

MACHINE LEARNING - SOLUTION/EXERCISES

Jan-Philipp Kolb

23 Mai, 2019

PART B2

TREES AND BAGGING

EXERCISE - RPART KYPHOSIS

Consider the Kyphosis data frame(`type help("kyphosis")` for more details), that contains:

- ▶ Kyphosis: a factor with levels `absent` `present` indicating if a kyphosis (a type of deformation) was present after the operation.
- ▶ Age: in months.
- ▶ Number: the number of vertebrae involved.
- ▶ Start: the number of the first (topmost) vertebra operated on.

1. Build a tree to classify Kyphosis from Age, Number and Start.

CONSIDER THE TREE BUILD ABOVE.

1. Which variables are used to explain Kyphosis presence?
2. How many observations contain the terminal nodes.

SOLUTION - RPART KYPHOSIS

```
library('rpart')

## Warning: package 'rpart' was built under R version 3.5.3

TREE <- rpart(Kyphosis ~ Age + Number + Start,
               data=kyphosis,method="class")

TREE

## n= 81
##
## node), split, n, loss, yval, (yprob)
##      * denotes terminal node
##
## 1) root 81 17 absent (0.79012346 0.20987654)
##    2) Start>=8.5 62  6 absent (0.90322581 0.09677419)
##      4) Start>=14.5 29  0 absent (1.00000000 0.00000000) *
##      5) Start< 14.5 33  6 absent (0.81818182 0.18181818)
##    10) Age< 55 12  0 absent (1.00000000 0.00000000) *
```