> # use those packages

> library(ggplot2)

> library(lattice)

> library(caret)

> library(rpart.plot)

> library(mboost)

Loading required package: parallel

Loading required package: stabs

This is mboost 2.9-2. See ‘package?mboost’ and ‘news(package = "mboost")’

for a complete list of changes.

Attaching package: ‘mboost’

The following object is masked from ‘package:ggplot2’:

%+%

> data("bodyfat", package="TH.data")

> attributes(bodyfat)

$names

[1] "age" "DEXfat" "waistcirc" "hipcirc" "elbowbreadth" "kneebreadth"

[7] "anthro3a" "anthro3b" "anthro3c" "anthro4"

$row.names

[1] "47" "48" "49" "50" "51" "52" "53" "54" "55" "56" "57" "58" "59" "60" "61"

[16] "62" "63" "64" "65" "66" "67" "68" "69" "70" "71" "72" "73" "74" "75" "76"

[31] "77" "78" "79" "80" "81" "82" "83" "84" "85" "86" "87" "88" "89" "90" "91"

[46] "92" "93" "94" "95" "96" "97" "98" "99" "100" "101" "102" "103" "104" "105" "106"

[61] "107" "108" "109" "110" "111" "112" "113" "114" "115" "116" "117"

$class

[1] "data.frame"

> str(bodyfat)

'data.frame': 71 obs. of 10 variables:

$ age : num 57 65 59 58 60 61 56 60 58 62 ...

$ DEXfat : num 41.7 43.3 35.4 22.8 36.4 ...

$ waistcirc : num 100 99.5 96 72 89.5 83.5 81 89 80 79 ...

$ hipcirc : num 112 116.5 108.5 96.5 100.5 ...

$ elbowbreadth: num 7.1 6.5 6.2 6.1 7.1 6.5 6.9 6.2 6.4 7 ...

$ kneebreadth : num 9.4 8.9 8.9 9.2 10 8.8 8.9 8.5 8.8 8.8 ...

$ anthro3a : num 4.42 4.63 4.12 4.03 4.24 3.55 4.14 4.04 3.91 3.66 ...

$ anthro3b : num 4.95 5.01 4.74 4.48 4.68 4.06 4.52 4.7 4.32 4.21 ...

$ anthro3c : num 4.5 4.48 4.6 3.91 4.15 3.64 4.31 4.47 3.47 3.6 ...

$ anthro4 : num 6.13 6.37 5.82 5.66 5.91 5.14 5.69 5.7 5.49 5.25 ...

> head(bodyfat)

age DEXfat waistcirc hipcirc elbowbreadth kneebreadth anthro3a anthro3b anthro3c anthro4

47 57 41.68 100.0 112.0 7.1 9.4 4.42 4.95 4.50 6.13

48 65 43.29 99.5 116.5 6.5 8.9 4.63 5.01 4.48 6.37

49 59 35.41 96.0 108.5 6.2 8.9 4.12 4.74 4.60 5.82

50 58 22.79 72.0 96.5 6.1 9.2 4.03 4.48 3.91 5.66

51 60 36.42 89.5 100.5 7.1 10.0 4.24 4.68 4.15 5.91

52 61 24.13 83.5 97.0 6.5 8.8 3.55 4.06 3.64 5.14

> # For checking the dimensions of our training data frame and testing data frame

> dim(bodyfat)

[1] 71 10

> attributes(bodyfat)

$names

[1] "age" "DEXfat" "waistcirc" "hipcirc" "elbowbreadth" "kneebreadth"

[7] "anthro3a" "anthro3b" "anthro3c" "anthro4"

$row.names

[1] "47" "48" "49" "50" "51" "52" "53" "54" "55" "56" "57" "58" "59" "60" "61"

[16] "62" "63" "64" "65" "66" "67" "68" "69" "70" "71" "72" "73" "74" "75" "76"

[31] "77" "78" "79" "80" "81" "82" "83" "84" "85" "86" "87" "88" "89" "90" "91"

[46] "92" "93" "94" "95" "96" "97" "98" "99" "100" "101" "102" "103" "104" "105" "106"

[61] "107" "108" "109" "110" "111" "112" "113" "114" "115" "116" "117"

$class

[1] "data.frame"

> bodyfat[1:5,]

age DEXfat waistcirc hipcirc elbowbreadth kneebreadth anthro3a anthro3b anthro3c anthro4

47 57 41.68 100.0 112.0 7.1 9.4 4.42 4.95 4.50 6.13

48 65 43.29 99.5 116.5 6.5 8.9 4.63 5.01 4.48 6.37

49 59 35.41 96.0 108.5 6.2 8.9 4.12 4.74 4.60 5.82

50 58 22.79 72.0 96.5 6.1 9.2 4.03 4.48 3.91 5.66

51 60 36.42 89.5 100.5 7.1 10.0 4.24 4.68 4.15 5.91

> set.seed(1234)

> ind <- sample(2, nrow(bodyfat), replace=TRUE, prob=c(0.7, 0.3))

> bodyfat.train <- bodyfat[ind==1,]

> bodyfat.test <- bodyfat[ind==2,]

> library(rpart)

> myFormula <- DEXfat ~ age + waistcirc + hipcirc + elbowbreadth + kneebreadth

> bodyfat\_rpart <- rpart(myFormula, data = bodyfat.train,

+ control = rpart.control(minsplit = 10))

> attributes(bodyfat\_rpart)

$names

[1] "frame" "where" "call" "terms"

[5] "cptable" "method" "parms" "control"

[9] "functions" "numresp" "splits" "variable.importance"

[13] "y" "ordered"

$xlevels

named list()

$class

[1] "rpart"

> print(bodyfat\_rpart$cptable)

CP nsplit rel error xerror xstd

1 0.67272638 0 1.00000000 1.0427457 0.19016187

2 0.09390665 1 0.32727362 0.5081173 0.11702581

3 0.06037503 2 0.23336696 0.4522296 0.09801847

4 0.03420446 3 0.17299193 0.3967005 0.09676249

5 0.01708278 4 0.13878747 0.3015476 0.07385485

6 0.01695763 5 0.12170469 0.2929969 0.06850104

7 0.01007079 6 0.10474706 0.2713231 0.06690466

8 0.01000000 7 0.09467627 0.2713231 0.06690466

> print(bodyfat\_rpart)

n= 56

node), split, n, deviance, yval

\* denotes terminal node

1) root 56 7265.0290000 30.94589

2) waistcirc< 88.4 31 960.5381000 22.55645

4) hipcirc< 96.25 14 222.2648000 18.41143

8) age< 60.5 9 66.8809600 16.19222 \*

9) age>=60.5 5 31.2769200 22.40600 \*

5) hipcirc>=96.25 17 299.6470000 25.97000

10) waistcirc< 77.75 6 30.7345500 22.32500 \*

11) waistcirc>=77.75 11 145.7148000 27.95818

22) hipcirc< 99.5 3 0.2568667 23.74667 \*

23) hipcirc>=99.5 8 72.2933500 29.53750 \*

3) waistcirc>=88.4 25 1417.1140000 41.34880

6) waistcirc< 104.75 18 330.5792000 38.09111

12) hipcirc< 109.9 9 68.9996200 34.37556 \*

13) hipcirc>=109.9 9 13.0832000 41.80667 \*

7) waistcirc>=104.75 7 404.3004000 49.72571 \*

> plot(bodyfat\_rpart)

> text(bodyfat\_rpart, use.n=T)

> opt <- which.min(bodyfat\_rpart$cptable[,"xerror"])

> cp <- bodyfat\_rpart$cptable[opt, "CP"]

> bodyfat\_prune <- prune(bodyfat\_rpart, cp = cp)

> print(bodyfat\_prune)

n= 56

node), split, n, deviance, yval

\* denotes terminal node

1) root 56 7265.02900 30.94589

2) waistcirc< 88.4 31 960.53810 22.55645

4) hipcirc< 96.25 14 222.26480 18.41143

8) age< 60.5 9 66.88096 16.19222 \*

9) age>=60.5 5 31.27692 22.40600 \*

5) hipcirc>=96.25 17 299.64700 25.97000

10) waistcirc< 77.75 6 30.73455 22.32500 \*

11) waistcirc>=77.75 11 145.71480 27.95818 \*

3) waistcirc>=88.4 25 1417.11400 41.34880

6) waistcirc< 104.75 18 330.57920 38.09111

12) hipcirc< 109.9 9 68.99962 34.37556 \*

13) hipcirc>=109.9 9 13.08320 41.80667 \*

7) waistcirc>=104.75 7 404.30040 49.72571 \*

> plot(bodyfat\_prune)

> text(bodyfat\_prune, use.n=T)

> DEXfat\_pred <- predict(bodyfat\_prune, newdata=bodyfat.test)

> xlim <- range(bodyfat$DEXfat)

> plot(DEXfat\_pred ~ DEXfat, data=bodyfat.test, xlab="Observed",

+ ylab="Predicted", ylim=xlim, xlim=xlim)

> abline(a=0, b=1)



