

Міністерство освіти і науки України
Національний університет «Львівська політехніка»

ПРАКТИЧНЕ МАШИННЕ НАВЧАННЯ.

ДЕРЕВА РІШЕНЬ

Звіт

до лабораторної роботи №4

з курсу «Методи і засоби комп'ютерного навчання»

Виконав:

студент гр. СПКм-12

Сергієнко В.Р.

Прийняв

Пукач А.І.

Львів 2014

Мета: оволодіти навиками з побудови дерев класифікації та регресійних дерев для моделей в системі R.

Короткі теоретичні відомості

Функція *tree*

Опис:

```
tree(formula, data, weights, subset,
na.action = na.pass,
control = tree.control(nobs, ...),
method = "recursive.partition",
split = c("deviance", "gini"),
model = FALSE, x = FALSE, y = TRUE, wts = TRUE, ...)
```

Функція *snip.tree*

snip.tree містить дві пов'язані функції. Якщо аргумент *nodes* визначений, тоді функція видаляє задані вузли і всіх їхніх нащадків з дерева.

Якщо аргумент *nodes* не визначений, користувачеві пропонується вибрати вузли вручну (на графіку); це має сенс, якщо дерево вже намальоване. Вузол вибирається клацанням по лівій кнопці миші: при цьому номер вузла, *deviance* поточного дерева і те, що могло б залишитися, якщо даний вузол видаляється, друкується. Якщо вибрати цей вузол ще раз, тоді він буде видалений (і лінії, що відповідають піддереву з коренем в даному вузлі, витираються). Клацання по інших кнопках миші припиняє процес вибору.

```
snip.tree(tree, nodes, xy.save = FALSE,
digits = getOption("digits") - 3)
```

Індивідуальне завдання

1. Завантажити набір даних *monica* з пакету *DAAG* (http://www.stats.uwo.ca/DAAG/DAAG_0.97.tag.gz). Побудувати дерево класифікації для моделі, що задається наступною формулою: *outcome ~.*, дати інтерпретацію отриманим результатам. Чи є побудоване дерево надлишковим? Якщо так, то виконати операцію «*snip off*» над відповідними вузлами

Виконання завдання

```
cat("\014")
```

```
install.packages("DAAG")
```

```
...
```

```
The downloaded source packages are in
```

```
‘C:\tmp\Rtmpw4a5lU\downloaded_packages’
```

```
fil=monica
```

```
der=tree(outcome~.,fil)
```

```
der
```

```
node), split, n, deviance, yval, (yprob)
```

```
* denotes terminal node
```

```
1) root 6367 8753.00 live ( 0.553636 0.446364 )
```

```
2) hosp: y 4442 4532.00 live ( 0.792886 0.207114 )
```

```
4) highbp: y,n 4199 3861.00 live ( 0.827578 0.172422 )
```

```
8) age < 58.5 1647 1046.00 live ( 0.903461 0.096539 ) *
```

```
9) age > 58.5 2552 2698.00 live ( 0.778605 0.221395 ) *
```

```
5) highbp: nk 243 238.70 dead ( 0.193416 0.806584 ) *
```

```
3) hosp: n 1925 44.78 dead ( 0.001558 0.998442 ) *
```

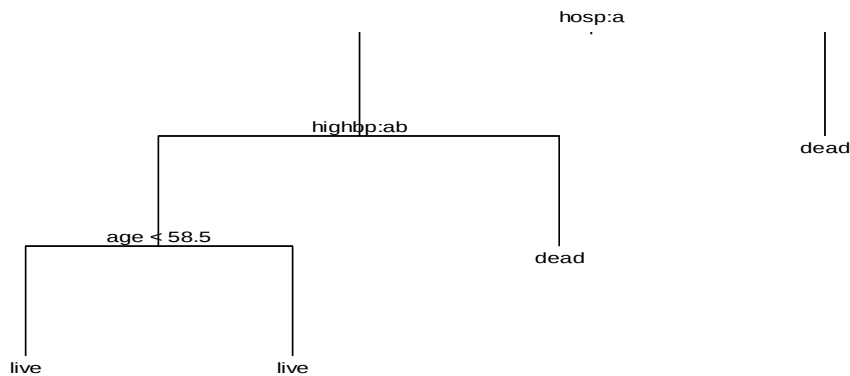
```
summary(der)
```

```
Classification tree:
```

```

tree(formula = outcome ~ ., data = fil)
Variables actually used in tree construction:
[1] "hosp" "highbp" "age"
Number of terminal nodes: 4
Residual mean deviance: 0.6329 = 4027 / 6363
Misclassification error rate: 0.1216 = 774 / 6367
plot(der,type="uniform")
text(der)

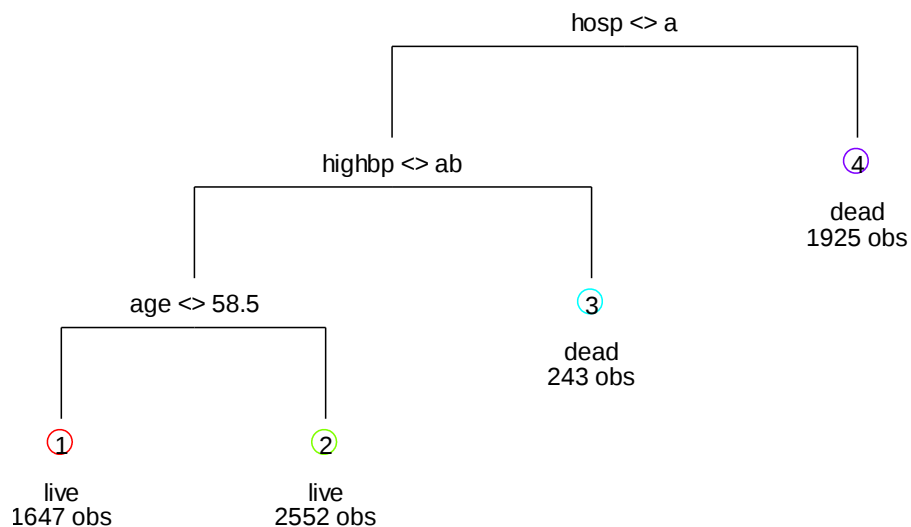
```



```

library(maptree)
draw.tree(der)

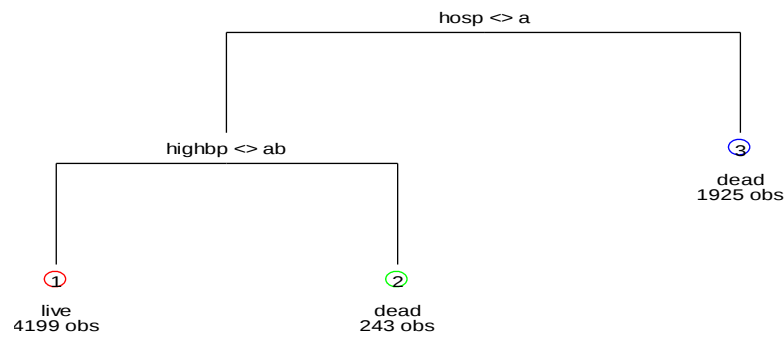
```



```

derevce=snip.tree(der,nodes=4)
derevce
node), split, n, deviance, yval, (yprob)
* denotes terminal node
1) root 6367 8753.00 live ( 0.553636 0.446364 )
2) hosp: y 4442 4532.00 live ( 0.792886 0.207114 )
4) highbp: y,n 4199 3861.00 live ( 0.827578 0.172422 ) *
5) highbp: nk 243 238.70 dead ( 0.193416 0.806584 ) *
3) hosp: n 1925 44.78 dead ( 0.001558 0.998442 ) *
draw.tree(derevce)

```



```
derevce2 <- prune.tree(der, k = 4)
```

```
derevce2
```

```
node), split, n, deviance, yval, (yprob)
```

```
* denotes terminal node
```

```
1) root 6367 8753.00 live ( 0.553636 0.446364 )
```

```
2) hosp: y 4442 4532.00 live ( 0.792886 0.207114 )
```

```
4) highbp: y,n 4199 3861.00 live ( 0.827578 0.172422 )
```

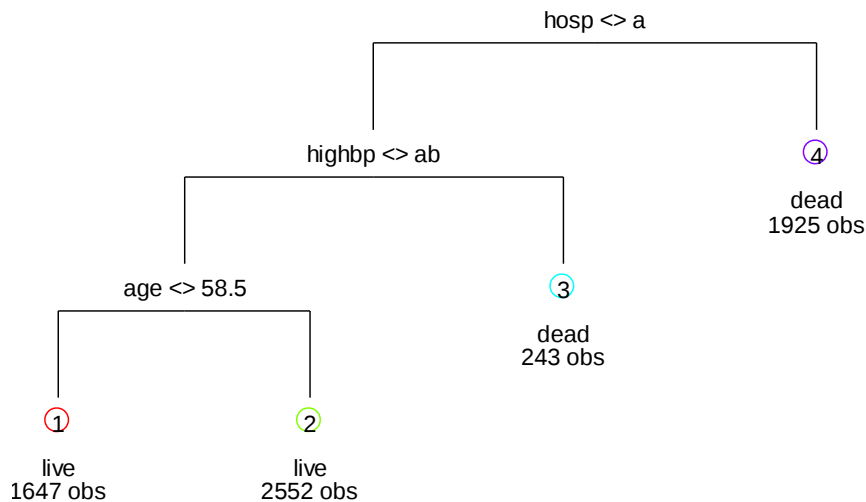
```
8) age < 58.5 1647 1046.00 live ( 0.903461 0.096539 ) *
```

```
9) age > 58.5 2552 2698.00 live ( 0.778605 0.221395 ) *
```

```
5) highbp: nk 243 238.70 dead ( 0.193416 0.806584 ) *
```

```
3) hosp: n 1925 44.78 dead ( 0.001558 0.998442 ) *
```

```
draw.tree(derevce2)
```



Висновок

Виконуючи лабораторну роботу, навчився будувати дерева класифікації та регресійні дерева, які використовуються для створення моделей в системі R. Дана система дозволяє швидко і зручно видаляти зайві вузли з дерев способом відсікання найменш важливих вузлів.