ходе выполнения расчетно-графической работы изучено байесовское решающее правило классификации. Получен опыт его программной реализации для нормально распределенной обучающей выборки. Исследовано влияние параметров классификатора на вероятность его ошибки. Закреплены навыки программирования на языке R, создания пользовательского интерфейса с помощью технологии html.

**7 Список использованных источников**

1. Система электронного обучения СФУ - https://e.sfu-kras.ru
2. Рубан, А. И. Методы анализа данных. Учебное пособие /А.И. Рубан, //уч.пособие. 2-е изд., исправл. и доп. Красноярск: ИПЦ КГТУ, 2004. 319 с
3. Рубан, А.И. Учебное пособие по циклу расчетно-графических работпо курсу «Методы анализа данных» /А.И. Рубан, А.В. Кузнецов. — 59 с