Practical Machine Learning

DainaE

Thursday, May 21, 2015

The goal of your project is to predict the manner in which they did the exercise. This is the "classe" variable in the training set. You may use any of the other variables to predict with.

Data Processing and Analysis

The training and testing datasets used in the analysis may be found as follows:

Training dataset: <https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv>

Testing dataset: <https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv>

We begin by loading the required libraries and reading in the training and testing datasets, assigning missing values to entries that are currently 'NA' or blank.

library(corrplot)

## Warning: package 'corrplot' was built under R version 3.1.3

library(caret)

## Warning: package 'caret' was built under R version 3.1.3

## Loading required package: lattice  
## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.1.2

setwd("C:/Personal/Coursera/Assignment/pml/PML-Assignment1")  
  
pml\_train <- data.frame()  
pml\_test <- data.frame()  
pml\_sub\_train <- data.frame()  
pml\_sub\_test <- data.frame()  
  
pml\_train <- read.csv(file = "pml-training.csv",header = TRUE,,na.strings=c("NA", ""))  
pml\_test <- read.csv(file = "pml-testing.csv",header = TRUE,na.strings = c("NA", ""))  
head(pml\_train)

## X user\_name raw\_timestamp\_part\_1 raw\_timestamp\_part\_2 cvtd\_timestamp  
## 1 1 carlitos 1323084231 788290 05/12/2011 11:23  
## 2 2 carlitos 1323084231 808298 05/12/2011 11:23  
## 3 3 carlitos 1323084231 820366 05/12/2011 11:23  
## 4 4 carlitos 1323084232 120339 05/12/2011 11:23  
## 5 5 carlitos 1323084232 196328 05/12/2011 11:23  
## 6 6 carlitos 1323084232 304277 05/12/2011 11:23  
## new\_window num\_window roll\_belt pitch\_belt yaw\_belt total\_accel\_belt  
## 1 no 11 1.41 8.07 -94.4 3  
## 2 no 11 1.41 8.07 -94.4 3  
## 3 no 11 1.42 8.07 -94.4 3  
## 4 no 12 1.48 8.05 -94.4 3  
## 5 no 12 1.48 8.07 -94.4 3  
## 6 no 12 1.45 8.06 -94.4 3  
## kurtosis\_roll\_belt kurtosis\_picth\_belt kurtosis\_yaw\_belt  
## 1 <NA> <NA> <NA>  
## 2 <NA> <NA> <NA>  
## 3 <NA> <NA> <NA>  
## 4 <NA> <NA> <NA>  
## 5 <NA> <NA> <NA>  
## 6 <NA> <NA> <NA>  
## skewness\_roll\_belt skewness\_roll\_belt.1 skewness\_yaw\_belt max\_roll\_belt  
## 1 <NA> <NA> <NA> NA  
## 2 <NA> <NA> <NA> NA  
## 3 <NA> <NA> <NA> NA  
## 4 <NA> <NA> <NA> NA  
## 5 <NA> <NA> <NA> NA  
## 6 <NA> <NA> <NA> NA  
## max\_picth\_belt max\_yaw\_belt min\_roll\_belt min\_pitch\_belt min\_yaw\_belt  
## 1 NA <NA> NA NA <NA>  
## 2 NA <NA> NA NA <NA>  
## 3 NA <NA> NA NA <NA>  
## 4 NA <NA> NA NA <NA>  
## 5 NA <NA> NA NA <NA>  
## 6 NA <NA> NA NA <NA>  
## amplitude\_roll\_belt amplitude\_pitch\_belt amplitude\_yaw\_belt  
## 1 NA NA <NA>  
## 2 NA NA <NA>  
## 3 NA NA <NA>  
## 4 NA NA <NA>  
## 5 NA NA <NA>  
## 6 NA NA <NA>  
## var\_total\_accel\_belt avg\_roll\_belt stddev\_roll\_belt var\_roll\_belt  
## 1 NA NA NA NA  
## 2 NA NA NA NA  
## 3 NA NA NA NA  
## 4 NA NA NA NA  
## 5 NA NA NA NA  
## 6 NA NA NA NA  
## avg\_pitch\_belt stddev\_pitch\_belt var\_pitch\_belt avg\_yaw\_belt  
## 1 NA NA NA NA  
## 2 NA NA NA NA  
## 3 NA NA NA NA  
## 4 NA NA NA NA  
## 5 NA NA NA NA  
## 6 NA NA NA NA  
## stddev\_yaw\_belt var\_yaw\_belt gyros\_belt\_x gyros\_belt\_y gyros\_belt\_z  
## 1 NA NA 0.00 0.00 -0.02  
## 2 NA NA 0.02 0.00 -0.02  
## 3 NA NA 0.00 0.00 -0.02  
## 4 NA NA 0.02 0.00 -0.03  
## 5 NA NA 0.02 0.02 -0.02  
## 6 NA NA 0.02 0.00 -0.02  
## accel\_belt\_x accel\_belt\_y accel\_belt\_z magnet\_belt\_x magnet\_belt\_y  
## 1 -21 4 22 -3 599  
## 2 -22 4 22 -7 608  
## 3 -20 5 23 -2 600  
## 4 -22 3 21 -6 604  
## 5 -21 2 24 -6 600  
## 6 -21 4 21 0 603  
## magnet\_belt\_z roll\_arm pitch\_arm yaw\_arm total\_accel\_arm var\_accel\_arm  
## 1 -313 -128 22.5 -161 34 NA  
## 2 -311 -128 22.5 -161 34 NA  
## 3 -305 -128 22.5 -161 34 NA  
## 4 -310 -128 22.1 -161 34 NA  
## 5 -302 -128 22.1 -161 34 NA  
## 6 -312 -128 22.0 -161 34 NA  
## avg\_roll\_arm stddev\_roll\_arm var\_roll\_arm avg\_pitch\_arm stddev\_pitch\_arm  
## 1 NA NA NA NA NA  
## 2 NA NA NA NA NA  
## 3 NA NA NA NA NA  
## 4 NA NA NA NA NA  
## 5 NA NA NA NA NA  
## 6 NA NA NA NA NA  
## var\_pitch\_arm avg\_yaw\_arm stddev\_yaw\_arm var\_yaw\_arm gyros\_arm\_x  
## 1 NA NA NA NA 0.00  
## 2 NA NA NA NA 0.02  
## 3 NA NA NA NA 0.02  
## 4 NA NA NA NA 0.02  
## 5 NA NA NA NA 0.00  
## 6 NA NA NA NA 0.02  
## gyros\_arm\_y gyros\_arm\_z accel\_arm\_x accel\_arm\_y accel\_arm\_z magnet\_arm\_x  
## 1 0.00 -0.02 -288 109 -123 -368  
## 2 -0.02 -0.02 -290 110 -125 -369  
## 3 -0.02 -0.02 -289 110 -126 -368  
## 4 -0.03 0.02 -289 111 -123 -372  
## 5 -0.03 0.00 -289 111 -123 -374  
## 6 -0.03 0.00 -289 111 -122 -369  
## magnet\_arm\_y magnet\_arm\_z kurtosis\_roll\_arm kurtosis\_picth\_arm  
## 1 337 516 <NA> <NA>  
## 2 337 513 <NA> <NA>  
## 3 344 513 <NA> <NA>  
## 4 344 512 <NA> <NA>  
## 5 337 506 <NA> <NA>  
## 6 342 513 <NA> <NA>  
## kurtosis\_yaw\_arm skewness\_roll\_arm skewness\_pitch\_arm skewness\_yaw\_arm  
## 1 <NA> <NA> <NA> <NA>  
## 2 <NA> <NA> <NA> <NA>  
## 3 <NA> <NA> <NA> <NA>  
## 4 <NA> <NA> <NA> <NA>  
## 5 <NA> <NA> <NA> <NA>  
## 6 <NA> <NA> <NA> <NA>  
## max\_roll\_arm max\_picth\_arm max\_yaw\_arm min\_roll\_arm min\_pitch\_arm  
## 1 NA NA NA NA NA  
## 2 NA NA NA NA NA  
## 3 NA NA NA NA NA  
## 4 NA NA NA NA NA  
## 5 NA NA NA NA NA  
## 6 NA NA NA NA NA  
## min\_yaw\_arm amplitude\_roll\_arm amplitude\_pitch\_arm amplitude\_yaw\_arm  
## 1 NA NA NA NA  
## 2 NA NA NA NA  
## 3 NA NA NA NA  
## 4 NA NA NA NA  
## 5 NA NA NA NA  
## 6 NA NA NA NA  
## roll\_dumbbell pitch\_dumbbell yaw\_dumbbell kurtosis\_roll\_dumbbell  
## 1 13.05217 -70.49400 -84.87394 <NA>  
## 2 13.13074 -70.63751 -84.71065 <NA>  
## 3 12.85075 -70.27812 -85.14078 <NA>  
## 4 13.43120 -70.39379 -84.87363 <NA>  
## 5 13.37872 -70.42856 -84.85306 <NA>  
## 6 13.38246 -70.81759 -84.46500 <NA>  
## kurtosis\_picth\_dumbbell kurtosis\_yaw\_dumbbell skewness\_roll\_dumbbell  
## 1 <NA> <NA> <NA>  
## 2 <NA> <NA> <NA>  
## 3 <NA> <NA> <NA>  
## 4 <NA> <NA> <NA>  
## 5 <NA> <NA> <NA>  
## 6 <NA> <NA> <NA>  
## skewness\_pitch\_dumbbell skewness\_yaw\_dumbbell max\_roll\_dumbbell  
## 1 <NA> <NA> NA  
## 2 <NA> <NA> NA  
## 3 <NA> <NA> NA  
## 4 <NA> <NA> NA  
## 5 <NA> <NA> NA  
## 6 <NA> <NA> NA  
## max\_picth\_dumbbell max\_yaw\_dumbbell min\_roll\_dumbbell min\_pitch\_dumbbell  
## 1 NA <NA> NA NA  
## 2 NA <NA> NA NA  
## 3 NA <NA> NA NA  
## 4 NA <NA> NA NA  
## 5 NA <NA> NA NA  
## 6 NA <NA> NA NA  
## min\_yaw\_dumbbell amplitude\_roll\_dumbbell amplitude\_pitch\_dumbbell  
## 1 <NA> NA NA  
## 2 <NA> NA NA  
## 3 <NA> NA NA  
## 4 <NA> NA NA  
## 5 <NA> NA NA  
## 6 <NA> NA NA  
## amplitude\_yaw\_dumbbell total\_accel\_dumbbell var\_accel\_dumbbell  
## 1 <NA> 37 NA  
## 2 <NA> 37 NA  
## 3 <NA> 37 NA  
## 4 <NA> 37 NA  
## 5 <NA> 37 NA  
## 6 <NA> 37 NA  
## avg\_roll\_dumbbell stddev\_roll\_dumbbell var\_roll\_dumbbell  
## 1 NA NA NA  
## 2 NA NA NA  
## 3 NA NA NA  
## 4 NA NA NA  
## 5 NA NA NA  
## 6 NA NA NA  
## avg\_pitch\_dumbbell stddev\_pitch\_dumbbell var\_pitch\_dumbbell  
## 1 NA NA NA  
## 2 NA NA NA  
## 3 NA NA NA  
## 4 NA NA NA  
## 5 NA NA NA  
## 6 NA NA NA  
## avg\_yaw\_dumbbell stddev\_yaw\_dumbbell var\_yaw\_dumbbell gyros\_dumbbell\_x  
## 1 NA NA NA 0  
## 2 NA NA NA 0  
## 3 NA NA NA 0  
## 4 NA NA NA 0  
## 5 NA NA NA 0  
## 6 NA NA NA 0  
## gyros\_dumbbell\_y gyros\_dumbbell\_z accel\_dumbbell\_x accel\_dumbbell\_y  
## 1 -0.02 0.00 -234 47  
## 2 -0.02 0.00 -233 47  
## 3 -0.02 0.00 -232 46  
## 4 -0.02 -0.02 -232 48  
## 5 -0.02 0.00 -233 48  
## 6 -0.02 0.00 -234 48  
## accel\_dumbbell\_z magnet\_dumbbell\_x magnet\_dumbbell\_y magnet\_dumbbell\_z  
## 1 -271 -559 293 -65  
## 2 -269 -555 296 -64  
## 3 -270 -561 298 -63  
## 4 -269 -552 303 -60  
## 5 -270 -554 292 -68  
## 6 -269 -558 294 -66  
## roll\_forearm pitch\_forearm yaw\_forearm kurtosis\_roll\_forearm  
## 1 28.4 -63.9 -153 <NA>  
## 2 28.3 -63.9 -153 <NA>  
## 3 28.3 -63.9 -152 <NA>  
## 4 28.1 -63.9 -152 <NA>  
## 5 28.0 -63.9 -152 <NA>  
## 6 27.9 -63.9 -152 <NA>  
## kurtosis\_picth\_forearm kurtosis\_yaw\_forearm skewness\_roll\_forearm  
## 1 <NA> <NA> <NA>  
## 2 <NA> <NA> <NA>  
## 3 <NA> <NA> <NA>  
## 4 <NA> <NA> <NA>  
## 5 <NA> <NA> <NA>  
## 6 <NA> <NA> <NA>  
## skewness\_pitch\_forearm skewness\_yaw\_forearm max\_roll\_forearm  
## 1 <NA> <NA> NA  
## 2 <NA> <NA> NA  
## 3 <NA> <NA> NA  
## 4 <NA> <NA> NA  
## 5 <NA> <NA> NA  
## 6 <NA> <NA> NA  
## max\_picth\_forearm max\_yaw\_forearm min\_roll\_forearm min\_pitch\_forearm  
## 1 NA <NA> NA NA  
## 2 NA <NA> NA NA  
## 3 NA <NA> NA NA  
## 4 NA <NA> NA NA  
## 5 NA <NA> NA NA  
## 6 NA <NA> NA NA  
## min\_yaw\_forearm amplitude\_roll\_forearm amplitude\_pitch\_forearm  
## 1 <NA> NA NA  
## 2 <NA> NA NA  
## 3 <NA> NA NA  
## 4 <NA> NA NA  
## 5 <NA> NA NA  
## 6 <NA> NA NA  
## amplitude\_yaw\_forearm total\_accel\_forearm var\_accel\_forearm  
## 1 <NA> 36 NA  
## 2 <NA> 36 NA  
## 3 <NA> 36 NA  
## 4 <NA> 36 NA  
## 5 <NA> 36 NA  
## 6 <NA> 36 NA  
## avg\_roll\_forearm stddev\_roll\_forearm var\_roll\_forearm avg\_pitch\_forearm  
## 1 NA NA NA NA  
## 2 NA NA NA NA  
## 3 NA NA NA NA  
## 4 NA NA NA NA  
## 5 NA NA NA NA  
## 6 NA NA NA NA  
## stddev\_pitch\_forearm var\_pitch\_forearm avg\_yaw\_forearm  
## 1 NA NA NA  
## 2 NA NA NA  
## 3 NA NA NA  
## 4 NA NA NA  
## 5 NA NA NA  
## 6 NA NA NA  
## stddev\_yaw\_forearm var\_yaw\_forearm gyros\_forearm\_x gyros\_forearm\_y  
## 1 NA NA 0.03 0.00  
## 2 NA NA 0.02 0.00  
## 3 NA NA 0.03 -0.02  
## 4 NA NA 0.02 -0.02  
## 5 NA NA 0.02 0.00  
## 6 NA NA 0.02 -0.02  
## gyros\_forearm\_z accel\_forearm\_x accel\_forearm\_y accel\_forearm\_z  
## 1 -0.02 192 203 -215  
## 2 -0.02 192 203 -216  
## 3 0.00 196 204 -213  
## 4 0.00 189 206 -214  
## 5 -0.02 189 206 -214  
## 6 -0.03 193 203 -215  
## magnet\_forearm\_x magnet\_forearm\_y magnet\_forearm\_z classe  
## 1 -17 654 476 A  
## 2 -18 661 473 A  
## 3 -18 658 469 A  
## 4 -16 658 469 A  
## 5 -17 655 473 A  
## 6 -9 660 478 A

## need to exclude 0 value columns and non measurement columns   
  
#exclude 0 columns  
csums <- colSums(is.na(pml\_train))  
csums\_log <- (csums == 0)  
pml\_sub\_train <- pml\_train[, (colSums(is.na(pml\_train)) == 0)]  
pml\_test <- pml\_test[, (colSums(is.na(pml\_train)) == 0)]  
  
#include only measurement columns  
del\_cols <- grepl("X|user\_name|timestamp|new\_window", colnames(pml\_sub\_train))  
pml\_sub\_train <- pml\_sub\_train[, !del\_cols]  
pml\_sub\_test <- pml\_test[, !del\_cols]  
  
dim(pml\_sub\_train)

## [1] 19622 54

dim(pml\_sub\_test)

## [1] 20 54

inTrain = createDataPartition(y = pml\_train$classe, p = 0.7, list = FALSE)  
pml\_train\_subset = pml\_sub\_train[inTrain, ]  
pml\_train\_t = pml\_sub\_train[-inTrain, ]

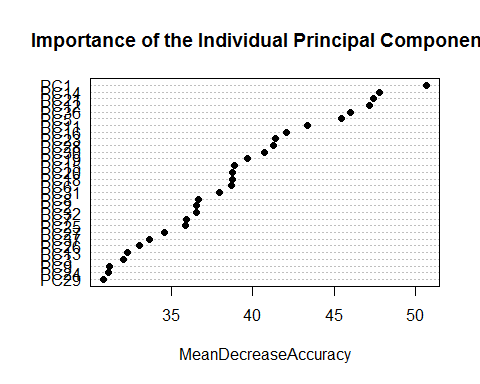
#Principal COmponent Analysis fitting  
preProc <- preProcess(pml\_train\_subset[, -54], method = "pca", thresh = 0.99)  
trainPC <- predict(preProc, pml\_train\_subset[, -54])  
valid\_testPC <- predict(preProc,pml\_train\_t[, -54])  
#RandomForest fitting  
modelFit <- train(pml\_train\_subset$classe ~ ., method = "rf", data = trainPC, trControl = trainControl(method = "cv", number = 4), importance = TRUE)

## Loading required package: randomForest

## Warning: package 'randomForest' was built under R version 3.1.3

## randomForest 4.6-10  
## Type rfNews() to see new features/changes/bug fixes.

#Depicting importance of the resulting principal components of the trained model  
  
varImpPlot(modelFit$finalModel, sort = TRUE, type = 1, pch = 19, col = 1, cex = 1,   
 main = "Importance of the Individual Principal Components")



#You should create a report describing how you built your model, how you used cross #validation, what you think the expected out of sample error is, and why you made the #choices you did. You will also use your prediction model to predict 20 different test #cases.   
  
pred\_valid\_rf <- predict(modelFit, valid\_testPC)  
confus <- confusionMatrix(pml\_train\_t$classe, pred\_valid\_rf)  
confus$table

## Reference  
## Prediction A B C D E  
## A 1668 3 2 1 0  
## B 25 1103 8 1 2  
## C 1 15 1003 7 0  
## D 0 0 43 918 3  
## E 0 1 3 4 1074

accur <- postResample(pml\_train\_t$classe, pred\_valid\_rf)  
model\_accuracy <- accur[[1]]  
model\_accuracy

## [1] 0.9797791

out\_of\_sample\_error <- 1 - model\_accuracy  
out\_of\_sample\_error

## [1] 0.0202209

testPC <- predict(preProc, pml\_sub\_test[, -54])  
pred\_final <- predict(modelFit, testPC)  
pred\_final

## [1] B A C A A C D B A A B C B A E E A B B B  
## Levels: A B C D E