Homework11

Yi Chen

4/10/2020

## Homework 11

In this homework, I will use the sport data.

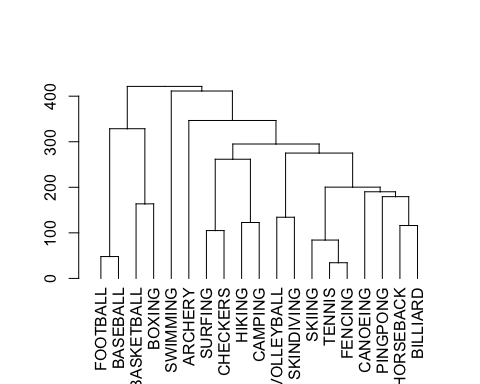
# load the data set  
sport\_name <- c("FOOTBALL","BASEBALL","BASKETBALL","TENNIS","CANOEING","SWIMMING","BOXING","VOLLEYBALL","SKIING","SURFING","FENCING","SKINDIVING","ARCHERY","PINGPONG","HORSEBACK","HIKING","BILLIARD","CAMPING","CHECKERS","JUMPROPE")  
temp <- scan('SPORTS.txt')  
sport <- matrix(NA, 19,19)  
sport[upper.tri(sport, diag = TRUE)] <- temp  
sport <- t(sport)  
colnames(sport) <- sport\_name[2:length(sport\_name)]  
rownames(sport) <- sport\_name[1:(length(sport\_name)-1)]  
sport\_dist <- dist(sport)

#### 1. fit an ultrametric tree to the data using the iterative projection method

library(clue)  
ultra\_sport <- ls\_fit\_ultrametric(sport\_dist,method=c("IP"), weights = 1, control = list())

## Warning in x - d: longer object length is not a multiple of shorter object  
## length

plot(ultra\_sport) # plots the fitted ultrametric tree



## squared distance  
datvec=as.vector(sport\_dist)  
modvec=as.vector(ultra\_sport)  
cor(datvec,modvec)

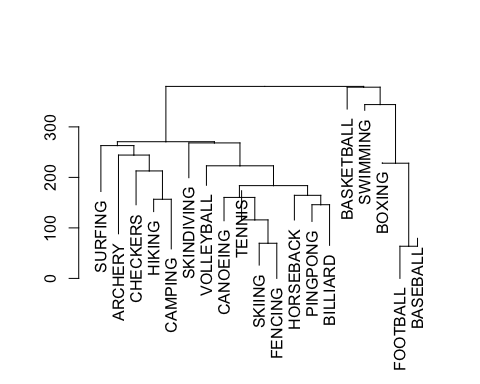
## [1] 0.6517794

#### 2. fit an additive tree to the data using the iterative projection method

ultra\_sport <- ls\_fit\_addtree(sport\_dist,method=c("IP"), weights = 1, control = list())

## Warning in any(diff(weights)): coercing argument of type 'double' to logical

plot(ultra\_sport) # plots the fitted ultrametric tree



## squared distance  
datvec=as.vector(sport\_dist)  
modvec=as.vector(ultra\_sport)  
cor(datvec,modvec)

## [1] 0.9383302

The correlation increase with additive tree. The result of additive tree and ultrametric tree are also different. In ultrametric football and baseball are differnt from other sports. In the additive tree, football, baseball and other sports require speific enviornment (e.g., swiming pool) are categorized together.

#### I cannot install the GTREE in the mac system.