project.R

duter

2021-03-02

##### Dependencies #####  
  
if (!require(tree))  
 install.packages('tree')

## Loading required package: tree

## Warning: package 'tree' was built under R version 4.0.4

library(tree)  
  
if (!require(randomForest))  
 install.packages("randomForest")

## Loading required package: randomForest

## Warning: package 'randomForest' was built under R version 4.0.4

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

library(randomForest)  
  
##### Pre-process Data #####  
  
data <-  
 read.csv('ObesityDataSet.csv',  
 na.strings = '?',  
 stringsAsFactors = T)  
data <- na.omit(data)  
  
# rename columns   
colnames(data)[6] <- 'eats\_high\_calor\_food'  
colnames(data)[7] <- 'eats\_veggies'  
colnames(data)[8] <- 'num\_meals'  
colnames(data)[9] <- 'eats\_snacks'  
colnames(data)[11] <- 'drinks\_water'  
colnames(data)[12] <- 'counts\_calories'  
colnames(data)[13] <- 'exercises\_often'  
colnames(data)[14] <- 'time\_using\_tech'  
colnames(data)[15] <- 'drinks\_alcohol'  
colnames(data)[16] <- 'method\_trans'  
colnames(data)[17] <- 'weight\_category'  
names(data)

## [1] "Gender" "Age"   
## [3] "Height" "Weight"   
## [5] "family\_history\_with\_overweight" "eats\_high\_calor\_food"   
## [7] "eats\_veggies" "num\_meals"   
## [9] "eats\_snacks" "SMOKE"   
## [11] "drinks\_water" "counts\_calories"   
## [13] "exercises\_often" "time\_using\_tech"   
## [15] "drinks\_alcohol" "method\_trans"   
## [17] "weight\_category"

head(data)

## Gender Age Height Weight family\_history\_with\_overweight eats\_high\_calor\_food  
## 1 Female 21 1.62 64.0 yes no  
## 2 Female 21 1.52 56.0 yes no  
## 3 Male 23 1.80 77.0 yes no  
## 4 Male 27 1.80 87.0 no no  
## 5 Male 22 1.78 89.8 no no  
## 6 Male 29 1.62 53.0 no yes  
## eats\_veggies num\_meals eats\_snacks SMOKE drinks\_water counts\_calories  
## 1 2 3 Sometimes no 2 no  
## 2 3 3 Sometimes yes 3 yes  
## 3 2 3 Sometimes no 2 no  
## 4 3 3 Sometimes no 2 no  
## 5 2 1 Sometimes no 2 no  
## 6 2 3 Sometimes no 2 no  
## exercises\_often time\_using\_tech drinks\_alcohol method\_trans  
## 1 0 1 no Public\_Transportation  
## 2 3 0 Sometimes Public\_Transportation  
## 3 2 1 Frequently Public\_Transportation  
## 4 2 0 Frequently Walking  
## 5 0 0 Sometimes Public\_Transportation  
## 6 0 0 Sometimes Automobile  
## weight\_category  
## 1 Normal\_Weight  
## 2 Normal\_Weight  
## 3 Normal\_Weight  
## 4 Overweight\_Level\_I  
## 5 Overweight\_Level\_II  
## 6 Normal\_Weight

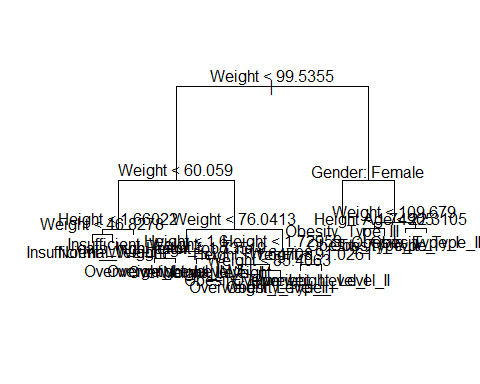
set.seed(1)  
  
# split sample and test  
train <- sample(1:nrow(data), nrow(data) / 2)  
test <- data[-train, ]  
  
##### Decision Tree #####  
  
# plant a tree   
data.tree <- tree(weight\_category ~ ., data = data)  
data.tree

## node), split, n, deviance, yval, (yprob)  
## \* denotes terminal node  
##   
## 1) root 2111 8202.000 Obesity\_Type\_I ( 0.128849 0.135955 0.166272 0.140692 0.153482 0.137376 0.137376 )   
## 2) Weight < 99.5355 1382 4493.000 Overweight\_Level\_I ( 0.196816 0.207670 0.175832 0.003618 0.000000 0.209841 0.206223 )   
## 4) Weight < 60.059 420 638.000 Insufficient\_Weight ( 0.633333 0.340476 0.000000 0.000000 0.000000 0.023810 0.002381 )   
## 8) Height < 1.66022 223 384.000 Normal\_Weight ( 0.408072 0.542601 0.000000 0.000000 0.000000 0.044843 0.004484 )   
## 16) Weight < 46.8278 92 54.360 Insufficient\_Weight ( 0.913043 0.086957 0.000000 0.000000 0.000000 0.000000 0.000000 ) \*  
## 17) Weight > 46.8278 131 135.600 Normal\_Weight ( 0.053435 0.862595 0.000000 0.000000 0.000000 0.076336 0.007634 ) \*  
## 9) Height > 1.66022 197 137.900 Insufficient\_Weight ( 0.888325 0.111675 0.000000 0.000000 0.000000 0.000000 0.000000 ) \*  
## 5) Weight > 60.059 962 2713.000 Overweight\_Level\_II ( 0.006237 0.149688 0.252599 0.005198 0.000000 0.291060 0.295218 )   
## 10) Weight < 76.0413 356 778.600 Overweight\_Level\_I ( 0.016854 0.353933 0.011236 0.000000 0.000000 0.494382 0.123596 )   
## 20) Height < 1.6 72 123.700 Overweight\_Level\_II ( 0.000000 0.000000 0.055556 0.000000 0.000000 0.402778 0.541667 )   
## 40) eats\_high\_calor\_food: no 40 9.353 Overweight\_Level\_II ( 0.000000 0.000000 0.000000 0.000000 0.000000 0.025000 0.975000 ) \*  
## 41) eats\_high\_calor\_food: yes 32 24.110 Overweight\_Level\_I ( 0.000000 0.000000 0.125000 0.000000 0.000000 0.875000 0.000000 ) \*  
## 21) Height > 1.6 284 485.100 Overweight\_Level\_I ( 0.021127 0.443662 0.000000 0.000000 0.000000 0.517606 0.017606 )   
## 42) Height < 1.73519 219 316.800 Overweight\_Level\_I ( 0.000000 0.315068 0.000000 0.000000 0.000000 0.662100 0.022831 ) \*  
## 43) Height > 1.73519 65 57.490 Normal\_Weight ( 0.092308 0.876923 0.000000 0.000000 0.000000 0.030769 0.000000 ) \*  
## 11) Weight > 76.0413 606 1431.000 Overweight\_Level\_II ( 0.000000 0.029703 0.394389 0.008251 0.000000 0.171617 0.396040 )   
## 22) Height < 1.72959 360 546.300 Obesity\_Type\_I ( 0.000000 0.000000 0.577778 0.013889 0.000000 0.002778 0.405556 )   
## 44) Height < 1.64708 126 63.630 Obesity\_Type\_I ( 0.000000 0.000000 0.944444 0.031746 0.000000 0.000000 0.023810 ) \*  
## 45) Height > 1.64708 234 334.700 Overweight\_Level\_II ( 0.000000 0.000000 0.380342 0.004274 0.000000 0.004274 0.611111 )   
## 90) Weight < 85.4063 148 90.180 Overweight\_Level\_II ( 0.000000 0.000000 0.074324 0.000000 0.000000 0.006757 0.918919 ) \*  
## 91) Weight > 85.4063 86 59.260 Obesity\_Type\_I ( 0.000000 0.000000 0.906977 0.011628 0.000000 0.000000 0.081395 ) \*  
## 23) Height > 1.72959 246 582.800 Overweight\_Level\_I ( 0.000000 0.073171 0.126016 0.000000 0.000000 0.418699 0.382114 )   
## 46) Weight < 91.0261 182 333.900 Overweight\_Level\_I ( 0.000000 0.098901 0.000000 0.000000 0.000000 0.565934 0.335165 ) \*  
## 47) Weight > 91.0261 64 88.660 Overweight\_Level\_II ( 0.000000 0.000000 0.484375 0.000000 0.000000 0.000000 0.515625 ) \*  
## 3) Weight > 99.5355 729 1522.000 Obesity\_Type\_III ( 0.000000 0.000000 0.148148 0.400549 0.444444 0.000000 0.006859 )   
## 6) Gender: Female 324 13.560 Obesity\_Type\_III ( 0.000000 0.000000 0.000000 0.003086 0.996914 0.000000 0.000000 ) \*  
## 7) Gender: Male 405 533.800 Obesity\_Type\_II ( 0.000000 0.000000 0.266667 0.718519 0.002469 0.000000 0.012346 )   
## 14) Weight < 109.679 153 241.300 Obesity\_Type\_I ( 0.000000 0.000000 0.594771 0.372549 0.000000 0.000000 0.032680 )   
## 28) Height < 1.74905 63 39.630 Obesity\_Type\_II ( 0.000000 0.000000 0.095238 0.904762 0.000000 0.000000 0.000000 ) \*  
## 29) Height > 1.74905 90 38.620 Obesity\_Type\_I ( 0.000000 0.000000 0.944444 0.000000 0.000000 0.000000 0.055556 ) \*  
## 15) Weight > 109.679 252 137.400 Obesity\_Type\_II ( 0.000000 0.000000 0.067460 0.928571 0.003968 0.000000 0.000000 )   
## 30) Age < 22.3105 25 38.950 Obesity\_Type\_I ( 0.000000 0.000000 0.640000 0.320000 0.040000 0.000000 0.000000 ) \*  
## 31) Age > 22.3105 227 12.850 Obesity\_Type\_II ( 0.000000 0.000000 0.004405 0.995595 0.000000 0.000000 0.000000 ) \*

summary(data.tree)

##   
## Classification tree:  
## tree(formula = weight\_category ~ ., data = data)  
## Variables actually used in tree construction:  
## [1] "Weight" "Height" "eats\_high\_calor\_food"  
## [4] "Gender" "Age"   
## Number of terminal nodes: 17   
## Residual mean deviance: 0.7234 = 1515 / 2094   
## Misclassification error rate: 0.1393 = 294 / 2111

plot(data.tree)  
text(data.tree, pretty = 0)



data.tree <- tree(weight\_category ~ ., data = data, subset = train)  
data.tree.pred = predict(data.tree, test, type = "class")  
  
table(data.tree.pred, test$weight\_category)

##   
## data.tree.pred Insufficient\_Weight Normal\_Weight Obesity\_Type\_I  
## Insufficient\_Weight 138 21 0  
## Normal\_Weight 3 65 0  
## Obesity\_Type\_I 0 0 134  
## Obesity\_Type\_II 0 0 15  
## Obesity\_Type\_III 0 0 0  
## Overweight\_Level\_I 0 44 0  
## Overweight\_Level\_II 0 0 28  
##   
## data.tree.pred Obesity\_Type\_II Obesity\_Type\_III Overweight\_Level\_I  
## Insufficient\_Weight 0 0 0  
## Normal\_Weight 0 0 5  
## Obesity\_Type\_I 8 0 0  
## Obesity\_Type\_II 134 1 0  
## Obesity\_Type\_III 1 176 0  
## Overweight\_Level\_I 0 0 121  
## Overweight\_Level\_II 0 0 17  
##   
## data.tree.pred Overweight\_Level\_II  
## Insufficient\_Weight 0  
## Normal\_Weight 1  
## Obesity\_Type\_I 9  
## Obesity\_Type\_II 0  
## Obesity\_Type\_III 0  
## Overweight\_Level\_I 11  
## Overweight\_Level\_II 124

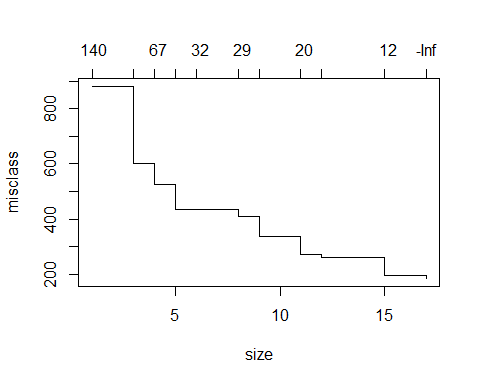
mean(data.tree.pred == test$weight\_category)

## [1] 0.844697

cv.data <- cv.tree(data.tree, FUN = prune.misclass)  
names(cv.data)

## [1] "size" "dev" "k" "method"

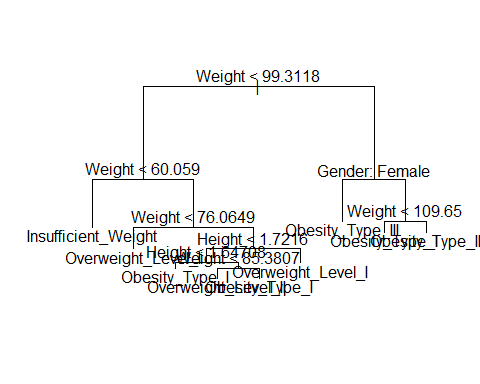
plot(cv.data)



cv.data

## $size  
## [1] 17 15 12 11 9 8 6 5 4 3 1  
##   
## $dev  
## [1] 185 196 262 273 337 408 433 433 527 601 881  
##   
## $k  
## [1] -Inf 11.5 19.0 20.0 23.5 29.0 32.5 35.0 67.0 111.0 141.0  
##   
## $method  
## [1] "misclass"  
##   
## attr(,"class")  
## [1] "prune" "tree.sequence"

prune.data <- prune.misclass(data.tree, best = 9)  
plot(prune.data)  
text(prune.data, pretty = 0)



prune.data.pred <- predict(prune.data, test, type = "class")  
table(prune.data.pred, test$weight\_category)

##   
## prune.data.pred Insufficient\_Weight Normal\_Weight Obesity\_Type\_I  
## Insufficient\_Weight 139 63 0  
## Normal\_Weight 0 0 0  
## Obesity\_Type\_I 0 0 135  
## Obesity\_Type\_II 0 0 14  
## Obesity\_Type\_III 0 0 0  
## Overweight\_Level\_I 2 67 22  
## Overweight\_Level\_II 0 0 6  
##   
## prune.data.pred Obesity\_Type\_II Obesity\_Type\_III Overweight\_Level\_I  
## Insufficient\_Weight 0 0 4  
## Normal\_Weight 0 0 0  
## Obesity\_Type\_I 38 0 0  
## Obesity\_Type\_II 104 1 0  
## Obesity\_Type\_III 1 176 0  
## Overweight\_Level\_I 0 0 139  
## Overweight\_Level\_II 0 0 0  
##   
## prune.data.pred Overweight\_Level\_II  
## Insufficient\_Weight 1  
## Normal\_Weight 0  
## Obesity\_Type\_I 9  
## Obesity\_Type\_II 0  
## Obesity\_Type\_III 0  
## Overweight\_Level\_I 74  
## Overweight\_Level\_II 61

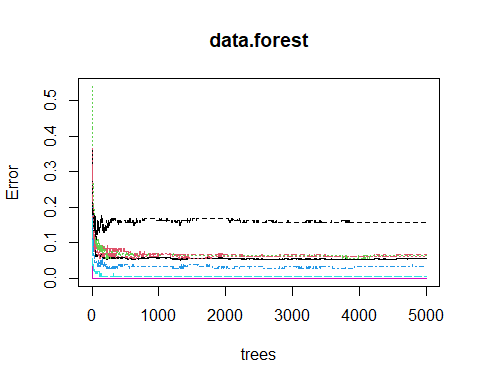
mean(prune.data.pred == test$weight\_category)

## [1] 0.7140152

data.forest <- randomForest(  
 weight\_category ~ .,  
 data = data,  
 subset = train,  
 ntree = 5000,  
 importance = T  
)  
data.forest

##   
## Call:  
## randomForest(formula = weight\_category ~ ., data = data, ntree = 5000, importance = T, subset = train)   
## Type of random forest: classification  
## Number of trees: 5000  
## No. of variables tried at each split: 4  
##   
## OOB estimate of error rate: 5.5%  
## Confusion matrix:  
## Insufficient\_Weight Normal\_Weight Obesity\_Type\_I  
## Insufficient\_Weight 122 9 0  
## Normal\_Weight 0 147 0  
## Obesity\_Type\_I 0 1 168  
## Obesity\_Type\_II 0 1 0  
## Obesity\_Type\_III 0 0 0  
## Overweight\_Level\_I 0 20 0  
## Overweight\_Level\_II 0 2 2  
## Obesity\_Type\_II Obesity\_Type\_III Overweight\_Level\_I  
## Insufficient\_Weight 0 0 0  
## Normal\_Weight 0 0 8  
## Obesity\_Type\_I 1 0 2  
## Obesity\_Type\_II 153 0 0  
## Obesity\_Type\_III 0 147 0  
## Overweight\_Level\_I 0 0 124  
## Overweight\_Level\_II 0 0 5  
## Overweight\_Level\_II class.error  
## Insufficient\_Weight 0 0.068702290  
## Normal\_Weight 2 0.063694268  
## Obesity\_Type\_I 2 0.034482759  
## Obesity\_Type\_II 0 0.006493506  
## Obesity\_Type\_III 0 0.000000000  
## Overweight\_Level\_I 3 0.156462585  
## Overweight\_Level\_II 136 0.062068966

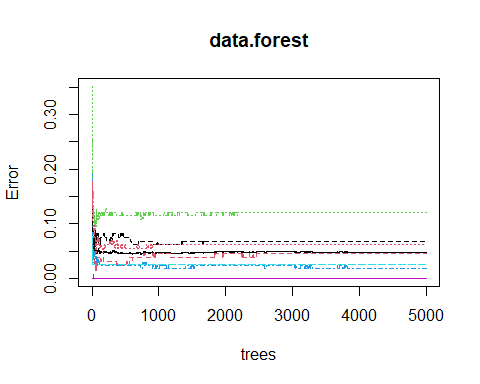
plot(data.forest)



data.forest <- randomForest(  
 weight\_category ~ .,  
 data = data,  
 subset = train,  
 ntree = 5000,  
 mtry = ncol(data) - 1,  
 importance = T  
)  
data.forest

##   
## Call:  
## randomForest(formula = weight\_category ~ ., data = data, ntree = 5000, mtry = ncol(data) - 1, importance = T, subset = train)   
## Type of random forest: classification  
## Number of trees: 5000  
## No. of variables tried at each split: 16  
##   
## OOB estimate of error rate: 4.83%  
## Confusion matrix:  
## Insufficient\_Weight Normal\_Weight Obesity\_Type\_I  
## Insufficient\_Weight 125 6 0  
## Normal\_Weight 8 138 0  
## Obesity\_Type\_I 0 0 171  
## Obesity\_Type\_II 0 0 4  
## Obesity\_Type\_III 0 0 0  
## Overweight\_Level\_I 0 5 0  
## Overweight\_Level\_II 0 0 8  
## Obesity\_Type\_II Obesity\_Type\_III Overweight\_Level\_I  
## Insufficient\_Weight 0 0 0  
## Normal\_Weight 0 0 11  
## Obesity\_Type\_I 1 0 0  
## Obesity\_Type\_II 150 0 0  
## Obesity\_Type\_III 0 147 0  
## Overweight\_Level\_I 0 0 137  
## Overweight\_Level\_II 0 0 1  
## Overweight\_Level\_II class.error  
## Insufficient\_Weight 0 0.04580153  
## Normal\_Weight 0 0.12101911  
## Obesity\_Type\_I 2 0.01724138  
## Obesity\_Type\_II 0 0.02597403  
## Obesity\_Type\_III 0 0.00000000  
## Overweight\_Level\_I 5 0.06802721  
## Overweight\_Level\_II 136 0.06206897

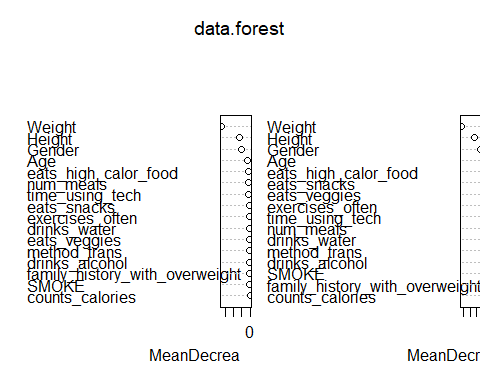
plot(data.forest)



importance(data.forest)

## Insufficient\_Weight Normal\_Weight Obesity\_Type\_I  
## Gender 33.567941 6.620219 1.966151e+02  
## Age 6.955835 21.969044 4.558723e+01  
## Height 192.270758 181.846909 3.600960e+02  
## Weight 789.845721 510.058705 7.635606e+02  
## family\_history\_with\_overweight 7.407269 5.099017 1.626893e+01  
## eats\_high\_calor\_food 5.663099 9.640023 2.234950e+01  
## eats\_veggies 25.962026 23.673422 2.407553e+01  
## num\_meals 43.273233 -1.686130 4.644282e+01  
## eats\_snacks 12.713725 27.770485 2.584563e+01  
## SMOKE 5.510698 7.371416 3.375441e-03  
## drinks\_water 17.480593 13.705093 3.235806e+01  
## counts\_calories -1.391688 -12.046501 2.229943e+00  
## exercises\_often 33.518625 2.515008 2.342145e+01  
## time\_using\_tech 56.256883 13.225996 3.021962e+01  
## drinks\_alcohol 4.976494 -5.180093 1.768500e+01  
## method\_trans 21.034882 2.990803 1.866125e+01  
## Obesity\_Type\_II Obesity\_Type\_III  
## Gender 399.435949 524.223206  
## Age 130.074372 6.036538  
## Height 91.795356 9.483723  
## Weight 836.583657 741.701388  
## family\_history\_with\_overweight 1.147465 0.000000  
## eats\_high\_calor\_food -1.054018 0.000000  
## eats\_veggies 12.526743 8.491329  
## num\_meals 18.060959 3.542797  
## eats\_snacks -4.620543 1.994911  
## SMOKE 0.000000 0.000000  
## drinks\_water 33.162233 4.906180  
## counts\_calories 0.000000 0.000000  
## exercises\_often 8.564850 4.189220  
## time\_using\_tech 9.752414 2.054273  
## drinks\_alcohol 18.770724 5.119241  
## method\_trans 1.732390 0.000000  
## Overweight\_Level\_I Overweight\_Level\_II  
## Gender 8.283746 35.6772526  
## Age 54.381308 49.4484004  
## Height 351.224199 410.1915457  
## Weight 633.261286 621.7230798  
## family\_history\_with\_overweight 9.362110 21.8945713  
## eats\_high\_calor\_food 8.027086 96.6361152  
## eats\_veggies 29.514952 21.8417121  
## num\_meals 31.585884 39.7168317  
## eats\_snacks 47.361689 29.0465217  
## SMOKE 19.167148 0.6177717  
## drinks\_water 15.906580 21.0812491  
## counts\_calories 10.407824 4.2376662  
## exercises\_often 34.580933 22.4536661  
## time\_using\_tech 16.347481 41.1421046  
## drinks\_alcohol 22.030244 23.1800904  
## method\_trans 20.180640 20.1115086  
## MeanDecreaseAccuracy MeanDecreaseGini  
## Gender 524.710784 137.2085709  
## Age 145.610809 25.7454345  
## Height 622.229227 233.9262225  
## Weight 1712.727676 420.7951060  
## family\_history\_with\_overweight 27.743069 1.4980195  
## eats\_high\_calor\_food 91.654534 13.1570509  
## eats\_veggies 34.824225 11.0976222  
## num\_meals 74.217117 9.8560343  
## eats\_snacks 60.636447 11.1964591  
## SMOKE 18.607139 1.5182872  
## drinks\_water 53.736385 8.6286447  
## counts\_calories -2.851231 0.4118914  
## exercises\_often 53.937537 10.1815509  
## time\_using\_tech 71.532712 9.9115028  
## drinks\_alcohol 32.608880 3.5248518  
## method\_trans 34.332685 3.7909938

varImpPlot(data.forest)



data.forest.pred <- predict(data.forest, test, typr = "class")  
table(data.forest.pred, test$weight\_category)

##   
## data.forest.pred Insufficient\_Weight Normal\_Weight Obesity\_Type\_I  
## Insufficient\_Weight 135 5 0  
## Normal\_Weight 6 112 0  
## Obesity\_Type\_I 0 0 169  
## Obesity\_Type\_II 0 0 5  
## Obesity\_Type\_III 0 0 0  
## Overweight\_Level\_I 0 13 0  
## Overweight\_Level\_II 0 0 3  
##   
## data.forest.pred Obesity\_Type\_II Obesity\_Type\_III Overweight\_Level\_I  
## Insufficient\_Weight 0 0 0  
## Normal\_Weight 0 0 2  
## Obesity\_Type\_I 5 1 0  
## Obesity\_Type\_II 137 0 0  
## Obesity\_Type\_III 1 176 0  
## Overweight\_Level\_I 0 0 140  
## Overweight\_Level\_II 0 0 1  
##   
## data.forest.pred Overweight\_Level\_II  
## Insufficient\_Weight 0  
## Normal\_Weight 0  
## Obesity\_Type\_I 5  
## Obesity\_Type\_II 0  
## Obesity\_Type\_III 0  
## Overweight\_Level\_I 14  
## Overweight\_Level\_II 126

mean(data.forest.pred == test$weight\_category)

## [1] 0.9422348

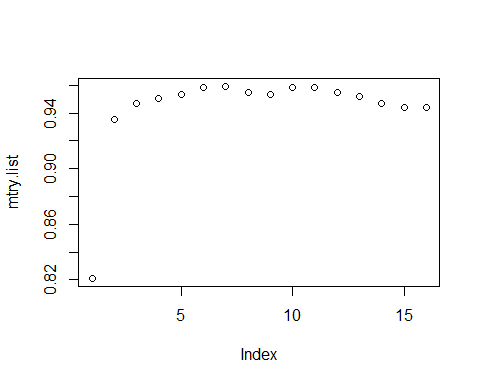
mtry.list <- c()  
i <- 1  
for (i in 1:ncol(data) - 1) {  
 data.forest.dummy <- randomForest(  
 weight\_category ~ .,  
 data = data,  
 subset = train,  
 ntree = 5000,  
 mtry = i,  
 importance = T  
 )  
 val <- predict(data.forest.dummy, test, typr = "class")  
 mtry.list[i] <- mean(val == test$weight\_category)  
}

## Warning in randomForest.default(m, y, ...): invalid mtry: reset to within valid  
## range

mtry.list

## [1] 0.8210227 0.9356061 0.9469697 0.9507576 0.9535985 0.9583333 0.9592803  
## [8] 0.9545455 0.9535985 0.9583333 0.9583333 0.9545455 0.9517045 0.9469697  
## [15] 0.9441288 0.9441288

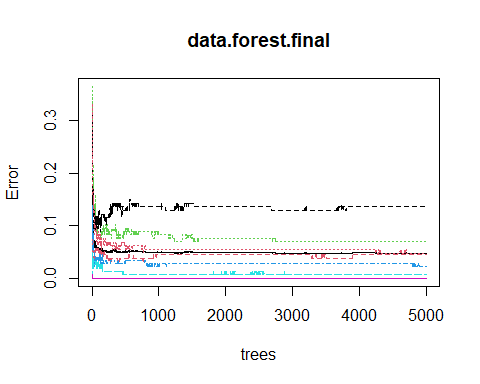
plot(mtry.list)



data.forest.final <- randomForest(  
 weight\_category ~ .,  
 data = data,  
 subset = train,  
 ntree = 5000,  
 mtry = which.max(mtry.list),  
 importance = T  
)  
data.forest.final

##   
## Call:  
## randomForest(formula = weight\_category ~ ., data = data, ntree = 5000, mtry = which.max(mtry.list), importance = T, subset = train)   
## Type of random forest: classification  
## Number of trees: 5000  
## No. of variables tried at each split: 7  
##   
## OOB estimate of error rate: 4.64%  
## Confusion matrix:  
## Insufficient\_Weight Normal\_Weight Obesity\_Type\_I  
## Insufficient\_Weight 125 6 0  
## Normal\_Weight 1 146 0  
## Obesity\_Type\_I 0 0 170  
## Obesity\_Type\_II 0 1 0  
## Obesity\_Type\_III 0 0 0  
## Overweight\_Level\_I 0 15 0  
## Overweight\_Level\_II 0 1 4  
## Obesity\_Type\_II Obesity\_Type\_III Overweight\_Level\_I  
## Insufficient\_Weight 0 0 0  
## Normal\_Weight 0 0 9  
## Obesity\_Type\_I 1 0 0  
## Obesity\_Type\_II 153 0 0  
## Obesity\_Type\_III 0 147 0  
## Overweight\_Level\_I 0 0 127  
## Overweight\_Level\_II 0 0 2  
## Overweight\_Level\_II class.error  
## Insufficient\_Weight 0 0.045801527  
## Normal\_Weight 1 0.070063694  
## Obesity\_Type\_I 3 0.022988506  
## Obesity\_Type\_II 0 0.006493506  
## Obesity\_Type\_III 0 0.000000000  
## Overweight\_Level\_I 5 0.136054422  
## Overweight\_Level\_II 138 0.048275862

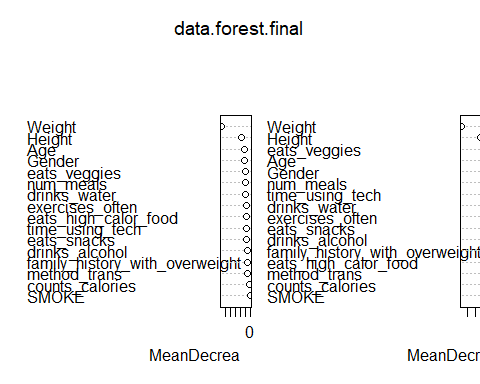
plot(data.forest.final)



importance(data.forest.final)

## Insufficient\_Weight Normal\_Weight Obesity\_Type\_I  
## Gender 34.008180 23.3260464 73.111373  
## Age 67.300879 28.5555892 76.463277  
## Height 83.506367 83.7874807 161.743558  
## Weight 392.921598 238.1283586 357.913319  
## family\_history\_with\_overweight 50.305535 9.2709582 53.116095  
## eats\_high\_calor\_food 14.762371 0.5458449 31.063287  
## eats\_veggies 64.284498 40.8091091 68.375897  
## num\_meals 62.650665 23.0169952 60.472030  
## eats\_snacks 57.273512 35.4157468 58.228157  
## SMOKE 6.023630 0.2724728 1.207259  
## drinks\_water 30.014647 29.8579245 54.213325  
## counts\_calories 1.752677 -14.3912650 5.018644  
## exercises\_often 37.018331 20.3396343 42.294096  
## time\_using\_tech 54.186503 28.2034232 45.746519  
## drinks\_alcohol 32.829456 -7.2360472 44.332289  
## method\_trans 39.922925 -4.7006330 25.988372  
## Obesity\_Type\_II Obesity\_Type\_III  
## Gender 69.3452462 142.370695  
## Age 95.2041802 31.888292  
## Height 40.4974531 28.727314  
## Weight 614.2929009 360.333059  
## family\_history\_with\_overweight 28.0386575 24.961635  
## eats\_high\_calor\_food 10.6111747 6.130873  
## eats\_veggies 69.8112540 85.171521  
## num\_meals 44.8107001 22.473425  
## eats\_snacks 26.8590390 29.248588  
## SMOKE 0.3032205 1.000100  
## drinks\_water 62.8427858 11.096973  
## counts\_calories 2.2335380 2.456470  
## exercises\_often 35.1220586 18.798790  
## time\_using\_tech 30.5854438 37.448782  
## drinks\_alcohol 40.3342154 22.018525  
## method\_trans 14.4328901 14.904263  
## Overweight\_Level\_I Overweight\_Level\_II  
## Gender 45.13889 52.2830048  
## Age 87.09885 88.2096511  
## Height 131.20420 157.7307347  
## Weight 288.28793 287.7778205  
## family\_history\_with\_overweight 47.57807 55.1892703  
## eats\_high\_calor\_food 37.57189 80.4449438  
## eats\_veggies 56.08817 56.5787793  
## num\_meals 58.36798 62.7766888  
## eats\_snacks 53.72144 57.0851357  
## SMOKE 12.56119 0.4510447  
## drinks\_water 42.00078 35.2825148  
## counts\_calories 26.13515 11.4248073  
## exercises\_often 49.44970 35.0090045  
## time\_using\_tech 37.46082 42.0853018  
## drinks\_alcohol 54.71310 58.8064346  
## method\_trans 38.72027 31.2036117  
## MeanDecreaseAccuracy MeanDecreaseGini  
## Gender 116.44828 58.029867  
## Age 124.66975 64.697234  
## Height 187.16857 121.733795  
## Weight 574.47890 376.476635  
## family\_history\_with\_overweight 61.93797 16.952076  
## eats\_high\_calor\_food 77.51063 13.094315  
## eats\_veggies 96.67007 71.360196  
## num\_meals 95.73860 34.327900  
## eats\_snacks 73.72160 26.352234  
## SMOKE 9.56560 1.284261  
## drinks\_water 92.42173 27.821827  
## counts\_calories 15.49854 2.090799  
## exercises\_often 78.15684 27.648344  
## time\_using\_tech 74.00320 29.516048  
## drinks\_alcohol 67.65241 19.746329  
## method\_trans 51.21212 11.158489

varImpPlot(data.forest.final)



data.forest.final.pred <-  
 predict(data.forest.final, test, typr = "class")  
table(data.forest.final.pred, test$weight\_category)

##   
## data.forest.final.pred Insufficient\_Weight Normal\_Weight Obesity\_Type\_I  
## Insufficient\_Weight 134 2 0  
## Normal\_Weight 7 123 0  
## Obesity\_Type\_I 0 0 173  
## Obesity\_Type\_II 0 0 3  
## Obesity\_Type\_III 0 0 0  
## Overweight\_Level\_I 0 5 1  
## Overweight\_Level\_II 0 0 0  
##   
## data.forest.final.pred Obesity\_Type\_II Obesity\_Type\_III Overweight\_Level\_I  
## Insufficient\_Weight 0 0 0  
## Normal\_Weight 0 0 8  
## Obesity\_Type\_I 1 1 0  
## Obesity\_Type\_II 141 0 0  
## Obesity\_Type\_III 1 176 0  
## Overweight\_Level\_I 0 0 131  
## Overweight\_Level\_II 0 0 4  
##   
## data.forest.final.pred Overweight\_Level\_II  
## Insufficient\_Weight 0  
## Normal\_Weight 5  
## Obesity\_Type\_I 2  
## Obesity\_Type\_II 0  
## Obesity\_Type\_III 0  
## Overweight\_Level\_I 5  
## Overweight\_Level\_II 133

mean(data.forest.final.pred == test$weight\_category)

## [1] 0.9573864

# rf.cv <- rfcv(data, data$weight\_category, cv.fold=10)  
#  
# with(rf.cv, plot(n.var, error.cv))