# Predicting Correct Weight Lifting Technique using Human Activity Recgonition Data

### Coursera: Practical Machine Learning

### December 2015

### Anthony L.

# Introduction

Using devices such as Jawbone Up, Nike FuelBand, and Fitbit it is now possible to collect a large amount of data about personal activity relatively inexpensively. These type of devices are part of the quantified self movement - a group of enthusiasts who take measurements about themselves regularly to improve their health, to find patterns in their behavior, or because they are tech geeks. One thing that people regularly do is quantify how much of a particular activity they do, but they rarely quantify how well they do it.

In this project, the goal will be to use data from accelerometers on the belt, forearm, arm, and dumbell of 6 participants. They were asked to perform barbell lifts correctly and incorrectly in 5 different ways. More information is available from the website here: http://groupware.les.inf.puc-rio.br/har (http://groupware.les.inf.puc-rio.br/har) (see the section on the Weight Lifting Exercise Dataset). From this data, a predictive model will be built to determine whether or not the barbell lifts are being performed correctly.

# Preparing Data

The data is downloaded and loaded for analysis. The outcome data is classe which is a factor with 5 levels: A, B, C, D, E. A represents the barbell lifts being performed correctly while the rest are all variations of an incorrect form. The predictors that were selected include measures of roll, pitch, yaw, gyration, acceleration, and magnet.

```
#Load Packages
library(caret)
library(randomForest)
library(ggplot2)
```

```
#Set WD
setwd("C:/Users/Anthony/Documents/Coursera/08_machine/course_project")
```

```
#Download Data
url1 <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv"
download.file(url=url1, destfile="pml-training.csv")
url2 <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv"
download.file(url=url2, destfile="pml-testing.csv")
```

```
#Load Data
pmltrain<-read.csv("pml-training.csv")
pmltest<-read.csv("pml-testing.csv")

#Distribution of Outcomes
table(pmltrain$classe)
```

```
##
##    A    B    C    D    E
## 5580 3797 3422 3216 3607
```

```
#Predictors
varlist <- grep("^roll|^pitch|^yaw|^gyro|^accel|^magnet|classe", names(pmltrain))
train <- pmltrain[,varlist]
```

Due to limitations in computational speed of the machine used in this project, a data partition of 10% was used to fit the model.

```
#Data Partition
inTrain <- createDataPartition(y=train$classe, p=0.1, list=F)
train <- train[inTrain,]
```

# Machine Learning Algorithm

The Random Forest machine learning algorithm is selected to predict the activity quality from the monitors. The model fit with a 92% accuracy rate.

```
##Random Forest Algorithm
modFit <- train(classe ~ .,
                data = train,
                prox=T,
                method = "rf")

modFit
```

A second model was created that used a 5-fold Cross Validation with 5 repetitions in order to enhance the model fit. This model fit with a 94% accuracy rate and was selected to use to predicting the test outcomes.

```
##5-fold Repeated Cross Validation
fitControl <- trainControl(
        method = "repeatedcv",
        number = 5,
        repeats = 5)

#Random Forest Algorithm
modFit1 <- train(classe ~ .,
                 data = train,
                 prox=T,
                 method = "rf",
                 trControl = fitControl)

modFit1
```

```
## Random Forest
##
## 1964 samples
##   48 predictor
##    5 classes: 'A', 'B', 'C', 'D', 'E'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold, repeated 5 times)
## Summary of sample sizes: 1571, 1571, 1571, 1572, 1571, 1571, ...
## Resampling results across tuning parameters:
##
##   mtry  Accuracy   Kappa      Accuracy SD  Kappa SD
##    2    0.9352260  0.9179377  0.01587382   0.02012124
##   25    0.9397088  0.9236441  0.01167580   0.01480751
##   48    0.9327869  0.9149059  0.01417615   0.01797680
##
## Accuracy was used to select the optimal model using  the largest value.
## The final value used for the model was mtry = 25.
```

# Out of Sample Error and Cross Validation

The model was fit with with two types of cross validation. The first was the default used by the random forest model which is a bootstrapped resampling with 25 reptitions which returned an out of sample error of 5.6%. The second was also with the random forest model a 10-fold cross validation resampling with 10 repititons which returned a slightly lower out of sample error of 5.1%.

```
modFit1$finalModel
```

```
##
## Call:
##  randomForest(x = x, y = y, mtry = param$mtry, proximity = ..1)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 25
##
##         OOB estimate of  error rate: 5.09%
## Confusion matrix:
##     A   B   C   D   E class.error
## A 550   3   2   3   0  0.01433692
## B  21 349   8   2   0  0.08157895
## C   0  10 329   4   0  0.04081633
## D   1   3  16 297   5  0.07763975
## E   2   4   7   9 339  0.06094183
```

# Test Cases

The random forest algorithm with 5-fold 5 reptition cross validation was used to predict the outcome of 20 test cases.

```
data.frame(pmltest$user_name, pmltest$cvtd_timestamp, predict(modFit1, pmltest))
```

```
##    pmltest.user_name pmltest.cvtd_timestamp predict.modFit1..pmltest.
## 1              pedro       05/12/2011 14:23                         B
## 2             jeremy       30/11/2011 17:11                         A
## 3             jeremy       30/11/2011 17:11                         A
## 4             adelmo       02/12/2011 13:33                         A
## 5             eurico       28/11/2011 14:13                         A
## 6             jeremy       30/11/2011 17:12                         E
## 7             jeremy       30/11/2011 17:12                         D
## 8             jeremy       30/11/2011 17:11                         B
## 9           carlitos       05/12/2011 11:24                         A
## 10           charles       02/12/2011 14:57                         A
## 11          carlitos       05/12/2011 11:24                         B
## 12            jeremy       30/11/2011 17:11                         C
## 13            eurico       28/11/2011 14:14                         B
## 14            jeremy       30/11/2011 17:10                         A
## 15            jeremy       30/11/2011 17:12                         E
## 16            eurico       28/11/2011 14:15                         E
## 17             pedro       05/12/2011 14:22                         A
## 18          carlitos       05/12/2011 11:24                         D
## 19             pedro       05/12/2011 14:23                         A
## 20            eurico       28/11/2011 14:14                         B
```