**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 *ИУ-КФ «Информатика и управление»***

**КАФЕДРА *ИУ4-КФ «Программное обеспечение ЭВМ, информационные***

***технологии»***

ЛАБОРАТОРНАЯ РАБОТА №1

**«Линейные классификаторы»**

# ДИСЦИПЛИНА: «Методы машинного обучения»

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

Калуга, 2022

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

# Вариант 2

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* = (*log*2(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.

1. 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*.

1. Допустим, что решающая функция линейного классификатора в упрощенном виде выглядит так:



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

# Реализация поставленных задач

**Листинг алгоритма для работы с фреймом:**

N = 30

Nrow = 1:N Name =

c("John","Kile","Emma","Oleg","Kate","Tolya","Max","Ann","Jim","Mary","John","Kile","Emma","Oleg","Kate","Tolya","Max","Ann","Jim","Mary","John","Kile","Emma","Oleg"," Kate","Tolya","Max","Ann","Jim","Mary")

BirthYear = round(runif(N, 1974, 1993)) EmployYear = round(BirthYear+17, 2014)

Salary = ifelse(BirthYear < 1989 , (log(2015-EmployYear)+1)\*7000, (log2(2015-EmployYear)+1)\*7000)

EyeColor = EyeColor <- c("1","4","1","0","1","2","0","1","2","3","0","1","4","0","3","4","0","1"

,"2","3","4","1","3","0","4","2","4","1","2","0")

ScinColor = ScinColor <- c("4","3","1","1","1","1","2","1","1","4","4","2","1","1","4","4","2","1"

,"1","3","1","2","3","1","4","1","3","2","1","1")

BloodType = round(runif(N, 1, 4)) HairColor = HairColor <-

c("5","9","3","8","2","1","3","1","4","3","1","5","1","1","4","5","2","3"

,"1","9","3","7","1","3","6","4","9","3","1","4")

frame = data.frame(Nrow, Name, BirthYear, EmployYear, EyeColor, Salary , ScinColor, BloodType, HairColor)

Результат работы программы: см. рис.1-3

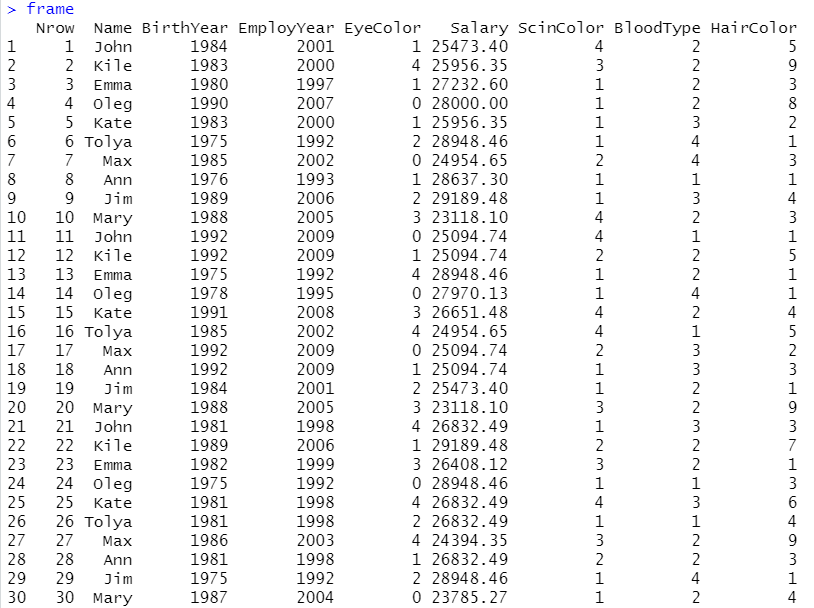


Рис.1. Созданный фрейм

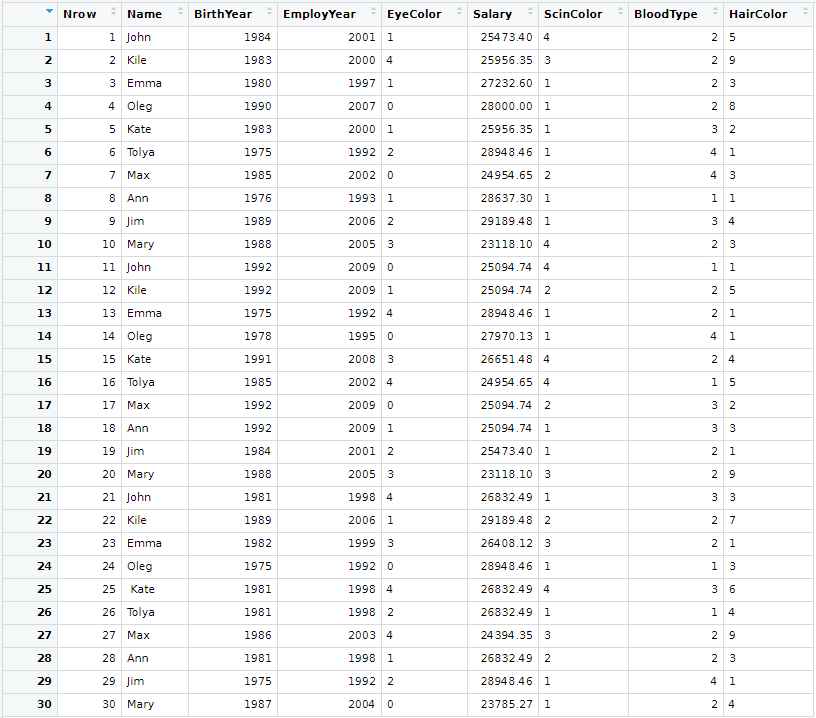


Рис.2. Открытие фрейма с помощью View() Подсчёт числа людей с Salary > 16000 в фрейме:

* countMy<- length(which(Salary>16000))
* countMy

Результат: 30

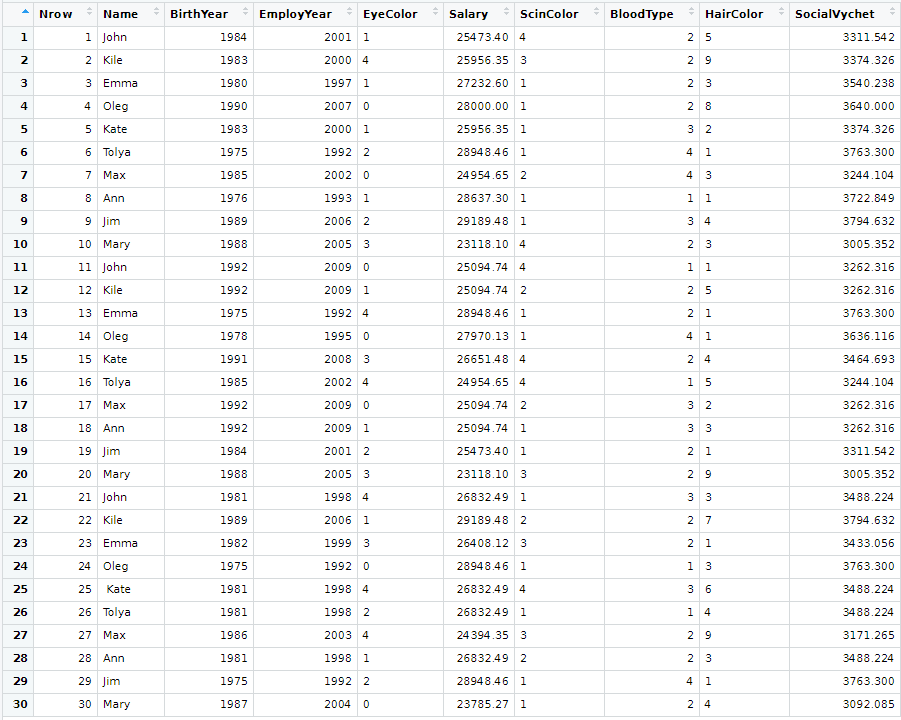


Рис.3. Модифицированный фрейм

# Листинг алгоритма классификации:

coun = 3

Mrow = 1:coun

Group = c(1, 2, 3) # 1 - Англичанин, 2 – Китаец, 3 - Немец

Eng\_min = c(1.30,1.14,2.03)

Chin\_min = c(2.80,1.14,3.43)

Ger\_min = c(0.20,1.64,1.03)

Eng\_max = c(3.30,3.54,4.03)

Chin\_max = c(4.00,2.14,4.03)

Ger\_max = c(3.10,3.14,4.03)

frame1 = data.frame(Mrow, Group, Eng\_min, Chin\_min, Ger\_min, Eng\_max, Chin\_max, Ger\_max)

ramFoT = frame1

library(e1071)

ramFo2 = frame1

x <- subset(ramFo2[3:8])

y <- ramFo2$Group

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

print(model)

pred <- predict(model, x)

pred

Результат работы программы: см. рис.4-5

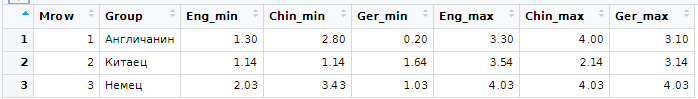


Рис.4. Загруженные данные



Рис.5. Результат классификации

rosenbrock <- function(v) {((v[2])^2 + (v[2])^2 - 1)^2 + (v[1] + v[2] - 1)^2}

result <- optim(

c(runif(1,-3,3), runif(1,-3,3)), method="Nelder-Mead",

control=c( # configure Nelder-Mead

maxit=100, # maximum iterations of 100

reltol=1e-8, # response tolerance over-one step alpha=1.0, # reflection factor

beta=0.5, # contraction factor gamma=2.0)) # expansion factor

print(result$par) print(result$value) print(result$counts)

x <- seq(-3, 3, length.out=100) y <- seq(-3, 3, length.out=100)

z <- rosenbrock(expand.grid(x, y))

contour(x, y, matrix(log10(z), length(x)), xlab="x",ylab="y") points(result$par[1], result$par[2], col="red", pch=19) rect(result$par[1]-0.2, result$par[2]-0.2, result$par[1]+0.2, result$par[2]+0.2, lwd=2)

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