**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. ИТД.Б-61 | (  (Подпись) | Бандурина Е.М.  (Ф.И.О.) | ) |
| Проверил: | (  (Подпись) | Кручинин И.И.  (Ф.И.О.) | ) |
| Дата сдачи (защиты): Результаты сдачи (защиты): |  | | |
| * Балльная оценка: * Оценка: | | | |

Калуга , 2020

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

# Вариант 3

1. Создайте фрейм данных из N = 27 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 [1972,1994], EmployYear распределен равномерно на отрезке [BirthY ear+19,2013], Salary для работников младше 1980 г.р. определяется по формуле Salary = (ln(2014 − EmployY ear) +1) ∗ 8000, для остальных Salary = (log2(2014 − EmployY ear) + 1) ∗ 8000.

Подсчитайте число сотрудников с зарплатой, большей 14000. 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 2014 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: Русский, Татарин, Португалец, Японец, Итальянец.

Для машины опорных векторов типа "C-classification" с радиальным 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:

*f* (*x* , *x* )  5  *x*2  5  *x*2  8  *x*  *x*

1 2 1 2 1 2

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

# Реализация поставленных задач

**Листинг алгоритма для работы с фреймом:**

N = 27

Nrow = 1:N Name =

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

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

BirthYear = round(runif(N, 1972, 1994)) EmployYear = round(BirthYear+19, 2013)

Salary = ifelse(BirthYear < 1989 , (log(2014-EmployYear)+1)\*8000, (log2(2014-EmployYear)+1)\*8000)

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")

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")

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")

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

Результат работы программы: см. рис.1-3

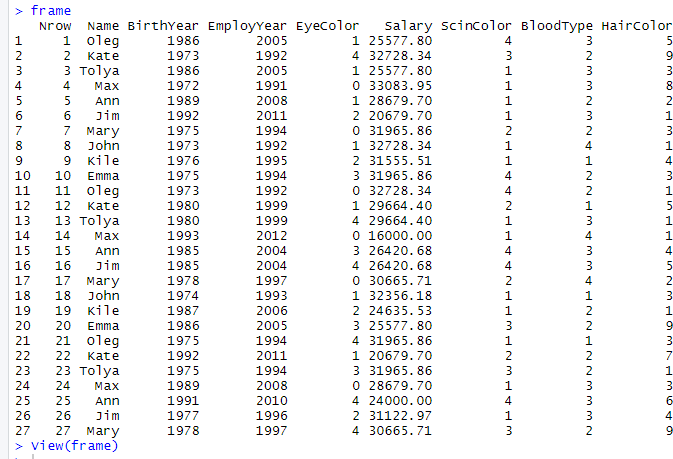


Рис.1. Созданный фрейм

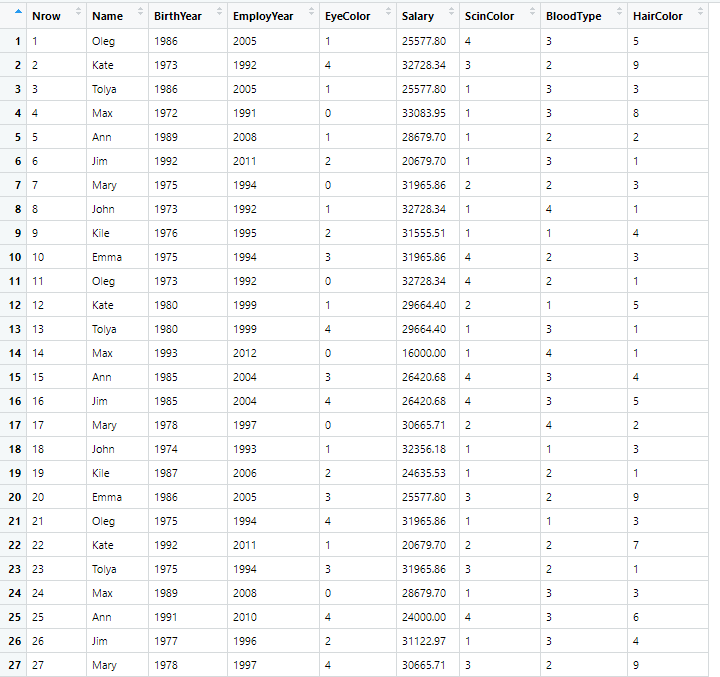


Рис.2. Открытие фрейма с помощью View() Подсчёт числа людей с Salary > 14000 в фрейме:

* countMy<- length(which(Salary>14000))
* countMy

Результат: 27

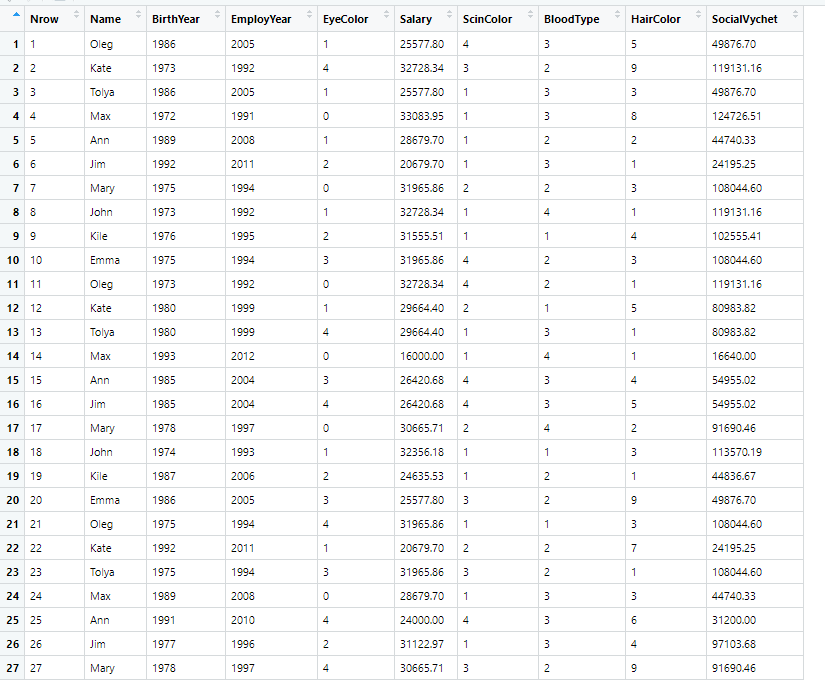


Рис.3. Модифицированный фрейм

# Листинг алгоритма классификации:

coun = 5

Mrow = 1:coun

Group = c("Русский","Татарин","Португалец","Японец","Итальянец") Rus\_min = c(0.20,1.64,1.03,2.0, 0.3)

Tat\_min = c(3.30,2.14,2.03,4.0, 0.12)

Port\_min = c(2.30,3.14,4.03,3.0, 0.41)

Jap\_min = c(2.80,1.14,3.43,4.0, 0.62)

Ita\_min = c(1.30,1.14,2.03,3.0, 0.32)

Rus\_max = c(3.10,3.14,4.03,6.0, 3.01)

Tat\_max = c(4.00,2.14,2.03,9.0, 5.62)

Port\_max = c(2.60,1.86,2.03,7.0, 5.32)

Jap\_max = c(4.00,2.14,4.03,0.0, 2.13)

Ita\_max = c(3.30,3.54,4.03,5.0, 4.24)

frame1 = data.frame(Mrow, Group, Rus\_min, Tat\_min, Port\_min, Jap\_min, Ita\_min, Rus\_max, Tat\_max, Port\_max, Jap\_max, Ita\_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="linaling")

ramFoT$Group print(model) summary(model)

x <- subset(ramFo2[7:9]) y <- ramFo2$Group

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

summary(model)

pred <- predict(model, x) table(pred, y)

Результат работы программы: см. рис.4-5



Рис.4. Загруженные данные



Рис.5. Результат классификации

rosenbrock <- function(v) {5\*(v[1])^2+5\*(v[2])^2-8\*v[1]\*v[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.