



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, *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.

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

```
> frame
```

	Nrow	Name	BirthYear	EmployYear	EyeColor	Salary	ScinColor	BloodType	HairColor
1	1	John	1984	2001	1	25473.40	4	2	5
2	2	Kile	1983	2000	4	25956.35	3	2	9
3	3	Emma	1980	1997	1	27232.60	1	2	3
4	4	Oleg	1990	2007	0	28000.00	1	2	8
5	5	Kate	1983	2000	1	25956.35	1	3	2
6	6	Tolya	1975	1992	2	28948.46	1	4	1
7	7	Max	1985	2002	0	24954.65	2	4	3
8	8	Ann	1976	1993	1	28637.30	1	1	1
9	9	Jim	1989	2006	2	29189.48	1	3	4
10	10	Mary	1988	2005	3	23118.10	4	2	3
11	11	John	1992	2009	0	25094.74	4	1	1
12	12	Kile	1992	2009	1	25094.74	2	2	5
13	13	Emma	1975	1992	4	28948.46	1	2	1
14	14	Oleg	1978	1995	0	27970.13	1	4	1
15	15	Kate	1991	2008	3	26651.48	4	2	4
16	16	Tolya	1985	2002	4	24954.65	4	1	5
17	17	Max	1992	2009	0	25094.74	2	3	2
18	18	Ann	1992	2009	1	25094.74	1	3	3
19	19	Jim	1984	2001	2	25473.40	1	2	1
20	20	Mary	1988	2005	3	23118.10	3	2	9
21	21	John	1981	1998	4	26832.49	1	3	3
22	22	Kile	1989	2006	1	29189.48	2	2	7
23	23	Emma	1982	1999	3	26408.12	3	2	1
24	24	Oleg	1975	1992	0	28948.46	1	1	3
25	25	Kate	1981	1998	4	26832.49	4	3	6
26	26	Tolya	1981	1998	2	26832.49	1	1	4
27	27	Max	1986	2003	4	24394.35	3	2	9
28	28	Ann	1981	1998	1	26832.49	2	2	3
29	29	Jim	1975	1992	2	28948.46	1	4	1
30	30	Mary	1987	2004	0	23785.27	1	2	4

Рис.1. Созданный фрейм

	▼	Nrow	Name	BirthYear	EmployYear	EyeColor	Salary	ScinColor	BloodType	HairColor
1		1	John	1984	2001	1	25473.40	4	2	5
2		2	Kile	1983	2000	4	25956.35	3	2	9
3		3	Emma	1980	1997	1	27232.60	1	2	3
4		4	Oleg	1990	2007	0	28000.00	1	2	8
5		5	Kate	1983	2000	1	25956.35	1	3	2
6		6	Tolya	1975	1992	2	28948.46	1	4	1
7		7	Max	1985	2002	0	24954.65	2	4	3
8		8	Ann	1976	1993	1	28637.30	1	1	1
9		9	Jim	1989	2006	2	29189.48	1	3	4
10		10	Mary	1988	2005	3	23118.10	4	2	3
11		11	John	1992	2009	0	25094.74	4	1	1
12		12	Kile	1992	2009	1	25094.74	2	2	5
13		13	Emma	1975	1992	4	28948.46	1	2	1
14		14	Oleg	1978	1995	0	27970.13	1	4	1
15		15	Kate	1991	2008	3	26651.48	4	2	4
16		16	Tolya	1985	2002	4	24954.65	4	1	5
17		17	Max	1992	2009	0	25094.74	2	3	2
18		18	Ann	1992	2009	1	25094.74	1	3	3
19		19	Jim	1984	2001	2	25473.40	1	2	1
20		20	Mary	1988	2005	3	23118.10	3	2	9
21		21	John	1981	1998	4	26832.49	1	3	3
22		22	Kile	1989	2006	1	29189.48	2	2	7
23		23	Emma	1982	1999	3	26408.12	3	2	1
24		24	Oleg	1975	1992	0	28948.46	1	1	3
25		25	Kate	1981	1998	4	26832.49	4	3	6
26		26	Tolya	1981	1998	2	26832.49	1	1	4
27		27	Max	1986	2003	4	24394.35	3	2	9
28		28	Ann	1981	1998	1	26832.49	2	2	3
29		29	Jim	1975	1992	2	28948.46	1	4	1
30		30	Mary	1987	2004	0	23785.27	1	2	4

Рис.2. Открытие фрейма с помощью View()

Подсчёт числа людей с Salary > 16000 в фрейме:

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

Результат: 30

▲	Nrow	Name	BirthYear	EmployYear	EyeColor	Salary	ScinColor	BloodType	HairColor	SocialVychet
1	1	John	1984	2001	1	25473.40	4	2	5	3311.542
2	2	Kile	1983	2000	4	25956.35	3	2	9	3374.326
3	3	Emma	1980	1997	1	27232.60	1	2	3	3540.238
4	4	Oleg	1990	2007	0	28000.00	1	2	8	3640.000
5	5	Kate	1983	2000	1	25956.35	1	3	2	3374.326
6	6	Tolya	1975	1992	2	28948.46	1	4	1	3763.300
7	7	Max	1985	2002	0	24954.65	2	4	3	3244.104
8	8	Ann	1976	1993	1	28637.30	1	1	1	3722.849
9	9	Jim	1989	2006	2	29189.48	1	3	4	3794.632
10	10	Mary	1988	2005	3	23118.10	4	2	3	3005.352
11	11	John	1992	2009	0	25094.74	4	1	1	3262.316
12	12	Kile	1992	2009	1	25094.74	2	2	5	3262.316
13	13	Emma	1975	1992	4	28948.46	1	2	1	3763.300
14	14	Oleg	1978	1995	0	27970.13	1	4	1	3636.116
15	15	Kate	1991	2008	3	26651.48	4	2	4	3464.693
16	16	Tolya	1985	2002	4	24954.65	4	1	5	3244.104
17	17	Max	1992	2009	0	25094.74	2	3	2	3262.316
18	18	Ann	1992	2009	1	25094.74	1	3	3	3262.316
19	19	Jim	1984	2001	2	25473.40	1	2	1	3311.542
20	20	Mary	1988	2005	3	23118.10	3	2	9	3005.352
21	21	John	1981	1998	4	26832.49	1	3	3	3488.224
22	22	Kile	1989	2006	1	29189.48	2	2	7	3794.632
23	23	Emma	1982	1999	3	26408.12	3	2	1	3433.056
24	24	Oleg	1975	1992	0	28948.46	1	1	3	3763.300
25	25	Kate	1981	1998	4	26832.49	4	3	6	3488.224
26	26	Tolya	1981	1998	2	26832.49	1	1	4	3488.224
27	27	Max	1986	2003	4	24394.35	3	2	9	3171.265
28	28	Ann	1981	1998	1	26832.49	2	2	3	3488.224
29	29	Jim	1975	1992	2	28948.46	1	4	1	3763.300
30	30	Mary	1987	2004	0	23785.27	1	2	4	3092.085

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

Mrow	Group	Eng_min	Chin_min	Ger_min	Eng_max	Chin_max	Ger_max
1	Англичанин	1.30	2.80	0.20	3.30	4.00	3.10
2	Китаец	1.14	1.14	1.64	3.54	2.14	3.14
3	Немец	2.03	3.43	1.03	4.03	4.03	4.03

Рис.4. Загруженные данные

```
> pred
      1      2      3
1.208830 1.913675 2.886325
```

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.