> rm(list=ls())

> A = read.csv("challenge.csv", header=T)

> summary(A)

response P1 P2 P3 P4 P5

A:50 Min. :1030 Min. :1121 Min. : 665 Min. :172.0 Min. :434.0

B:50 1st Qu.:1380 1st Qu.:1465 1st Qu.: 942 1st Qu.:233.8 1st Qu.:602.5

C:48 Median :1486 Median :1566 Median :1016 Median :256.0 Median :669.0

Mean :1490 Mean :1558 Mean :1021 Mean :256.9 Mean :662.9

3rd Qu.:1592 3rd Qu.:1661 3rd Qu.:1106 3rd Qu.:282.0 3rd Qu.:716.5

Max. :1893 Max. :1945 Max. :1315 Max. :332.0 Max. :893.0

NA's :1 NA's :2 NA's :1 NA's :24 NA's :1

P6 P7 P8 P9 P10

Min. :141.0 Min. :121.0 Min. :303.0 Min. : 640.0 Min. :190.0

1st Qu.:214.8 1st Qu.:161.5 1st Qu.:411.0 1st Qu.: 824.5 1st Qu.:230.0

Median :233.5 Median :179.0 Median :440.0 Median : 879.0 Median :239.0

Mean :232.8 Mean :179.7 Mean :441.4 Mean : 876.6 Mean :239.4

3rd Qu.:251.5 3rd Qu.:192.0 3rd Qu.:470.2 3rd Qu.: 926.0 3rd Qu.:249.0

Max. :308.0 Max. :299.0 Max. :547.0 Max. :1090.0 Max. :290.0

NA's :1 NA's :1

P11 P12 P13 P14 P15

Min. :173.0 Min. :435.0 Min. : 13.0 Min. : 60.00 Min. : 856

1st Qu.:247.0 1st Qu.:611.0 1st Qu.:100.8 1st Qu.: 84.75 1st Qu.:1155

Median :268.0 Median :650.0 Median :125.0 Median : 94.50 Median :1242

Mean :271.8 Mean :650.9 Mean :123.5 Mean : 94.51 Mean :1247

3rd Qu.:294.0 3rd Qu.:698.0 3rd Qu.:151.0 3rd Qu.:104.00 3rd Qu.:1347

Max. :371.0 Max. :798.0 Max. :216.0 Max. :137.00 Max. :1568

NA's :3 NA's :11 NA's :12

P16 P17 P18

Min. :101 Min. :132.0 Min. :473.0

1st Qu.:130 1st Qu.:181.8 1st Qu.:648.8

Median :138 Median :194.5 Median :700.0

Mean :139 Mean :195.9 Mean :698.8

3rd Qu.:148 3rd Qu.:210.2 3rd Qu.:751.5

Max. :169 Max. :271.0 Max. :880.0

> sapply(A, function(x) sum(is.na(x)))

response P1 P2 P3 P4 P5 P6 P7 P8

0 1 2 1 24 1 0 1 0

P9 P10 P11 P12 P13 P14 P15 P16 P17

1 0 3 11 0 0 12 0 0

P18

0

> sum(complete.cases(A))

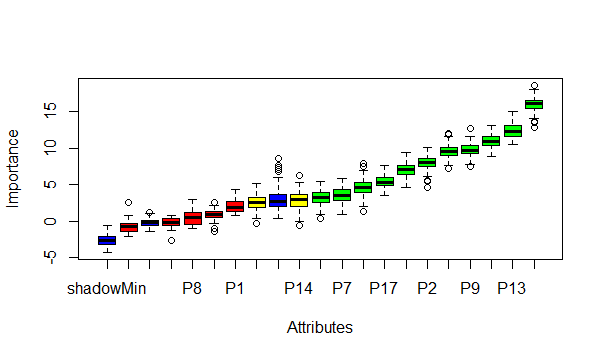
[1] 101

> library(Boruta)

> compData<-A[complete.cases(A),]

> b<-Boruta(x=compData[,2:ncol(A)],y=compData[,1],maxRuns=200)

> plot(b)



> print(b)

Boruta performed 199 iterations in 22.31646 secs.

11 attributes confirmed important: P12, P13, P16, P17, P18 and 6 more.

5 attributes confirmed unimportant: P1, P10, P11, P3, P8.

2 tentative attributes left: P14, P15.

> impVar<-getSelectedAttributes(b,withTentative=F)

> impVar

[1] "P2" "P4" "P5" "P6" "P7" "P9" "P12" "P13" "P16" "P17" "P18"

> library(randomForest)

> newData<-A[complete.cases(A[,impVar]),]

> fm<-as.formula(paste("response",paste(impVar,collapse="+"),sep="~"))

> rf<-randomForest(fm,newData,ntree=500)

> print(rf)

Call:

randomForest(formula = fm, data = newData, ntree = 500)

Type of random forest: classification

Number of trees: 500

No. of variables tried at each split: 3

OOB estimate of error rate: 18.42%

Confusion matrix:

A B C class.error

A 45 0 0 0.0000000

B 2 29 8 0.2564103

C 3 8 19 0.3666667

> library(mice)

> input<-c(1115, NA, 748, 182, NA, NA, 178, 311, 756, 226, NA, NA, NA, 48, 1009, NA, 204, 593)

> B<-rbind(A[,2:ncol(A)],input)

> tempData <- mice(B,m=5,maxit=50,meth='pmm',seed=500)

Not shown because it is too long

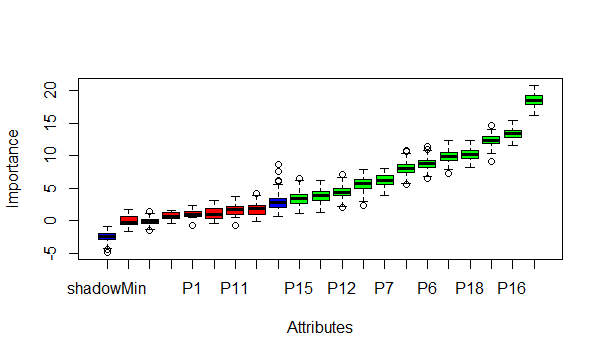
> completedData <- complete(tempData,1)

> A[,2:ncol(A)]<-completedData[1:(nrow(completedData)-1), ]

> compData<-A[complete.cases(A),]

> b<-Boruta(x=compData[,2:ncol(A)],y=compData[,1],maxRuns=200)

> plot(b)



> print(b)

Boruta performed 158 iterations in 25.72804 secs.

12 attributes confirmed important: P12, P13, P15, P16, P17 and 7 more.

6 attributes confirmed unimportant: P1, P10, P11, P14, P3 and 1 more.

> impVar<-getSelectedAttributes(b,withTentative=F)

> impVar

[1] "P2" "P4" "P5" "P6" "P7" "P9" "P12" "P13" "P15" "P16" "P17" "P18"

> newData<-A[complete.cases(A[,impVar]),]

> fm<-as.formula(paste("response",paste(impVar,collapse="+"),sep="~"))

> rf<-randomForest(fm,newData,ntree=500)

> print(rf)

Call:

randomForest(formula = fm, data = newData, ntree = 500)

Type of random forest: classification

Number of trees: 500

No. of variables tried at each split: 3

OOB estimate of error rate: 33.78%

Confusion matrix:

A B C class.error

A 47 0 3 0.0600000

B 2 28 20 0.4400000

C 3 22 23 0.5208333

> idx<-which(names(A)%in%impVar)

> inputData<-input[idx]

> predData<-completedData[nrow(completedData),impVar,drop=F]

> result<-predict(rf,newdata=predData)

> result

149

C

Levels: A B C