**Лабораторная работа №3**

по дисциплине “Машинное обучение”

**“Деревья решений”**

|  |  |
| --- | --- |
| Выполнил  студент гр. 33504/2 | Лелюхин Д.О. |
| Руководитель | Селин И.А. |

**Оглавление**

[**Задание** 3](#_Toc510978692)

[**Ход работы:** 4](#_Toc510978693)

[**Задание 1** 4](#_Toc510978694)

[**Задание 2** 5](#_Toc510978695)

[**Задание 3** 6](#_Toc510978696)

[**Задание 4** 8](#_Toc510978697)

[**Задание 5** 9](#_Toc510978698)

# **Задание**

1) Загрузите набор данных Glass из пакета “mlbench”. Набор данных (признаки, классы) был изучен в работе «Метод ближайших соседей». Постройте дерево классификации для модели, задаваемой следующей формулой: **Type ~ .**, дайте интерпретацию полученным результатам. При рисовании дерева используйте параметр cex=0.7 для уменьшения размера текста на рисунке, например, text(bc.tr,cex=0.7) или draw.tree(bc.tr,cex=0.7). Является ли построенное дерево избыточным? Выполните все операции оптимизации дерева.

2) Загрузите набор данных spam7 из пакета DAAG. Постройте дерево классификации для модели, задаваемой следующей формулой: **yesno ~.**, дайте интерпретацию полученным результатам. Запустите процедуру “**cost-complexity prunning**” с выбором параметра **k** по умолчанию, **method = ’misclass’**, выведите полученную последовательность деревьев. Какое из полученных деревьев, на Ваш взгляд, является оптимальным? Объясните свой выбор.

3) Загрузите набор данных nsw74psid1 из пакета DAAG. Постройте регрессионное дерево для модели, задаваемой следующей формулой: **re78 ~.**. Постройте регрессионную модель и SVM-регрессию для данной формулы. Сравните качество построенных моделей, выберите оптимальную модель и объясните свой выбор.

4) Загрузите набор данных Lenses Data Set из файла Lenses.txt:

3 класса (последний столбец): 1 : пациенту следует носить жесткие контактные линзы, 2 : пациенту следует носить мягкие контактные линзы, 3 : пациенту не следует носить контактные линзы.

Признаки (категориальные):

1. возраст пациента: (1) молодой, (2) предстарческая дальнозоркость, (3) старческая дальнозоркость 2. состояние зрения: (1) близорукий, (2) дальнозоркий

3. астигматизм: (1) нет, (2) да 4. состояние слезы: (1) сокращенная, (2) нормальная

Постройте дерево решений. Какие линзы надо носить при предстарческой дальнозоркости, близорукости, при наличии астигматизма и сокращенной слезы?

5) Постройте дерево решений для обучающего множества **Glass**, данные которого характеризуются 10-ю признаками:

1. Id number: 1 to 214; 2. RI: показатель преломления; 3. Na: сода (процент содержания в соотвествующем оксиде); 4. Mg; 5. Al; 6. Si; 7. K; 8. Ca; 9. Ba; 10. Fe.

Классы характеризуют тип стекла:

(1) окна зданий, плавильная обработка

(2) окна зданий, не плавильная обработка

(3) автомобильные окна, плавильная обработка

(4) автомобильные окна, не плавильная обработка (нет в базе)

(5) контейнеры

(6) посуда

(7) фары

Посмотрите заголовки признаков и классов. Перед построением классификатора необходимо также удалить первый признак Id number, который не несет никакой информационной нагрузки. Это выполняется командой **glass <- glass[,-1]**.

Определите, к какому типу стекла относится экземпляр с характеристиками

RI =1.516 Na =11.7 Mg =1.01 Al =1.19 Si =72.59 K=0.43 Ca =11.44 Ba =0.02 Fe =0.1

# **Ход работы:**

## **Задание 1**

library(mlbench)

library(tree)

data(Glass)

m <- dim(Glass)[1]

Glass.tr <- tree(Type ~., Glass)

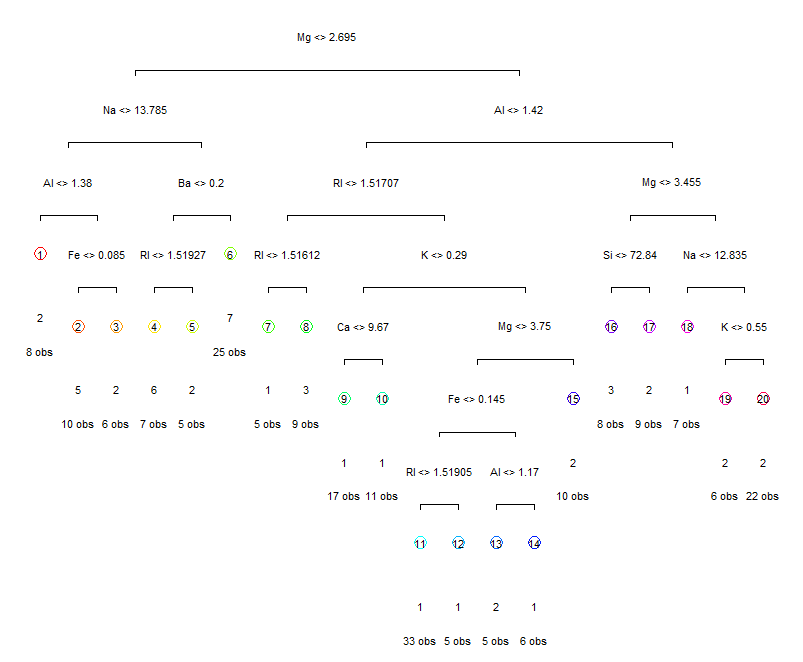
library(maptree)

draw.tree(Glass.tr, cex=0.7)

Glass.tr

Glass.tr1 <- prune.tree(Glass.tr, k = 56)

draw.tree(Glass.tr1, cex=0.7)



Оптимизация путем усечения вершин Fe < 0.145, Na > 12.835 и K < 0.29:

## **Задание 2**

library(DAAG)

library(tree)

data(spam7)

sp.tr <- tree(yesno ~., spam7)

library(maptree)

draw.tree(sp.tr, cex = 0.7)

tr1 <- prune.tree(spam\_tree,method = "misclass")

for(i in 2:4)

{

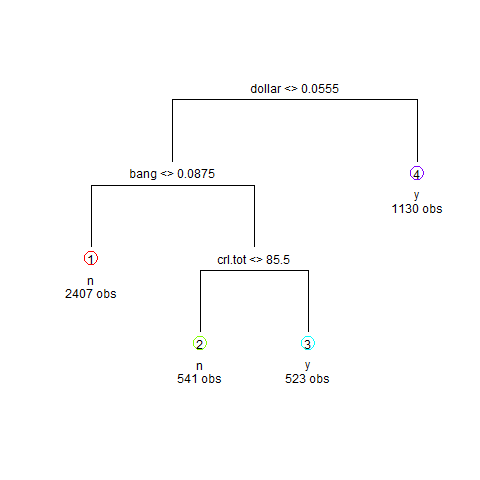
tr2 <- prune.tree(spam\_tree,k=tr1$k[i],method = "misclass")

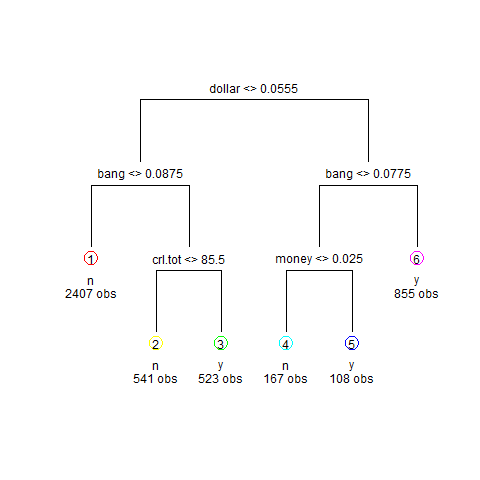
png(filename=paste(toString(i),'.jpg'))

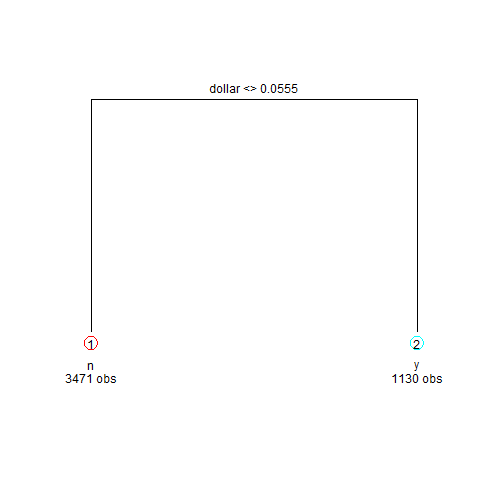
draw.tree(tr2)

dev.off()

}







На мой взгляд, оптимальным является 1 дерево, та как оно более упрощенное, и получает наиболее точную оценку классификации.

## **Задание 3**

library("tree")

library("DAAG")

library("maptree")

library("kernlab")

data(nsw74psid1)

tr <- tree(re78 ~.,nsw74psid1)

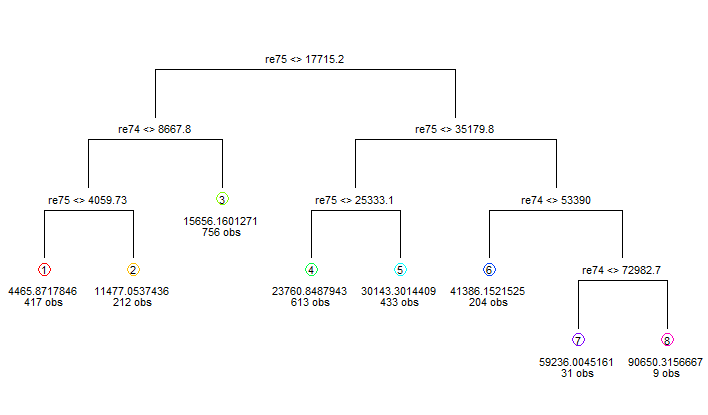
draw.tree(tr, cex = 0.7)

res <- lm(re78 ~ ., data = nsw74psid1)

summary(res)

confint(res)

ksvm(re78 ~ ., data=nsw74psid1)



Call:

lm(formula = re78 ~ ., data = nsw74psid1)

Residuals:

Min 1Q Median 3Q Max

-64870 -4302 -435 3786 110412

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -129.74276 1688.51706 -0.077 0.9388

trt 751.94643 915.25723 0.822 0.4114

age -83.56559 20.81380 -4.015 6.11e-05 \*\*\*

educ 592.61020 103.30278 5.737 1.07e-08 \*\*\*

black -570.92797 495.17772 -1.153 0.2490

hisp 2163.28118 1092.29036 1.981 0.0478 \*

marr 1240.51952 586.25391 2.116 0.0344 \*

nodeg 590.46695 646.78417 0.913 0.3614

re74 0.27812 0.02792 9.960 < 2e-16 \*\*\*

re75 0.56809 0.02756 20.613 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10070 on 2665 degrees of freedom

Multiple R-squared: 0.5864, Adjusted R-squared: 0.585

F-statistic: 419.8 on 9 and 2665 DF, p-value: < 2.2e-16

2.5 % 97.5 %

(Intercept) -3440.6790999 3181.1935733

trt -1042.7398553 2546.6327191

age -124.3784146 -42.7527563

educ 390.0484762 795.1719147

black -1541.8994503 400.0435060

hisp 21.4586646 4305.1037006

marr 90.9608911 2390.0781545

nodeg -677.7827244 1858.7166144

re74 0.2233653 0.3328768

re75 0.5140503 0.6221322

Support Vector Machine object of class "ksvm"

SV type: eps-svr (regression)

parameter : epsilon = 0.1 cost C = 1

Gaussian Radial Basis kernel function.

Hyperparameter : sigma = 0.146155276067426

Number of Support Vectors : 2023

Objective Function Value : -799.1829

Training error : 0.387178

Оптимальная модель – построенная при помощи *svm*, так как обладает наименьшей ошибкой при обучении.

## **Задание 4**

library("tree")

A\_raw <- read.table("Lenses.txt", sep = ',', stringsAsFactors = TRUE)

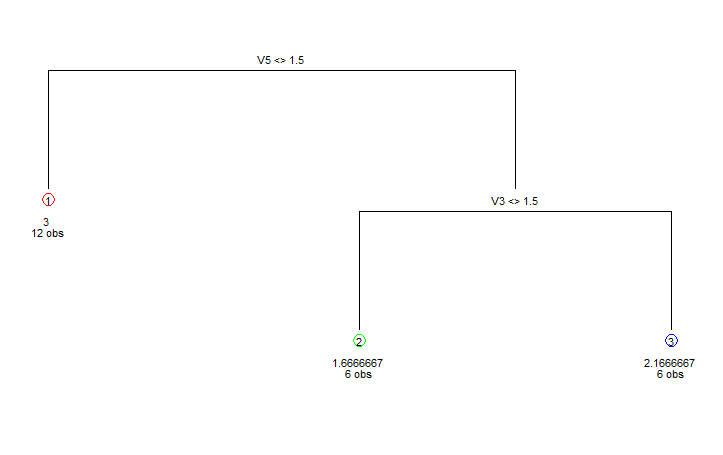
m <- dim(A\_raw)[1]

A\_raw <- A\_raw[,-1]

A\_raw.tr <- tree(V6 ~., A\_raw)

library(maptree)

draw.tree(A\_raw.tr, cex = 0.7)



Пациенту не следует носить линзы.

## **Задание 5**

library(kknn)

data(glass)

gl <- glass[,-1]

tr <- tree(Type~ .,gl)

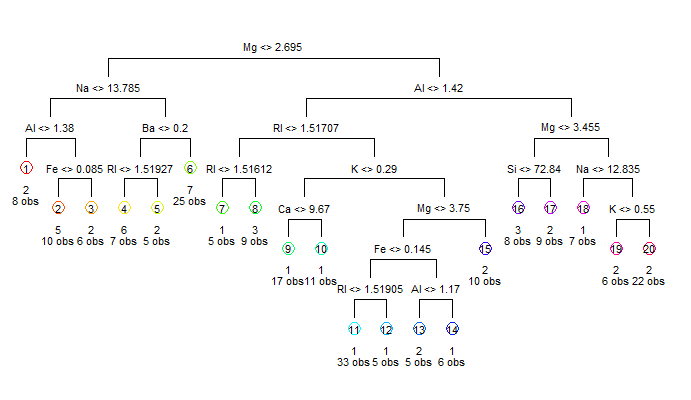
library(maptree)

draw.tree(tr,cex=0.7)

ex<-data.frame(1.516, 11.7, 1.01,1.19,72.59,0.43,11.44,0.02,0.1)

l <- c("RI", "Na", "Mg", "Al","Si","K","Ca","Ba","Fe")

colnames(ex) <- l



Относиться к типу (2) – окна зданий, не плавильная обработка.