|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **Министерство науки и высшего образования Российской Федерации**  Калужский филиал  федерального государственного бюджетного  образовательного учреждения высшего образования  ***«Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»***  ***(КФ МГТУ им. Н.Э. Баумана)*** |

**ФАКУЛЬТЕТ** ***ИУК «Информатика и управление»***

**КАФЕДРА** \_\_***ИУК4 «Программное обеспечение ЭВМ, информационные технологии»***

**ЛАБОРАТОРНАЯ РАБОТА №1**

**«Линейные классификаторы»**

**ДИСЦИПЛИНА: «Методы машинного обучения»**

|  |  |  |
| --- | --- | --- |
| Выполнил: студент гр. ИУК4-62Б | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ( Калашников А.С. )  (Подпись) (Ф.И.О.) |
| Проверил: | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (      Кручинин И.И. )  (Подпись) (Ф.И.О.) |
| Дата сдачи (защиты):  Результаты сдачи (защиты): | | |
|  | - Балльная оценка:  - Оценка: | |

Калуга, 2023

**Цель:** сформировать практические навыки разработки программ с использованием основ линейного классификатора на языке R.

**Вариант 2**

**Задание:**

**№ 1.** Создайте фрейм данных из N = 30 записей со следующими полями: Nrow – номер записи, Name – имя сотрудника, BirthYear – год рождения, EmployYear – год приема на работу, Salary – зарплата. EyEColor – цвет глаз, SkinColor – цвет кожи, BloodType – группа крови, HairColor – цвет волос на голове. Заполните данный фрейм данными так, что Nrow изменяется от 1 до N, Name задается произвольно, BithYear распределен равномерно (случайно) на отрезке [1974,1993], EmployYear распределен равномерно на отрезке [BirthY ear+17,2014], Salary для работников младше 1989 г.р. определяется по формуле Salary = (ln(2015 − EmployY ear) +1) ∗ 7000, для остальных Salary = (log2(2015 − EmployY ear) + 1) ∗ 7000.

Подсчитайте число сотрудников с зарплатой, большей 16000. Добавьте в таблицу поле, соответствующее суммарному подоходному налогу (ставка 13%), выплаченному сотрудником за время работы в организации, если его зарплата за каждый год начислялась согласно формулам для Salary, где вместо 2015 следует последовательно подставить каждый год работы сотрудника в организации.

**№ 2.** Постройте линейный классификатор для классификации сотрудников данной международной организации (признаки классификации: группа крови, цвет волос, глаз и цвет кожи). Использовать машину опорных векторов и алгоритм персептрона. Полученные результаты сравнить. Цвет глаз, кожи или волос можно закодировать определенным числом. В данном варианте использовать национальности: Англичанин, Китаец, Немец.

Для машины опорных векторов типа "C-classification" с сигмоидальным ядром, добейтесь нулевой ошибки сначала на обучающей выборке, а затем на тестовой, путем изменения параметра C.

**№ 3.** Допустим, что решающая функция линейного классификатора в упрощенном виде выглядит так:

Найти координаты и значение функции в точке минимума методом наискорейшего градиентного спуска.

**Листинг алгоритма для работы с фреймом:**

N = 30

Nrow = 1:N

Name = c("Liza","Kate","Sonya","Ann","Stepan","Jim","Mary","John","Kile","Emma",

  "Liza","Kate","Sonya","Ann","Stepan","Jim","Mary","John","Kile","Emma","Liza",

  "Kate","Sonya","Ann","Stepan","Jim","Mary","John","Kile","Emma")

BirthYear = round(runif(N, 1974, 1993))

EmployYear = round(BirthYear+17, 1993)

Salary = ifelse(BirthYear > 1989 , (log(2015-EmployYear)+1)\*7000, (log10(2015-EmployYear)+1)\*7000)

EyeColor = EyeColor <- c("0","1","2","0","1","2","0","1","2","2","0","1", "2","0","1","2","0","1","2","2","0","1","2","0","1","2","0","1","2","2")

ScinColor = ScinColor <- c("1","2","1","1","1","1","2","1","1","1","1","2","1","1","1","1","2","1","1","1","1","2","1","1","1","1","2","1","1","1")

BloodType = round(runif(N, 1, 4))

HairColor = HairColor <- c("1","2","3","2","2","1","3","1","1","3","1","3","1","1","3","3","2","3","1","1","3","2","1","3","3","1","2","1","1","3")

frame = data.frame(Nrow, Name, BirthYear, EmployYear, EyeColor, Salary ,

ScinColor, BloodType, HairColor)

countNeedSalary<- length(which(Salary > 16000))

> countNeedSalary

frame$SocialVychet = frame$Salary \* 0.13\*(2019-EmployYear)

**Листинг алгоритма классификации:**

coun = 3

Mrow = 1:coun

Group = c("англичанин", "китаец", "немец")

English\_min = c(0.30,1.14,2.03)

Chineese\_min = c(0.14,0.33,0.73)

German\_min = c(0.00,1.14,1.73)

English\_max = c(1.2,2.0,1.94)

Chineese\_max = c(2.0,0.02,2.96)

German\_max = c(1.2,1.2,1.94)

frame1 = data.frame(Mrow, Group, English\_min, Chineese\_max , German\_min,

English\_max, Chineese\_min, German\_max)

View(frame1)

ramFoT = frame1

library(kohonen)

library(RSNNS)

library(class)

library(gmodels)

library(modeest)

library(nnet)

library(e1071)

ramFoT$Group

ramFo2 = frame

ramFo2 <- cbind(ramFo2, Group)

model <- svm(ramFo2$Group ~ ., data = ramFo2, type = "C-classification",

kernel="sigmoid")

ramFoT$Group

print(model)

summary(model)

x <- subset(ramFo2[7:9])

y <- ramFo2$Group

model <- svm(x, y,type = "C-classification", kernel="sigmoid ")

print(model)

summary(model)

pred <- predict(model, x)

table(pred, y)

**Листинг алгоритма градиентного спуска:**

f <- function(x1, x2) {

    (x2^2 + x1^2 - 1)^2 + (x1 + x2 - 1)^2

}

x = seq(-3, 3, by = 0.1)

y = seq(-3, 3, by = 0.1)

z <- scale(outer(x,y,f))

par(mar=rep(2,4))

view <- persp(x, y, z, phi = 10, theta = 10, xlim=c(-3,3), ylim=c(-3,3),

              xlab = "X", ylab = "Y", zlab = "Z", scale = FALSE,

              main = "F(u,v)", col="yellow", ticktype = "detailed")

set.seed(2)

pts <- data.frame(x = sample(x, 60),

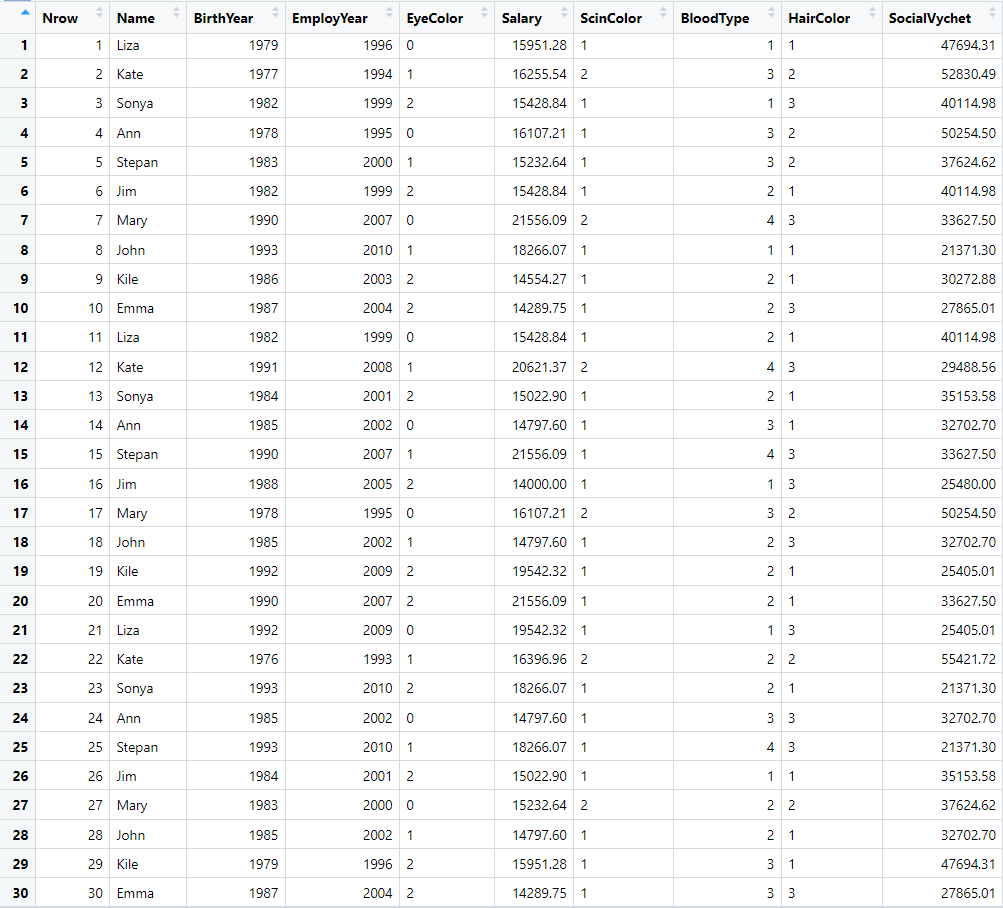
                  y = sample(y, 60),

                  z = sample(z, 60))

points(trans3d(x = pts$x, y = pts$y, z = pts$z, pmat = view), pch = 16)

lines(trans3d(x = pts$x, y = pts$y, z = pts$z, pmat = view))

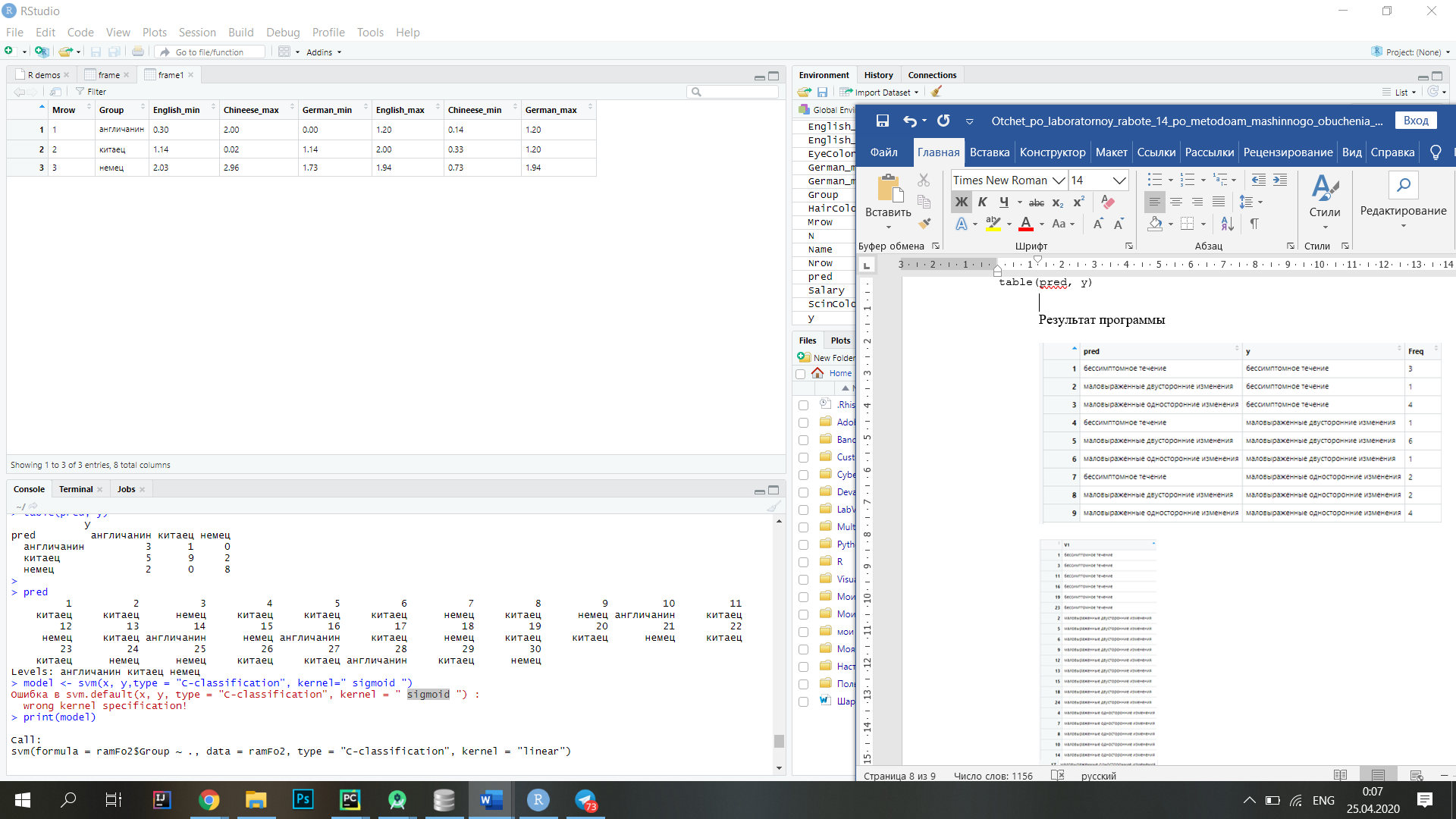
**Демонстрация работы алгоритма:**



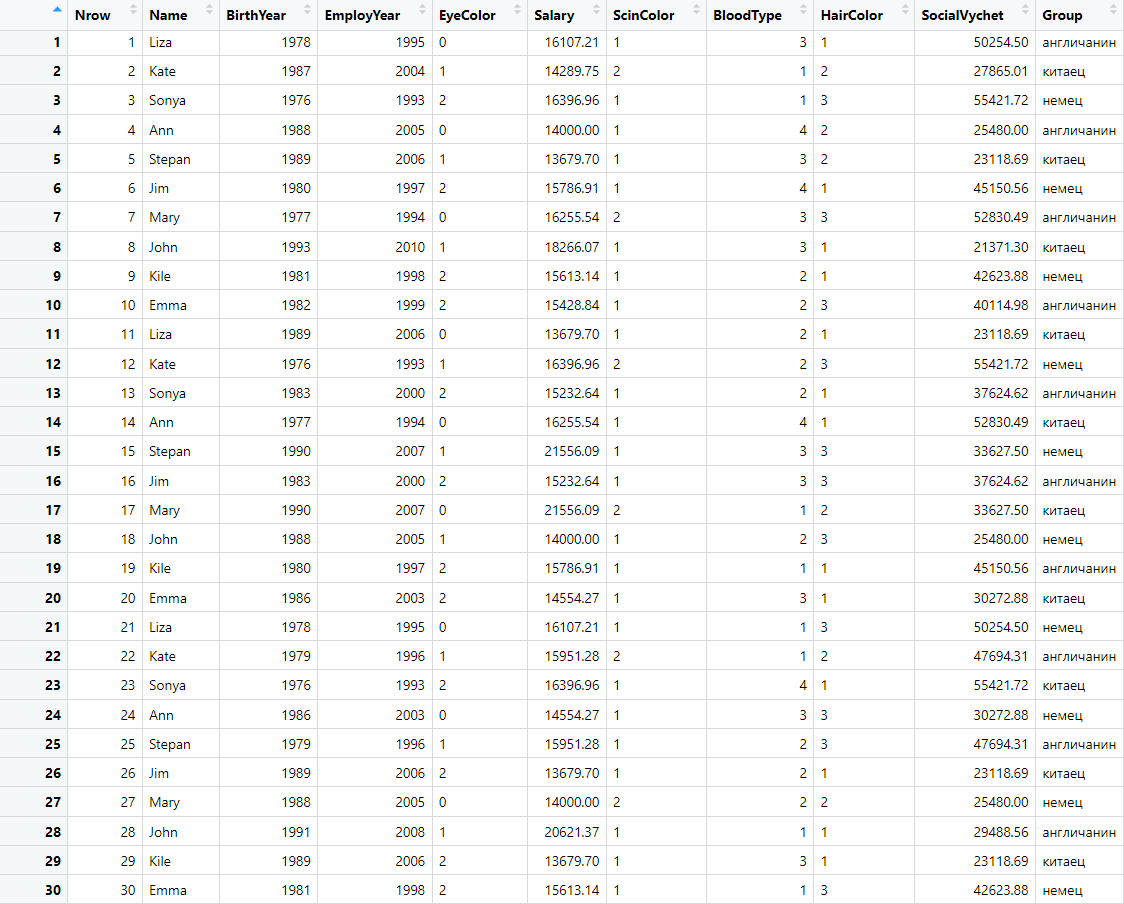
**Рис.1.** Демонстрация основного фрейма с добавлением столбца,   
содержащего значения суммарного подходного налога



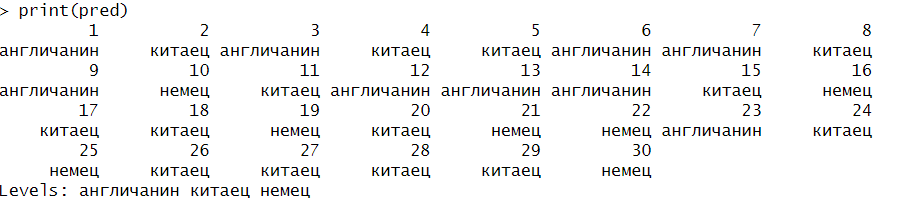
**Рис.2.** Демонстрация подсчёта числа сотрудников с зарплатой более 16000



**Рис. 3.** Полученная матрица

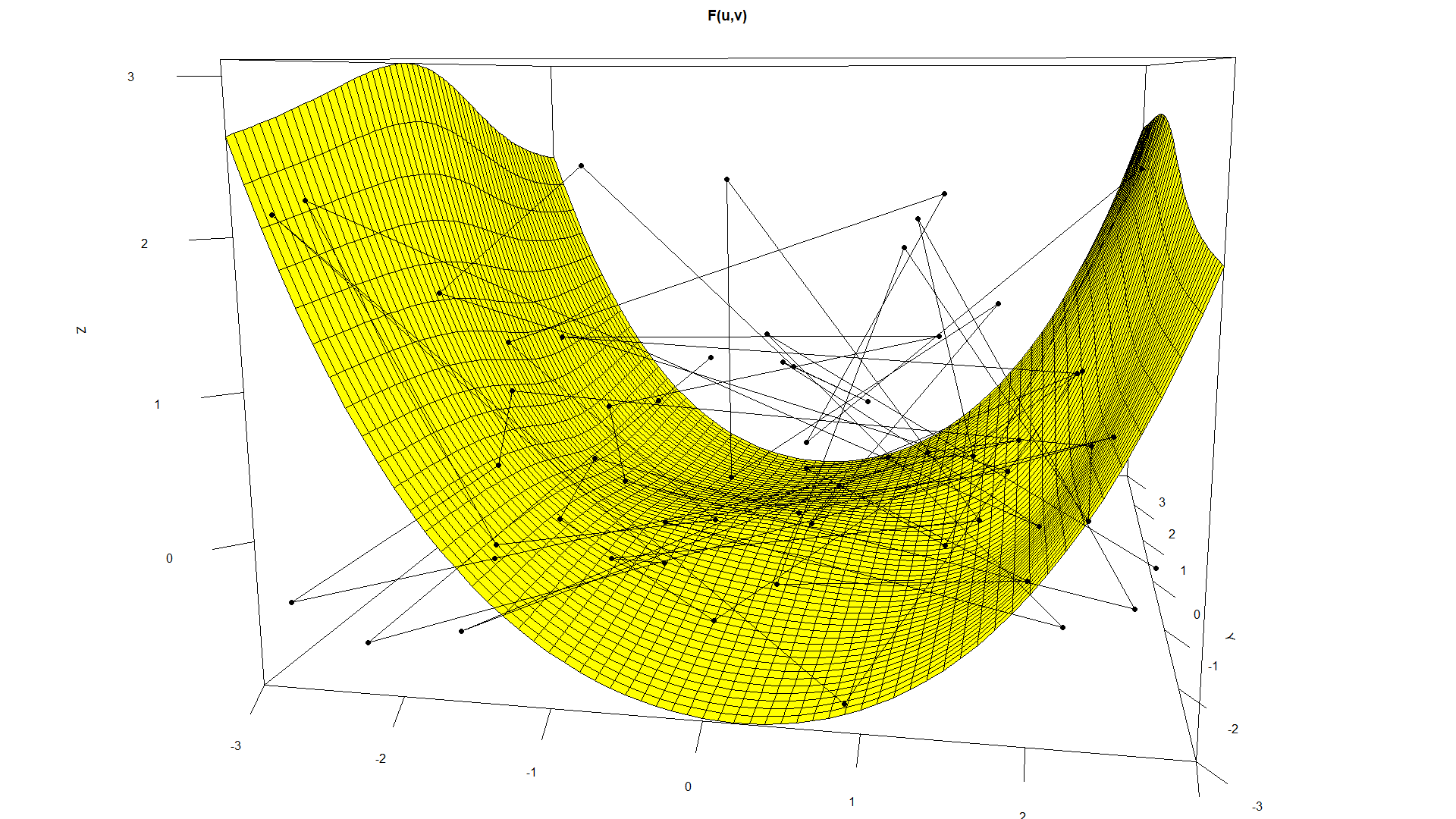
****

**Рис.4.** Полученный фрейм

****

**Рис.5.** Результат классификации

**Демонстрация работы алгоритма:**



**Рис. 6.** Построение функции с точками градиентного спуска

**Вывод:** в ходе выполнения данной лабораторной работы были получены навыки разработки программ с использованием основ линейного классификатора на языке R.