



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

_____ (Подпись) (Калашников А.С.)
(Ф.И.О.)

Проверил:

_____ (Подпись) (Кручинин И.И.)
(Ф.И.О.)

Дата сдачи (защиты):

Результаты сдачи (защиты):

- Балльная оценка:

- Оценка:

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, BirthYear распределен равномерно (случайно) на отрезке [1974,1993], EmployYear распределен равномерно на отрезке [BirthYear+17,2014], Salary для работников младше 1989 г.р. определяется по формуле $Salary = (\ln(2015 - EmployYear) + 1) * 7000$, для остальных $Salary = (\log_2(2015 - EmployYear) + 1) * 7000$.

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:

$$f(x_1, x_2) = (x_2^2 + x_1^2 - 1)^2 + (x_1 + x_2 - 1)^2$$

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", "2", "1", "1", "1", "1", "2", "1", "1", "1", "1", "1", "2", "1", "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)
Chinese_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)
Chinese_max = c(2.0,0.02,2.96)
German_max = c(1.2,1.2,1.94)
frame1 = data.frame(Mrow, Group, English_min, Chinese_max, German_min,
English_max, Chinese_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:

	Nrow	Name	BirthYear	EmployYear	EyeColor	Salary	ScinColor	BloodType	HairColor	SocialVychet
1	1	Liza	1979	1996	0	15951.28	1	1	1	47694.31
2	2	Kate	1977	1994	1	16255.54	2	3	2	52830.49
3	3	Sonya	1982	1999	2	15428.84	1	1	3	40114.98
4	4	Ann	1978	1995	0	16107.21	1	3	2	50254.50
5	5	Stepan	1983	2000	1	15232.64	1	3	2	37624.62
6	6	Jim	1982	1999	2	15428.84	1	2	1	40114.98
7	7	Mary	1990	2007	0	21556.09	2	4	3	33627.50
8	8	John	1993	2010	1	18266.07	1	1	1	21371.30
9	9	Kile	1986	2003	2	14554.27	1	2	1	30272.88
10	10	Emma	1987	2004	2	14289.75	1	2	3	27865.01
11	11	Liza	1982	1999	0	15428.84	1	2	1	40114.98
12	12	Kate	1991	2008	1	20621.37	2	4	3	29488.56
13	13	Sonya	1984	2001	2	15022.90	1	2	1	35153.58
14	14	Ann	1985	2002	0	14797.60	1	3	1	32702.70
15	15	Stepan	1990	2007	1	21556.09	1	4	3	33627.50
16	16	Jim	1988	2005	2	14000.00	1	1	3	25480.00
17	17	Mary	1978	1995	0	16107.21	2	3	2	50254.50
18	18	John	1985	2002	1	14797.60	1	2	3	32702.70
19	19	Kile	1992	2009	2	19542.32	1	2	1	25405.01
20	20	Emma	1990	2007	2	21556.09	1	2	1	33627.50
21	21	Liza	1992	2009	0	19542.32	1	1	3	25405.01
22	22	Kate	1976	1993	1	16396.96	2	2	2	55421.72
23	23	Sonya	1993	2010	2	18266.07	1	2	1	21371.30
24	24	Ann	1985	2002	0	14797.60	1	3	3	32702.70
25	25	Stepan	1993	2010	1	18266.07	1	4	3	21371.30
26	26	Jim	1984	2001	2	15022.90	1	1	1	35153.58
27	27	Mary	1983	2000	0	15232.64	2	2	2	37624.62
28	28	John	1985	2002	1	14797.60	1	2	1	32702.70
29	29	Kile	1979	1996	2	15951.28	1	3	1	47694.31
30	30	Emma	1987	2004	2	14289.75	1	3	3	27865.01

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

```
> countNeedSalary<- length(which(Salary > 16000))
> countNeedSalary
[1] 13
```

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

	Mrow	Group	English_min	Chinese_max	German_min	English_max	Chinese_min	German_max
1	1	англичанин	0.30	2.00	0.00	1.20	0.14	1.20
2	2	китаец	1.14	0.02	1.14	2.00	0.33	1.20
3	3	немец	2.03	2.96	1.73	1.94	0.73	1.94

Рис. 3. Полученная матрица

	Nrow	Name	BirthYear	EmployYear	EyeColor	Salary	ScinColor	BloodType	HairColor	SocialVychet	Group
1	1	Liza	1978	1995	0	16107.21	1	3	1	50254.50	англичанин
2	2	Kate	1987	2004	1	14289.75	2	1	2	27865.01	китаец
3	3	Sonya	1976	1993	2	16396.96	1	1	3	55421.72	немец
4	4	Ann	1988	2005	0	14000.00	1	4	2	25480.00	англичанин
5	5	Stepan	1989	2006	1	13679.70	1	3	2	23118.69	китаец
6	6	Jim	1980	1997	2	15786.91	1	4	1	45150.56	немец
7	7	Mary	1977	1994	0	16255.54	2	3	3	52830.49	англичанин
8	8	John	1993	2010	1	18266.07	1	3	1	21371.30	китаец
9	9	Kile	1981	1998	2	15613.14	1	2	1	42623.88	немец
10	10	Emma	1982	1999	2	15428.84	1	2	3	40114.98	англичанин
11	11	Liza	1989	2006	0	13679.70	1	2	1	23118.69	китаец
12	12	Kate	1976	1993	1	16396.96	2	2	3	55421.72	немец
13	13	Sonya	1983	2000	2	15232.64	1	2	1	37624.62	англичанин
14	14	Ann	1977	1994	0	16255.54	1	4	1	52830.49	китаец
15	15	Stepan	1990	2007	1	21556.09	1	3	3	33627.50	немец
16	16	Jim	1983	2000	2	15232.64	1	3	3	37624.62	англичанин
17	17	Mary	1990	2007	0	21556.09	2	1	2	33627.50	китаец
18	18	John	1988	2005	1	14000.00	1	2	3	25480.00	немец
19	19	Kile	1980	1997	2	15786.91	1	1	1	45150.56	англичанин
20	20	Emma	1986	2003	2	14554.27	1	3	1	30272.88	китаец
21	21	Liza	1978	1995	0	16107.21	1	1	3	50254.50	немец
22	22	Kate	1979	1996	1	15951.28	2	1	2	47694.31	англичанин
23	23	Sonya	1976	1993	2	16396.96	1	4	1	55421.72	китаец
24	24	Ann	1986	2003	0	14554.27	1	3	3	30272.88	немец
25	25	Stepan	1979	1996	1	15951.28	1	2	3	47694.31	англичанин
26	26	Jim	1989	2006	2	13679.70	1	2	1	23118.69	китаец
27	27	Mary	1988	2005	0	14000.00	2	2	2	25480.00	немец
28	28	John	1991	2008	1	20621.37	1	1	1	29488.56	англичанин
29	29	Kile	1989	2006	2	13679.70	1	3	1	23118.69	китаец
30	30	Emma	1981	1998	2	15613.14	1	1	3	42623.88	немец

Рис.4. Полученный фрейм

```
> print(pred)
      1      2      3      4      5      6      7      8
англичанин китаец англичанин китаец китаец англичанин англичанин китаец
      9     10     11     12     13     14     15     16
англичанин немец китаец англичанин англичанин англичанин китаец немец
     17     18     19     20     21     22     23     24
китаец китаец немец китаец немец немец англичанин китаец
     25     26     27     28     29     30
немец китаец китаец китаец китаец немец
Levels: англичанин китаец немец
```

Рис.5. Результат классификации

Демонстрация работы алгоритма:

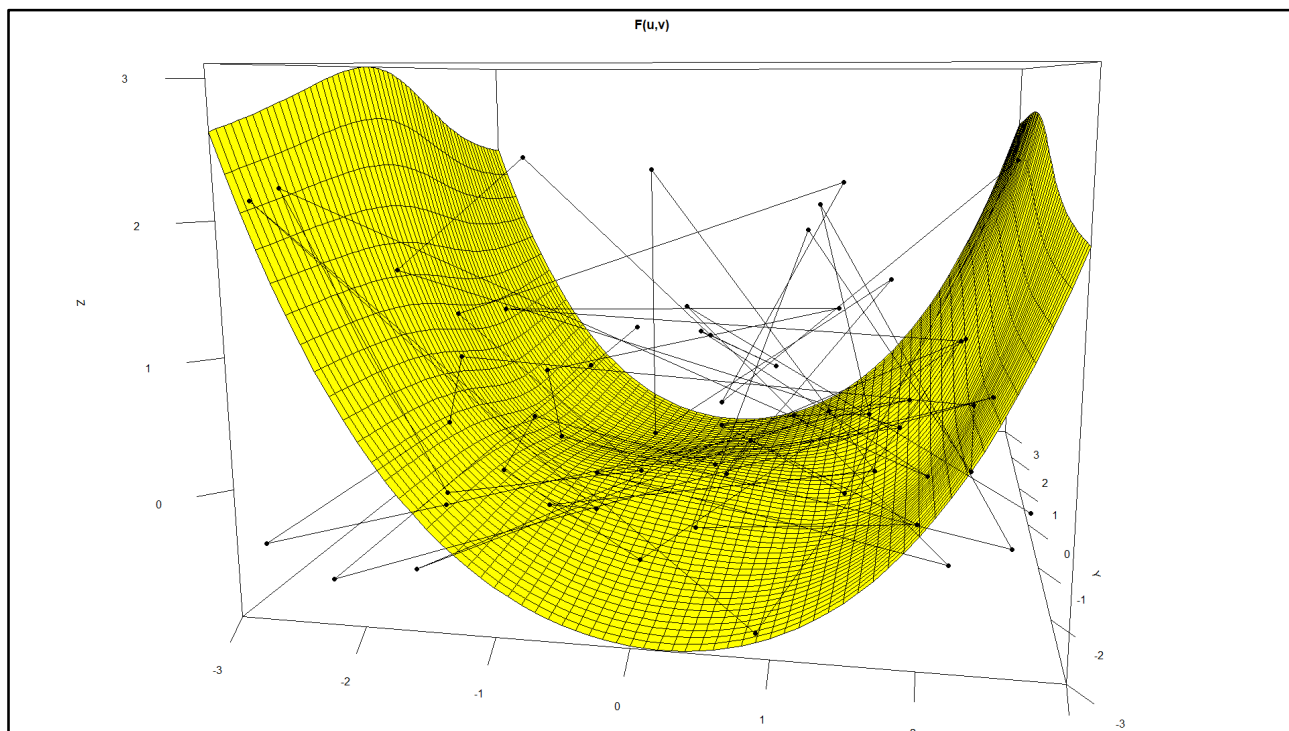


Рис. 6. Построение функции с точками градиентного спуска

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