

Белорусский государственный университет
информатики и радиоэлектроники

Кафедра ЭВМ

**Отчет по лабораторной работе №4
по курсу ХиУИ**

Вариант #30

Выполнил:
студент группы 7М2432
Пантелеев В.В.

Проверил:
Марченко В.В.

МИНСК 2017

Выполнение работы

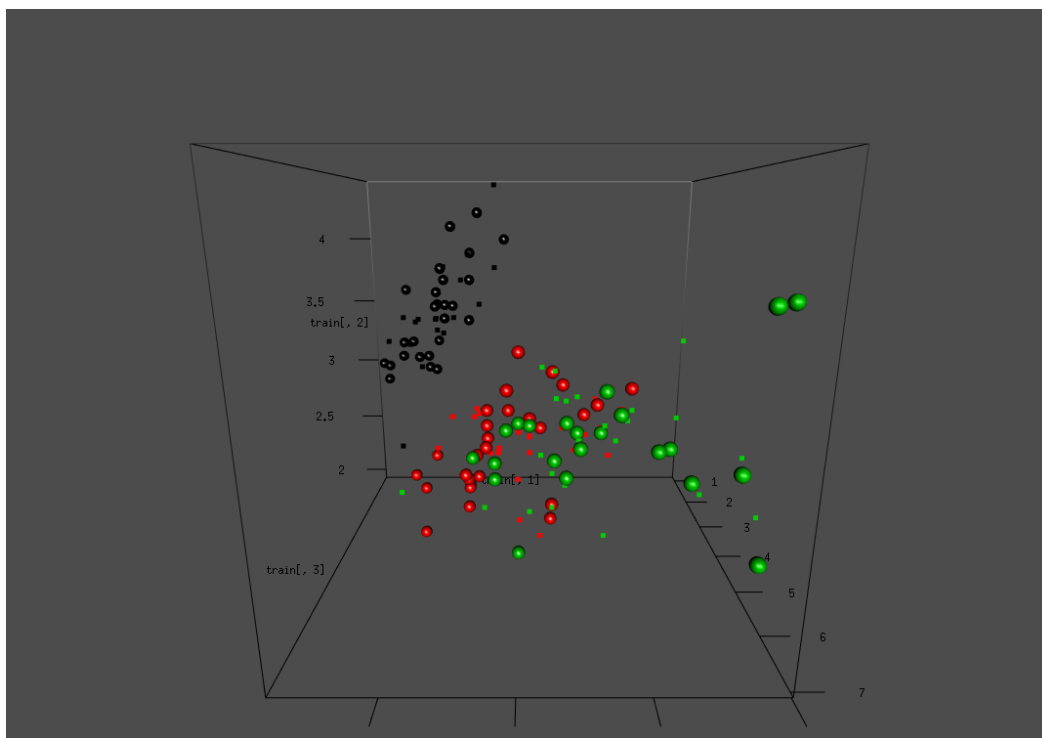
Листинг программы:

```
require(class)
require(MASS)
require(rgl)
plot_points <- function(train, test, clazz.train,
                        clazz.test) {
  rgl.open()
  plot3d(train[, 1], train[, 2], train[, 3],
         col=clazz.train, type='p', size=5, add=FALSE)
  plot3d(test[, 1], test[, 2], test[, 3],
         col=clazz.test, type='s', size=1, add=TRUE)
}
analyse_knn <- function(dat, clazz) {
  n <- nrow(dat)
  rnd.num <- sample(1 : n)
  train.num <- rnd.num[1 : (n %% 2)]
  test.num <- rnd.num[(n %% 2 + 1) : n]
  train <- dat[train.num,]
  test <- dat[test.num,]
  clazz.train <- clazz[train.num]
  clazz.test <- clazz[test.num]
  clazz.knn <- knn(train, test, clazz.train)
  print(sum(clazz.test != clazz.knn) / n)
  plot_points(train, test, clazz.train, clazz.test)
  plot_points(train, test, clazz.train, clazz.knn)
}
fileName = "/home/panteleev/Documents/Data/iris.data.txt"
irisData <- read.table(fileName, header = FALSE, sep = ",")
head(irisData)
analyse_knn(cbind(irisData$V1, irisData$V2,
                  irisData$V3), unclass(irisData$V5))

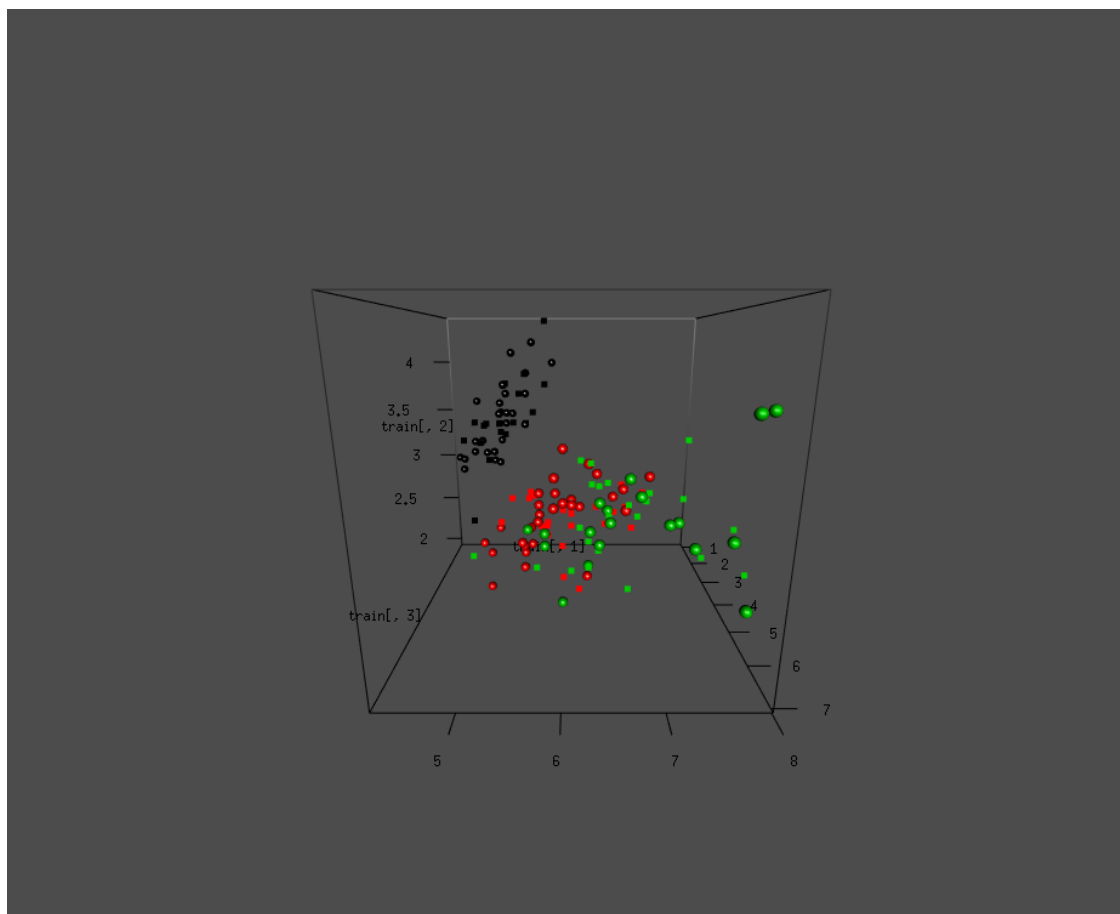
n1 <- 1000
a1 <- c(7, 7, 7)
r1 <- cbind(c(3, -1, 1), c(-1, 3, 1), c(1, 1, 2))
n2 <- 100
a2 <- c(10, 4, 10)
r2 <- cbind(c(4, -1, -1), c(-1, 2, 0.1), c(-1, 0.1, 2))
dat <- rbind(mvrnorm(n1, a1, r1), mvrnorm(n2, a2, r2))
analyse_knn(dat, c(rep(1, n1), rep(2, n2)))
```

Результат выполнения программы:

Визуализация данных из файла *iris.data.txt* до кластеризации:



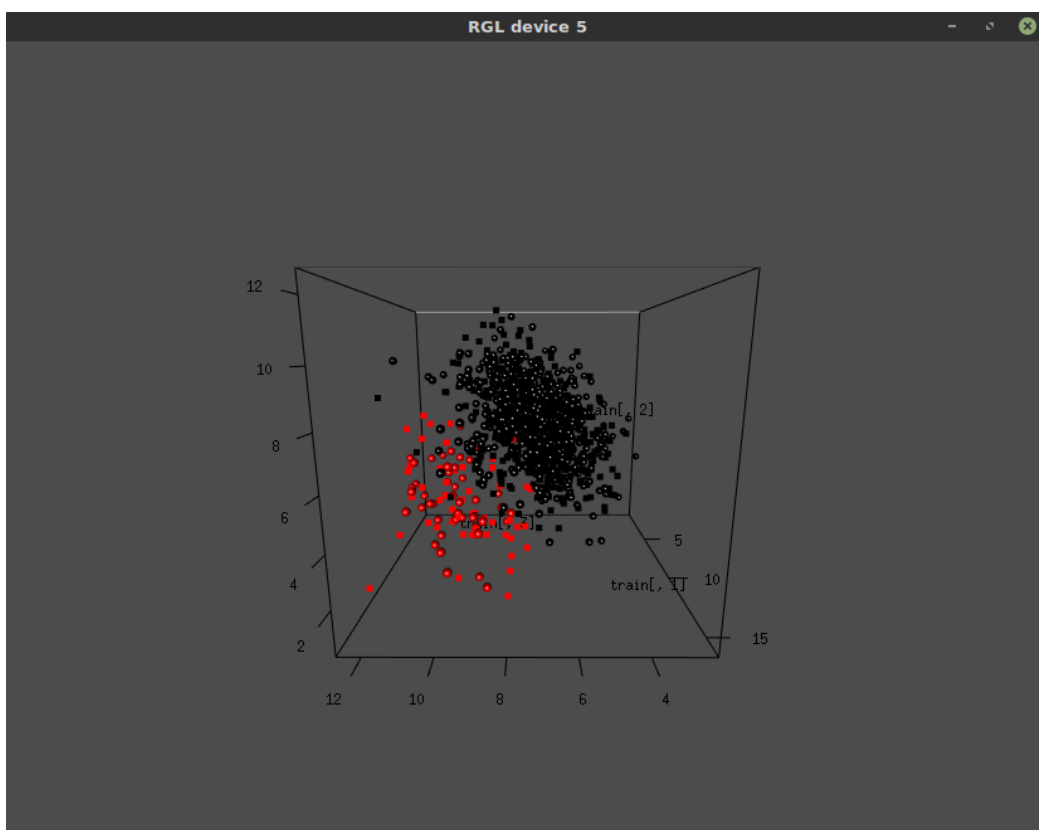
Визуализация данных из файла *iris.data.txt* после кластеризации:



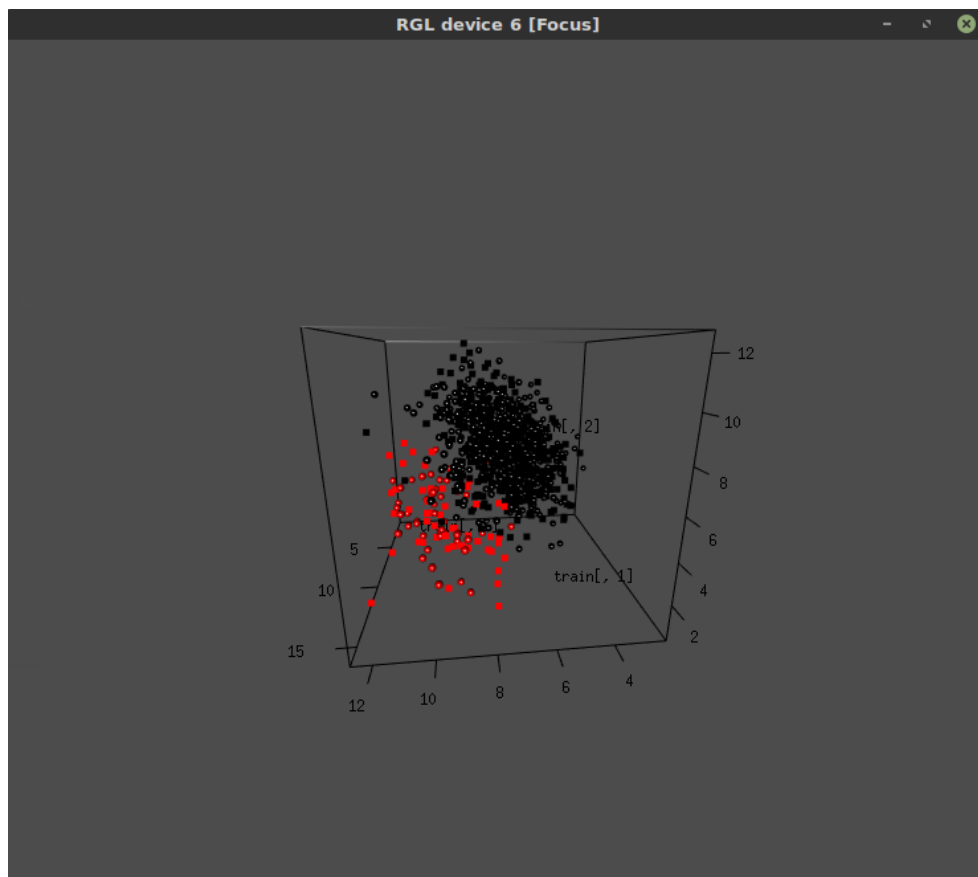
Вероятность ошибочной кластеризации — 0.04

Результат выполнения программы:

Визуализация сгенерированных данных до кластеризации:



Визуализация сгенерированных данных после кластеризации:



Вероятность ошибочной кластеризации — 0.009

Выводы

Вероятность ошибочной классификации очень мала и не превышает 4 процентов для реальных данных и 0.9 процентов для сгенерированных.