HUDK5124Assignment10

Zhuqian Zhou

2017/4/9

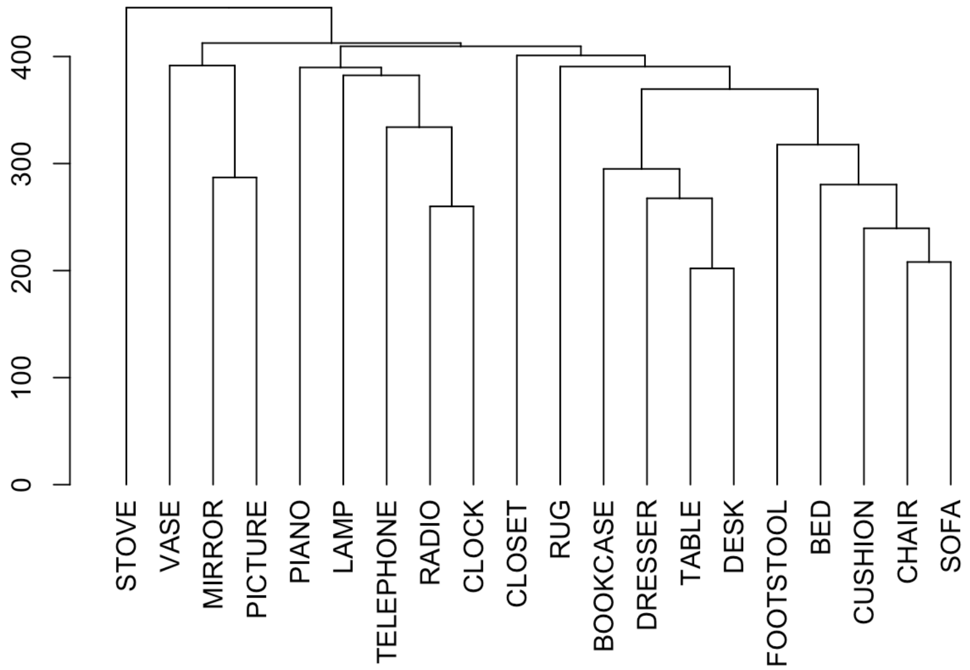
1

For either the sports data or the furniture data (both rated similarity, collected by Rosch et al.) posted in this session, fit an ultrametric tree to the data using the iterative projection method offered in R's "ls_fit_ultrametric" routine (see method below, also see posted help file "R_fitting trees to prox data.txt"). Calculate the (squared) correlation of the model distances to the data, RSQ.

```
library(clue)
f1 <- read.table("FURN_prx_revised.txt", fill = TRUE, header = TR
UE) # add zero on diagonal before read the table
rownames(f1) <- names(f1)
f2 <- as.matrix(f1)
ultraD=ls_fit_ultrametric(f2,method=c("IP"), weights = 1, control
= list()) # fit an ultrametric tree (LS solution, by iterative pr
ojection) to a data set
```

```
## Warning in x - d: 长的对象长度不是短的对象长度的整倍数
```

```
hc <- hclust(ultraD, "ave") # method="average"
dend1 <- as.dendrogram(hc) # "print()" method
plot(dend1)
```




```
# calculate the "cophenetic correlation" - the corr of data & mod
el distances
library(gdata)
```

```
## gdata: read.xls support for 'XLS' (Excel 97-2004) files ENABLE
D.
```

```
##
```

```
## gdata: read.xls support for 'XLSX' (Excel 2007+) files ENABLED
.
```

```
##
## Attaching package: 'gdata'
```

```
## The following object is masked from 'package:stats':
##
##     nobs
```

```
## The following object is masked from 'package:utils':
##
##     object.size
```

```
## The following object is masked from 'package:base':
##
##     startsWith
```

```
datvec = as.vector(lowerTriangle(f2, diag=FALSE)) # Function lowe
rTriangle() is in Package "gdata"
modvec = as.vector(ultraD)
cor(datvec,modvec)
```

```
## [1] 0.8035247
```

The correlation of data distances and model distances of the ultrametic tree is 0.7975155 for the furniture data.

2

Again using the iterative projection method offered in R, fit an additive tree to the same data set (using “ls_fit_addtree”). Calculate the (squared) correlation of the model distances to the data, RSQ.

```
addD=ls_fit_addtree(f2,method=c("IP"), weights = 1, control = lis
t()) # fit an additive tree (LS solution, by "iterative projectio
n" method to the data set
```

```
## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值
```

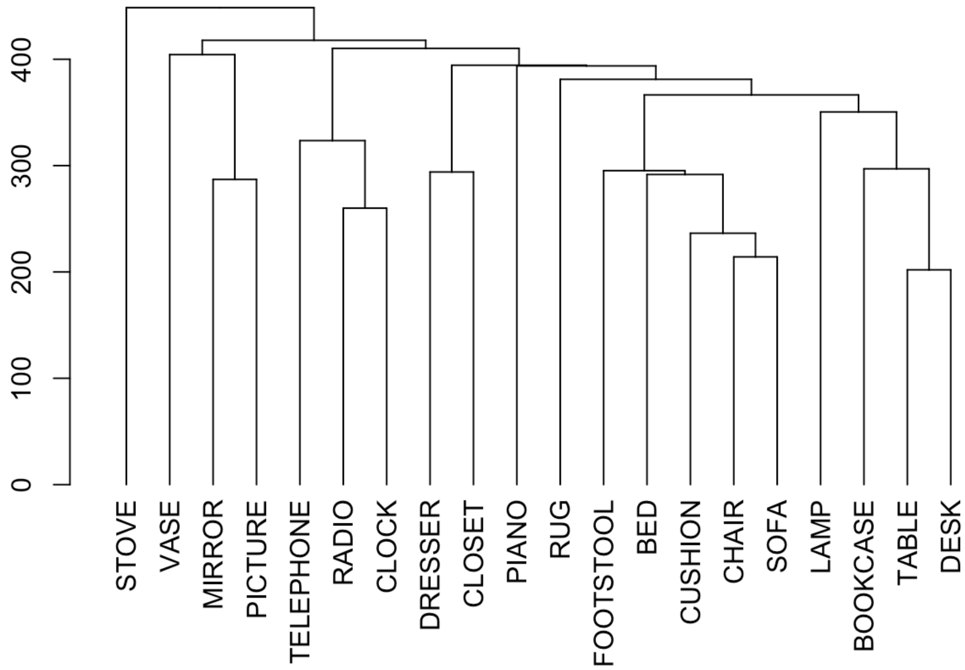
```
addD
```

```
## Dissimilarities using Additive tree distances:
##              CHAIR     SOFA    TABLE  DRESSER     DESK       BE
D    STOVE
## SOFA      214.1667
## TABLE     326.3816 341.0777
## DRESSER   356.6694 371.3655 318.4558
```

```
## DESK       325.6594 340.3554 202.0000 317.7336
## BED        272.6413 287.3374 332.1855 362.4734 331.4633
## STOVE      439.7254 454.4215 401.5118 406.0981 400.7896 445.529
4
## BOOKCASE   365.8734 380.5695 297.3611 357.9477 296.6389 371.677
4 441.0037
## FOOTSTOOL 274.5898 289.2859 353.9292 384.2170 353.2070 300.188
9 467.2730
## LAMP       375.7569 390.4529 337.5433 338.7913 336.8210 381.560
8 425.1856
## PIANO      389.6795 404.3756 351.4659 360.3503 350.7436 395.483
4 443.4063
## MIRROR     412.0675 426.7636 373.8539 378.4402 373.1317 417.871
4 461.1259
## RUG        360.0412 374.7373 362.8570 393.1448 362.1348 365.845
2 476.2009
## RADIO      404.8277 419.5237 366.6140 375.4985 365.8918 410.631
6 458.5545
## CLOCK      404.3277 419.0237 366.1140 374.9985 365.3918 410.131
6 458.0545
## PICTURE    414.4008 429.0969 376.1872 380.7735 375.4650 420.204
8 463.4592
## VASE       407.1538 421.8499 368.9402 373.5265 368.2180 412.957
8 456.2122
## TELEPHONE 412.8129 427.5090 374.5993 383.4838 373.8771 418.616
9 466.5398
## CLOSET     425.3916 440.0877 387.1780 294.0000 386.4558 431.195
6 474.8203
## CUSHION    241.8333 231.0000 368.7443 399.0322 368.0221 315.004
1 482.0882
##            BOOKCASE FOOTSTOOL     LAMP    PIANO   MIRROR       R
UG    RADIO
## SOFA
## TABLE
## DRESSER
## DESK
## BED
## STOVE
## BOOKCASE
## FOOTSTOOL 393.4210
## LAMP       377.0351  403.3045
## PIANO      390.9577  417.2271 379.4377
## MIRROR     413.3457  439.6151 397.5276 415.7483
## RUG        402.3488  387.5888 412.2323 426.1549 448.5429
## RADIO      406.1059  432.3752 394.5859 391.8178 430.8965 441.30
```

```
31
## CLOCK     405.6059  431.8752 394.0859 391.3178 430.3965 440.80
31 260.0000
## PICTURE   415.6790  441.9484 399.8609 418.0817 287.0000 450.87
62 433.2299
## VASE      408.4321  434.7014 392.6139 410.8347 403.1988 443.62
92 425.9829
## TELEPHONE 414.0912  440.3605 402.5712 399.8030 438.8818 449.28
84 323.7500
## CLOSET    426.6699  452.9392 407.5136 429.0725 447.1624 461.86
71 444.2207
## CUSHION   408.2362  316.9526 418.1196 432.0422 454.4302 402.40
40 447.1904
##              CLOCK  PICTURE     VASE TELEPHONE   CLOSET
## SOFA
## TABLE
## DRESSER
## DESK
## BED
## STOVE
## BOOKCASE
## FOOTSTOOL
## LAMP
## PIANO
## MIRROR
## RUG
## RADIO
## CLOCK
## PICTURE   432.7299
## VASE      425.4829 405.5322
## TELEPHONE 323.2500 441.2152 433.9682
## CLOSET    443.7207 449.4957 442.2487  452.2060
## CUSHION   446.6904 456.7636 449.5166  455.1757 467.7544
```

```
hc2 <- hclust(addD, "ave")
dend2 <- as.dendrogram(hc2)
plot(dend2)
```




```
modvec2 = as.vector(addD)
cor(datvec,modvec2) # The correlation of data distances and model
distances of the additive tree is 0.9067304 for the furniture dat
a at this run.
```

```
## [1] 0.9012762
```

```
# However, IP uses a semi-randomized start. It is susceptible to
local minima, so run multiple starts.
corvector <- c(rep(0,25))
for (i in 1:25)
{addD3=ls_fit_addtree(f2,method=c("IP"), weights = 1, control = l
ist())
modvec3=as.vector(addD3)
corvector[i]=cor(datvec,modvec3) # calculate linear fit = corr of
data & model distances
}
```

```
## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值
```

```
## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值
```

```
## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值

## Warning in any(diff(weights)): 将种类为'double'的参数强迫转化为逻辑
值
```

```
corvector # the vector of fits shows that local minima are very c
ommon (cf. Smith, 1998)
```

```
##  [1] 0.9086624 0.9030464 0.9105745 0.9065246 0.9040460 0.90290
98 0.9004377
##  [8] 0.9092728 0.9063303 0.9093261 0.9091252 0.8974086 0.90649
95 0.9034255
## [15] 0.9052995 0.9041801 0.9001116 0.9072837 0.9011703 0.89757
60 0.9047619
## [22] 0.9054518 0.9087903 0.9036018 0.8991714
```

3

Also fit an additive tree to the data using GTREE. Compare the fit of this solution to the results using the R methods. Comment. Although I cannot run the executable file “gtree.exe” anyhow, I will set down the furthest step I am currently able to get. 1) Download Pascal Compiler for Mac: go to Website free pascal (http://www.freepascal.org/down/i386/macosx-hungary.var) and download fpc-3.0.2.intel-macosx.dmg (ftp://ftp.hu.freepascal.org/pub/fpc/dist/3.0.2/i386-macosx/fpc-3.0.2.intel-macosx.dmg) 2) Install the Mac OS X installer package named “fpc-3.0.2.intel-macosx.pkg” 3) Download the file gtree.pas (http://netlib.sandia.gov/mds/gtree.pas) 4) Open “Terminal” on Mac and Type the following lines after the dollar sign: $ *cd* address_of_the_*repository*_where_I_saved_gtree.pas $ *fpc* gtree.pas 5) Open the folder where I saved “gtree.pas”, I saw a Unix executable file named “gtree” 6) Double click that file and start fitting the additive tree. (But it always shows error 2 (fild not found). That’s where I cannot get anything further.)

References Smith, T. J. (1998). A comparison of three additive tree algorithms that rely on a leastsquares