|  |  |
| --- | --- |
| **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** ***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 | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (\_\_\_\_Бурцев В.А.\_\_)  (Подпись) (Ф.И.О.) |
| Проверил: | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (\_\_\_\_ Кручинин И.И.        )  (Подпись) (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.

**Задание на лабораторную работу**

**Вариант 4**

1. Создайте фрейм данных из *N* = 24 записей со следующими полями: *Nrow* – номер записи, *Name* – имя пациента, *BirthYear* – год рождения, *Employ* – место работы, *Salary* – зарплата. Cost – стоимость лечения, Albumin – содержание альбумина в крови, TransFerrin – содержание трансферрина в крови, Ferritin – содержание ферритина в крови. Заполните данный фрейм данными так, что *Nrow* изменяется от 1 до *N*,

Name задается произвольно, *BithYear* распределен равномерно (случайно) на отрезке [1971*,*1997], , Cost для пациентов младше 1991 г.р. определяется по формуле Cost= (*ln*(2013 − Birth*Y ear*) +1) ∗ 11000, для остальных Cost= (*log*2(2013 − Birth*Y ear*) + 1) ∗ 10000.

Ранжируйте пациентов по стоимости лечения, начиная с минимальной суммы. Добавьте в таблицу поле, соответствующее общему социальному вычету за лечение (ставка 13%), выплаченному пациенту, если стоимость лечения за каждый год начислялась согласно формулам для Cost, где вместо 2013 следует последовательно подставить каждый год нахождения пациента под наблюдением.

1. Постройте линейный классификатор для классификации пациентов по степени развития болезни Паркинсона (3 степени - бессимптомное течение, маловыраженные односторонние изменения, маловыраженные двусторонние изменения) на основе трех признаков – содержания в крови альбумина, трансферрина, ферритина. Использовать машину опорных векторов и алгоритм персептрона. Полученные результаты сравнить с использованием функций языка R - Predict и Table.

Для машины опорных векторов типа "C-classification" с линейным 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 градиентного спуска.

Таблица показателей анализа крови

|  |  |  |  |
| --- | --- | --- | --- |
| № | Показатель | Диапазон значений считающихся нормальными | Примечание |
| 1 | Альбумин | 35-50 |  |
| 2 | Трансферрин | 2-4 |  |
| 3 | Ферритин | 20-250 |  |
| 4 | Билирубин | 8,6-20,5 |  |
| 5 | Глобулин | 40-60 |  |
| 6 | ГЕМОГЛОБИН | 120-160 |  |
| 7 | ЭРИТРОЦИТЫ | 3.7-5.1 |  |
| 8 | ЛЕЙКОЦИТЫ | 4-9 |  |
| 9 | лимфоциты, | 40.7 |  |
| 10 | нейтрофилы, , | 2-11 |  |
| 11 | Базофилы | 0.3 |  |
| 12 | плазмоциты | 3-11 |  |
| 13 | Эозинофилы | 2.4 |  |
| 14 | ТРОМБОЦИТЫ | 160-320 |  |
| 15 | ГЛЮКОЗА | 3.5-6.5 |  |
| 16 | ОБЩИЙ БЕЛОК | 60-80 |  |
| 17 | КРЕАТИНИН | 0.18 |  |
| 18 | Гематокрит | 36-48 |  |
| 19 | Ретикулоциты | 2-12 |  |
| 20 | Моноциты | 10 |  |
|  |  |  |  |

Формы заболевания болезни Паркинсона: Дрожательно-ригидная, Акинетико-ригидная, Акинетическая форма, Дрожательная форма, ригидная форма. Степень заболевания характеризуется этапами от 0 до 5 (бессимптомное течение, маловыраженные односторонние изменения. маловыраженные двусторонние изменения, двусторонние изменения с появлением первых заметных двигательных ограничений, ощутимые ограничения, пациент сам не справляется, полная потеря трудоспособности и самостоятельной жизнедеятельности).

**Ход выполнения работы**

1. Создайте фрейм данных из *N* = 24 записей со следующими полями: *Nrow* – номер записи, *Name* – имя пациента, *BirthYear* – год рождения, *Employ* – место работы, *Salary* – зарплата. Cost – стоимость лечения, Albumin – содержание альбумина в крови, TransFerrin – содержание трансферрина в крови, Ferritin – содержание ферритина в крови. Заполните данный фрейм данными так, что *Nrow* изменяется от 1 до *N*,

Name задается произвольно, *BithYear* распределен равномерно (случайно) на отрезке [1971*,*1997], , Cost для пациентов младше 1991 г.р. определяется по формуле Cost= (*ln*(2013 − Birth*Y ear*) +1) ∗ 11000, для остальных Cost= (*log*2(2013 − Birth*Y ear*) + 1) ∗ 10000.

Ранжируем пациентов по стоимости лечения, начиная с минимальной суммы.

Добавьте в таблицу поле, соответствующее общему социальному вычету за лечение (ставка 13%), выплаченному пациенту, если стоимость лечения за каждый год начислялась согласно формулам для Cost, где вместо 2013 следует последовательно подставить каждый год нахождения пациента под наблюдением.

Листинг программы на R

#подготовка данных для фрейма

N = 24

Nrow = 1:N

Name = c("Jane","Jack","John","Bill","Harry","Jim","Mary","Ann","Kate","Emma","Marco","Jane","Jack","John","Bill","Harry","Jim","Mary","Ann","Kate","Emma","Marco","Emma","Marco")

BirthYear = round(runif(N, 1971, 1997))

Employ = Employ <- c("School","University","Bank","Hospital","IT-office","School","University","Bank","Hospital","IT-office","School","University","Bank","Hospital","IT-office","School","University","Bank","Hospital","IT-office","School","University","Bank","IT-office")

Salary = round(runif(N, 100000, 200000))

Cost = ifelse(BirthYear > 1991 ,(log(2013 - BirthYear) +1) \* 11000, (log2(2013 - BirthYear) + 1) \* 10000)

Albumin = round(runif(N, 200, 900)) \* 10^-3

TransFerrin = round(runif(N, 200, 1000)) \* 10^-3

Ferritin = round(runif(N, 1000, 10000)) \* 10^-3

#созданиефрейма

frame = data.frame(Nrow, Name, BirthYear, Employ, Salary,Cost,Albumin,TransFerrin,Ferritin)

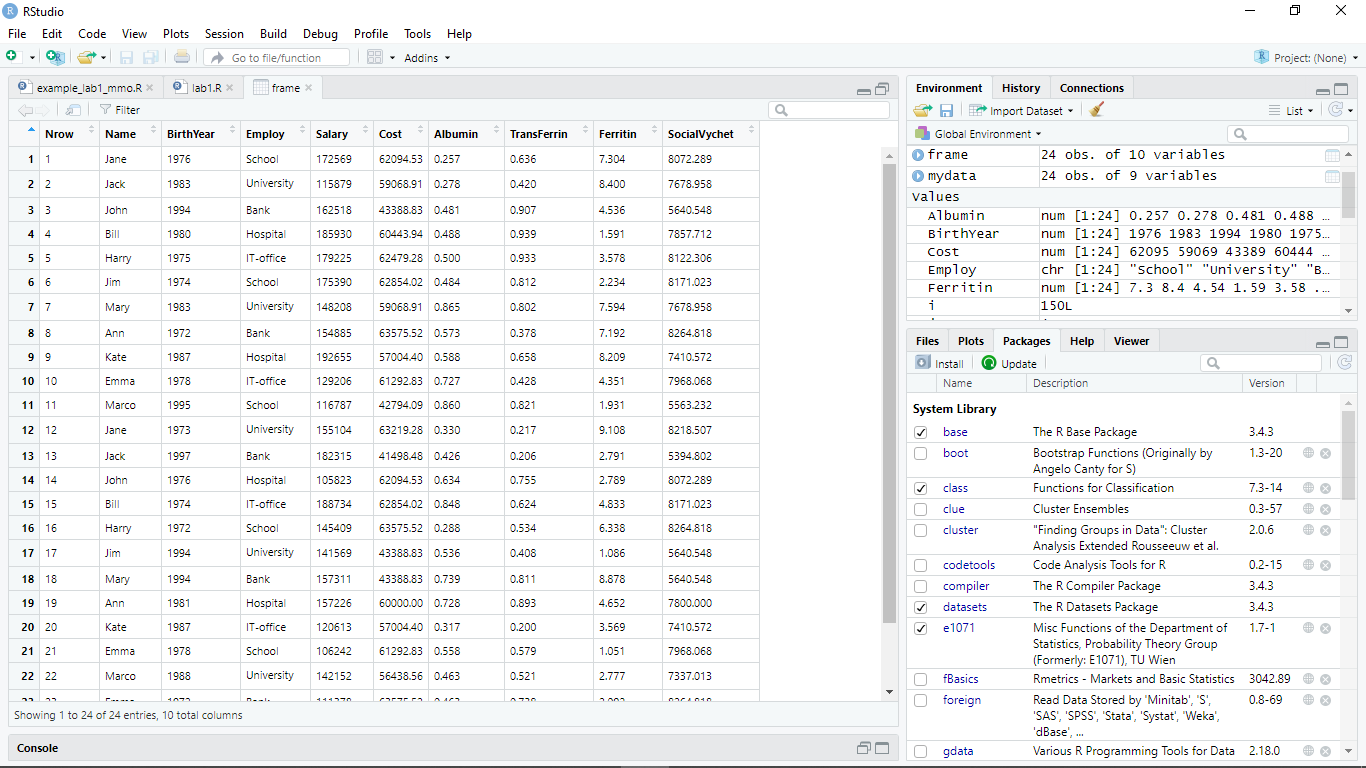
print(frame)

frame[order(frame$Cost),]

frame$SocialVychet = frame$Cost \* 0.13

View(frame)

Результат программы – созданный фрейм



1. Постройте линейный классификатор для классификации пациентов по степени развития болезни Паркинсона (3 степени - бессимптомное течение, маловыраженные односторонние изменения, маловыраженные двусторонние изменения) на основе трех признаков – содержания в крови альбумина, трансферрина, ферритина. Использовать машину опорных векторов и алгоритм персептрона. Полученные результаты сравнить с использованием функций языка R - Predict и Table.

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

Листинг программы на R

#подготовка данных для вспомогательного фрейма

coun = 3

Mrow = 1:coun

Group = c("бессимптомное течение","маловыраженные односторонние изменения","маловыраженные двусторонние изменения")

Albumin\_min = c(0.23,0.45,0.68)

TransFerrin\_min = c(0.14,0.33,0.73)

Ferritin\_min = c(1.41,2.09,4.44)

Albumin\_max = c(0.44,0.67,0.89)

TransFerrin\_max = c(0.32,0.72,0.96)

Ferritin\_max = c(2.08,4.43,6.53)

frame1 = data.frame(Mrow, Group, Albumin\_min, Albumin\_max , TransFerrin\_min,TransFerrin\_max,Ferritin\_min,Ferritin\_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="linear")

ramFoT$Group

print(model)

summary(model)

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

y <- ramFo2$Group

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

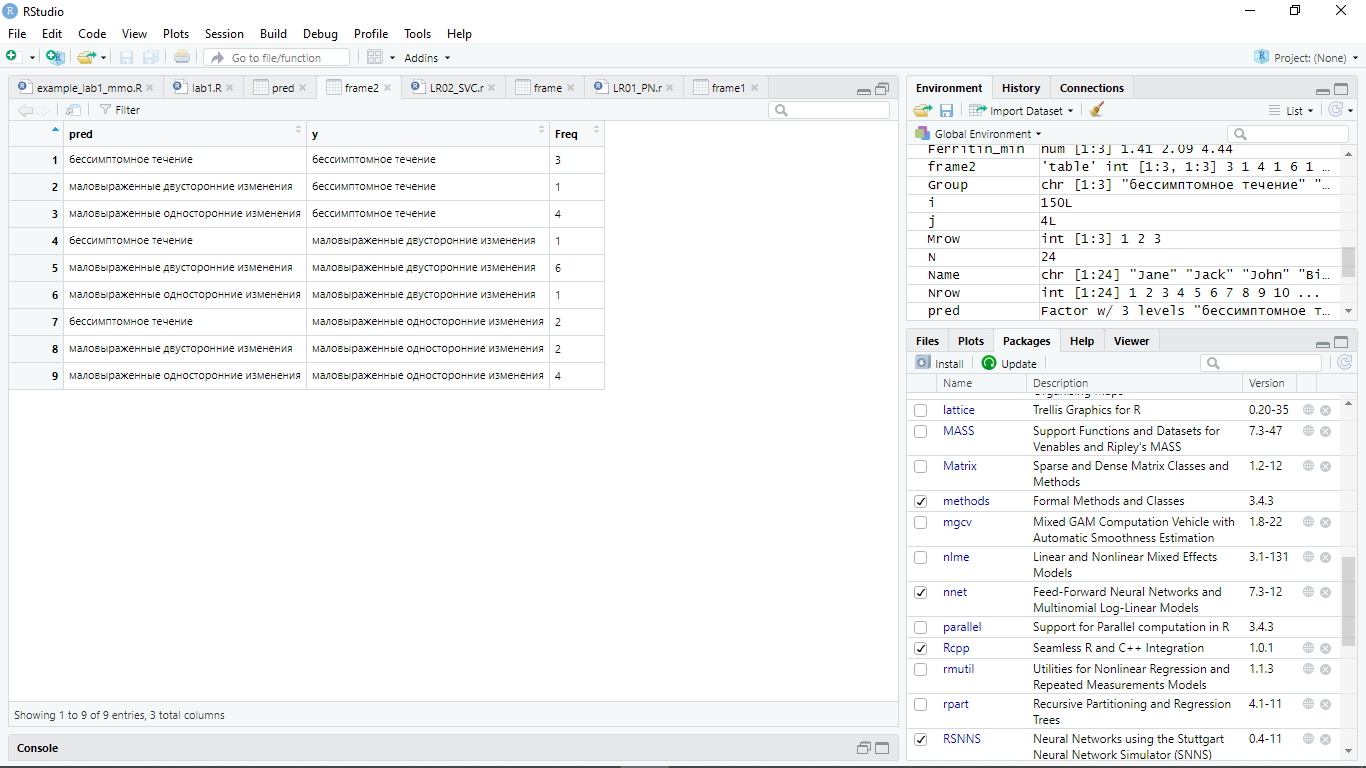
print(model)

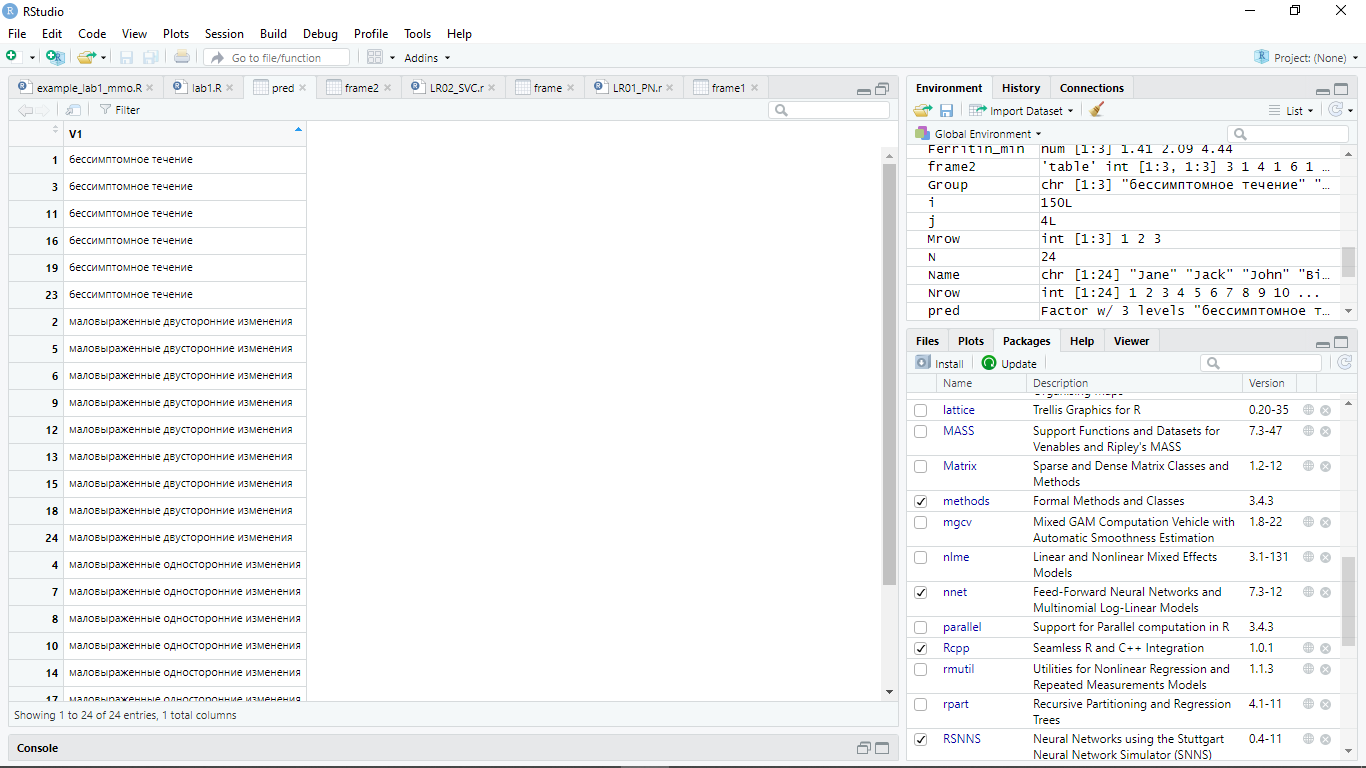
summary(model)

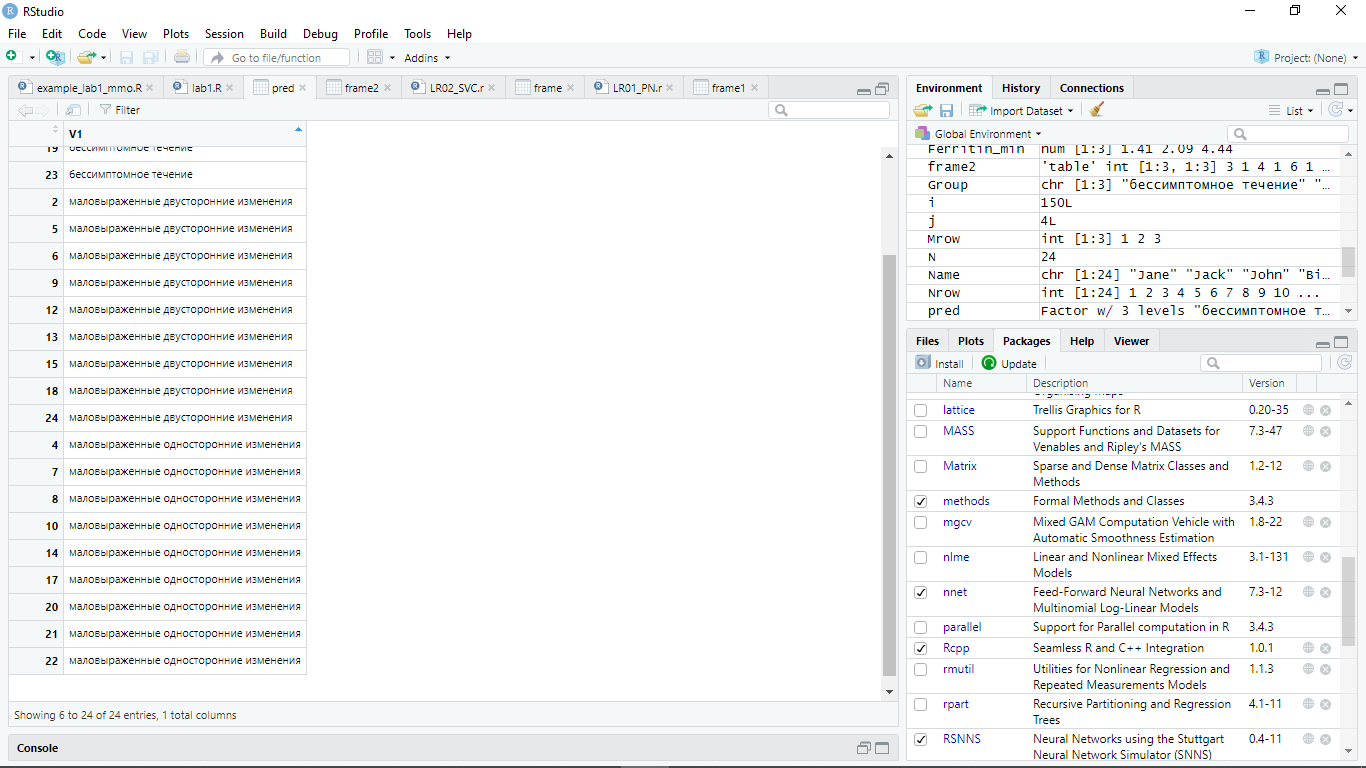
pred <- predict(model, x)

table(pred, y)

Результат программы







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



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

Листинг программы на Python

Результат программы

**Вывод:**

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