Practical Machine Learning Project

Angelika

December 11, 2019

Problem Description (from Instructor)

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 types 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 is to use data from accelerometers on the belt, forearm, arm, and dumbbell 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 (see the section on the Weight Lifting Exercise Dataset).

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.

Loading the data

The code below creates a "coursera_practical_machine_learning" directory within user's current working directory, downloads both the train and test datafiles directly from the source, and reads them in using the data.table package. The first few entries of the dataset are presented in the output below.

```
library("data.table")


if (!file.exists("coursera_practical_machine_learning")){dir.create("coursera_practical_machine_learning")}

fileUrl <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv"

download.file(fileUrl, destfile = "./coursera_practical_machine_learning/pml-training.csv", method = "curl")


train<-read.csv("./coursera_practical_machine_learning/pml-training.csv", header=TRUE)


fileUrl <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv"

download.file(fileUrl, destfile = "./coursera_practical_machine_learning/pml-testing.csv", method = "curl")
```

```r
test<-read.csv("./coursera_practical_machine_learning/pml-testing.csv", heade
r=TRUE)
```

```r
head(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
## 2
## 3
## 4
## 5
## 6
##   skewness_roll_belt skewness_roll_belt.1 skewness_yaw_belt max_roll_belt
## 1                                                                      NA
## 2                                                                      NA
## 3                                                                      NA
## 4                                                                      NA
## 5                                                                      NA
## 6                                                                      NA
##   max_picth_belt max_yaw_belt min_roll_belt min_pitch_belt min_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
##   amplitude_roll_belt amplitude_pitch_belt amplitude_yaw_belt
## 1                  NA                   NA
## 2                  NA                   NA
## 3                  NA                   NA
## 4                  NA                   NA
## 5                  NA                   NA
## 6                  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
## 2          337          513
## 3          344          513
## 4          344          512
## 5          337          506
## 6          342          513
##   kurtosis_yaw_arm skewness_roll_arm skewness_pitch_arm skewness_yaw_arm
## 1
## 2
## 3
## 4
## 5
## 6
##   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
## 2      13.13074      -70.63751    -84.71065
## 3      12.85075      -70.27812    -85.14078
## 4      13.43120      -70.39379    -84.87363
## 5      13.37872      -70.42856    -84.85306
```

```
## 6      13.38246      -70.81759    -84.46500
##   kurtosis_picth_dumbbell kurtosis_yaw_dumbbell skewness_roll_dumbbell
## 1
## 2
## 3
## 4
## 5
## 6
##   skewness_pitch_dumbbell skewness_yaw_dumbbell max_roll_dumbbell
## 1                                                              NA
## 2                                                              NA
## 3                                                              NA
## 4                                                              NA
## 5                                                              NA
## 6                                                              NA
##   max_picth_dumbbell max_yaw_dumbbell min_roll_dumbbell min_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
##   min_yaw_dumbbell amplitude_roll_dumbbell amplitude_pitch_dumbbell
## 1                                       NA                       NA
## 2                                       NA                       NA
## 3                                       NA                       NA
## 4                                       NA                       NA
## 5                                       NA                       NA
## 6                                       NA                       NA
##   amplitude_yaw_dumbbell total_accel_dumbbell var_accel_dumbbell
## 1                                          37                 NA
## 2                                          37                 NA
## 3                                          37                 NA
## 4                                          37                 NA
```

```
## 5                                          37                 NA
## 6                                          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
## 2         28.3         -63.9        -153
## 3         28.3         -63.9        -152
## 4         28.1         -63.9        -152
## 5         28.0         -63.9        -152
## 6         27.9         -63.9        -152
##   kurtosis_picth_forearm kurtosis_yaw_forearm skewness_roll_forearm
## 1
## 2
## 3
## 4
## 5
## 6
##   skewness_pitch_forearm skewness_yaw_forearm max_roll_forearm
## 1                                                           NA
## 2                                                           NA
## 3                                                           NA
## 4                                                           NA
## 5                                                           NA
## 6                                                           NA
##   max_picth_forearm max_yaw_forearm min_roll_forearm min_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
##   min_yaw_forearm amplitude_roll_forearm amplitude_pitch_forearm
## 1                                     NA                      NA
## 2                                     NA                      NA
```

```
## 3                                     NA                       NA
## 4                                     NA                       NA
## 5                                     NA                       NA
## 6                                     NA                       NA
##   amplitude_yaw_forearm total_accel_forearm var_accel_forearm
## 1                                        36                NA
## 2                                        36                NA
## 3                                        36                NA
## 4                                        36                NA
## 5                                        36                NA
## 6                                        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
```

Date Cleaning and Pre-Processing

There appear to be some columns that have a lot of null values, perhaps even the entire column is null. We remove such columns for both the test and train datasets.

```
library(dplyr)
## Warning: package 'dplyr' was built under R version 3.5.3
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:data.table':
##
##     between, first, last
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
dim(train)
## [1] 19622   160
not_any_na <- function(x) all(!is.na(x))
train<-train %>% select_if(not_any_na)
test<-test %>% select_if(not_any_na)
```

```
head(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
## 2
## 3
## 4
## 5
## 6
##   skewness_roll_belt skewness_roll_belt.1 skewness_yaw_belt max_yaw_belt
## 1
## 2
## 3
## 4
## 5
## 6
##   min_yaw_belt amplitude_yaw_belt gyros_belt_x gyros_belt_y gyros_belt_z
## 1                                         0.00         0.00        -0.02
## 2                                         0.02         0.00        -0.02
## 3                                         0.00         0.00        -0.02
## 4                                         0.02         0.00        -0.03
```

```
## 5                                        0.02         0.02        -0.02
## 6                                        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 gyros_arm_x
## 1          -313     -128      22.5    -161              34        0.00
## 2          -311     -128      22.5    -161              34        0.02
## 3          -305     -128      22.5    -161              34        0.02
## 4          -310     -128      22.1    -161              34        0.02
## 5          -302     -128      22.1    -161              34        0.00
## 6          -312     -128      22.0    -161              34        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
## 2          337          513
## 3          344          513
## 4          344          512
## 5          337          506
## 6          342          513
##   kurtosis_yaw_arm skewness_roll_arm skewness_pitch_arm skewness_yaw_arm
## 1
## 2
## 3
```

```
## 4
## 5
## 6
##   roll_dumbbell pitch_dumbbell yaw_dumbbell kurtosis_roll_dumbbell
## 1      13.05217      -70.49400    -84.87394
## 2      13.13074      -70.63751    -84.71065
## 3      12.85075      -70.27812    -85.14078
## 4      13.43120      -70.39379    -84.87363
## 5      13.37872      -70.42856    -84.85306
## 6      13.38246      -70.81759    -84.46500
##   kurtosis_picth_dumbbell kurtosis_yaw_dumbbell skewness_roll_dumbbell
## 1
## 2
## 3
## 4
## 5
## 6
##   skewness_pitch_dumbbell skewness_yaw_dumbbell max_yaw_dumbbell
## 1
## 2
## 3
## 4
## 5
## 6
##   min_yaw_dumbbell amplitude_yaw_dumbbell total_accel_dumbbell
## 1                                                           37
## 2                                                           37
## 3                                                           37
## 4                                                           37
## 5                                                           37
## 6                                                           37
##   gyros_dumbbell_x gyros_dumbbell_y gyros_dumbbell_z accel_dumbbell_x
## 1                0            -0.02             0.00             -234
## 2                0            -0.02             0.00             -233
```

```
## 3                0            -0.02            0.00              -232
## 4                0            -0.02           -0.02              -232
## 5                0            -0.02            0.00              -233
## 6                0            -0.02            0.00              -234
##   accel_dumbbell_y accel_dumbbell_z magnet_dumbbell_x magnet_dumbbell_y
## 1               47             -271              -559               293
## 2               47             -269              -555               296
## 3               46             -270              -561               298
## 4               48             -269              -552               303
## 5               48             -270              -554               292
## 6               48             -269              -558               294
##   magnet_dumbbell_z roll_forearm pitch_forearm yaw_forearm
## 1               -65         28.4         -63.9        -153
## 2               -64         28.3         -63.9        -153
## 3               -63         28.3         -63.9        -152
## 4               -60         28.1         -63.9        -152
## 5               -68         28.0         -63.9        -152
## 6               -66         27.9         -63.9        -152
##   kurtosis_roll_forearm kurtosis_picth_forearm kurtosis_yaw_forearm
## 1
## 2
## 3
## 4
## 5
## 6
##   skewness_roll_forearm skewness_pitch_forearm skewness_yaw_forearm
## 1
## 2
## 3
## 4
## 5
## 6
##   max_yaw_forearm min_yaw_forearm amplitude_yaw_forearm
## 1
```

```
## 2
## 3
## 4
## 5
## 6
##   total_accel_forearm gyros_forearm_x gyros_forearm_y gyros_forearm_z
## 1                  36            0.03            0.00           -0.02
## 2                  36            0.02            0.00           -0.02
## 3                  36            0.03           -0.02            0.00
## 4                  36            0.02           -0.02            0.00
## 5                  36            0.02            0.00           -0.02
## 6                  36            0.02           -0.02           -0.03
##   accel_forearm_x accel_forearm_y accel_forearm_z magnet_forearm_x
## 1             192             203            -215              -17
## 2             192             203            -216              -18
## 3             196             204            -213              -18
## 4             189             206            -214              -16
## 5             189             206            -214              -17
## 6             193             203            -215               -9
##   magnet_forearm_y magnet_forearm_z classe
## 1              654              476      A
## 2              661              473      A
## 3              658              469      A
## 4              658              469      A
## 5              655              473      A
## 6              660              478      A
```

We started out with 160 variables and now have 93. Additionally, the first few columns do not contain information regarding how well an activity is performed and mostly serve the role of primary keys for these data. For the purposes of the analysis, they can be ignored and, hence, removed from the dataset.

```
train<- train[,-c(1:7)]
test<- test[,-c(1:7)]
dim(train)
```

```
## [1] 19622    86
```

```
dim(test)
```

```
## [1] 20 53
```

We are not left with 86 variables in the train set and 52 variables in the test set. Since we need to have the same variables in both the training and testing steps, we remove the columns from the train set that are not part of the test set.

```
names<-names(test)
names<-append(names, "classe")
train<-train[,names(train) %in% names]
dim(train)
## [1] 19622    53
```

We now have comparable datasets and now we can move onto modeling stage. We note that the train data includes the variable “classe”, which we are attempting to predict.

Splitting Data into Train and Validate Sets

We split training data into the train and validation sets in the traditional 70/30 manner. The validation set will be used to estimate an out-of-sample error.

```
library(caret)
## Warning: package 'caret' was built under R version 3.5.3
## Loading required package: lattice
## Loading required package: ggplot2
inTrain = createDataPartition(train$classe, p = 0.70)[[1]]
training= train[ inTrain,]
validation = train[-inTrain,]
```

Modeling Using Classification Tree

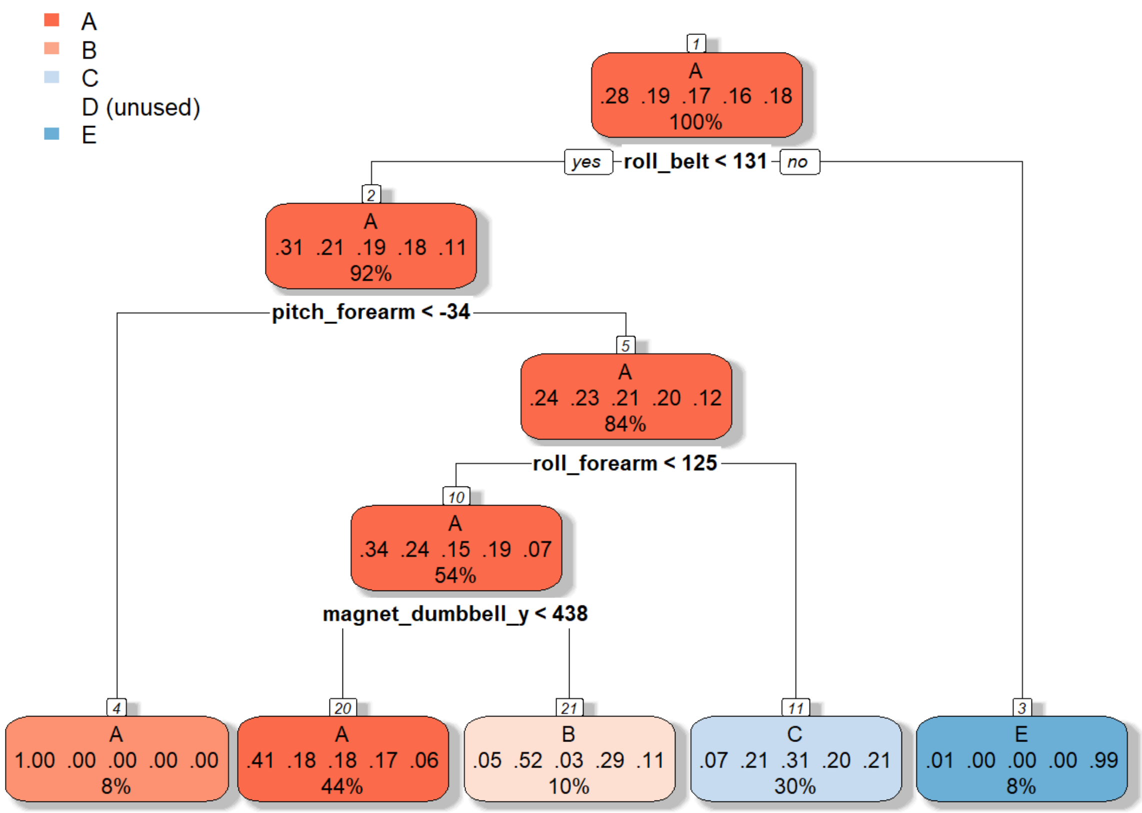
We first fit a classification tree model to these data.

```
library(rpart)
## Warning: package 'rpart' was built under R version 3.5.3
library(rpart.plot)
## Warning: package 'rpart.plot' was built under R version 3.5.3
tree_model<- train(classe~., data=training, method="rpart")
tree_model$finalModel
## n= 13737
##
```

```
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
##  1) root 13737 9831 A (0.28 0.19 0.17 0.16 0.18)
##    2) roll_belt< 130.5 12599 8703 A (0.31 0.21 0.19 0.18 0.11)
##      4) pitch_forearm< -33.95 1083    3 A (1 0.0028 0 0 0) *
##      5) pitch_forearm>=-33.95 11516 8700 A (0.24 0.23 0.21 0.2 0.12)
##       10) roll_forearm< 124.5 7375 4836 A (0.34 0.24 0.15 0.19 0.07)
##         20) magnet_dumbbell_y< 437.5 6068 3590 A (0.41 0.18 0.18 0.17 0.06
2) *
##         21) magnet_dumbbell_y>=437.5 1307  622 B (0.047 0.52 0.031 0.29 0.
11) *
##       11) roll_forearm>=124.5 4141 2875 C (0.067 0.21 0.31 0.2 0.21) *
##    3) roll_belt>=130.5 1138   10 E (0.0088 0 0 0 0.99) *
```

Next, we visualize the tree model via a graphic.

```
rpart.plot(tree_model$finalModel, box.palette="RdBu", shadow.col="gray", nn=T
RUE)
```



Note that rpart automatically performs 10-fold cross-validation when growing a tree. The variables used in the final tree are magnet_dumbbell_y, pitch_forearm, roll_bel, roll_forearm with the relative error of 71 %.

We can estimate the out of sample error by applying the trained model on the validation set.

```
validate_tree <- predict(tree_model$finalModel, validation, type="class")
confusionMatrix(validation$classe, validate_tree)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1520   25  125    0    4
##          B  480  277  382    0    0
##          C  503   13  510    0    0
##          D  425  167  372    0    0
##          E  146   73  360    0  503
##
```

```
## Overall Statistics
##
##                Accuracy : 0.4775
##                  95% CI : (0.4647, 0.4903)
##     No Information Rate : 0.5223
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.3174
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.4945  0.49910  0.29160       NA  0.99211
## Specificity            0.9452  0.83827  0.87524   0.8362  0.89234
## Pos Pred Value         0.9080  0.24320  0.49708       NA  0.46488
## Neg Pred Value         0.6310  0.94142  0.74501       NA  0.99917
## Prevalence             0.5223  0.09431  0.29720   0.0000  0.08615
## Detection Rate         0.2583  0.04707  0.08666   0.0000  0.08547
## Detection Prevalence   0.2845  0.19354  0.17434   0.1638  0.18386
## Balanced Accuracy      0.7198  0.66869  0.58342       NA  0.94222
```

The accuracy does not look too good but overall, the model is statistically significant, so we use it for the final prediction.

```
test_tree <- predict(tree_model$finalModel,test, type="class")
test_tree
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##  C  A  C  A  A  C  C  A  A  A  C  C  C  A  C  A  A  A  A  C
## Levels: A B C D E
```

Unfortunately, after submitting these results to the quiz engine, I did not pass the assignment.

Modeling with Random Forest

So, I try a random forest next because it uses an ensemble of trees so maybe it will work better. First, I train the model

```
random_forest <- train(classe ~ ., data = train, method = "rf")
print(random_forest)
```

```
## Random Forest
##
## 19622 samples
##    52 predictor
##     5 classes: 'A', 'B', 'C', 'D', 'E'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 19622, 19622, 19622, 19622, 19622, 19622, ...
## Resampling results across tuning parameters:
##
##   mtry  Accuracy   Kappa
##    2    0.9930247  0.9911747
##   27    0.9926283  0.9906736
##   52    0.9846718  0.9806069
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 2.
```

Then, estimate an out-of-sample error on the validation set and display the confusion matrix.

```
validate <- predict(random_forest, validation)
# Show prediction result
confusionMatrix(validation$classe, validate)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1674    0    0    0    0
##          B    0 1139    0    0    0
##          C    0    0 1026    0    0
```

```
##          D    0    0    0  964    0
##          E    0    0    0    0 1082
##
## Overall Statistics
##
##                Accuracy : 1
##                  95% CI : (0.9994, 1)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 1
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            1.0000   1.0000   1.0000   1.0000   1.0000
## Specificity            1.0000   1.0000   1.0000   1.0000   1.0000
## Pos Pred Value         1.0000   1.0000   1.0000   1.0000   1.0000
## Neg Pred Value         1.0000   1.0000   1.0000   1.0000   1.0000
## Prevalence             0.2845   0.1935   0.1743   0.1638   0.1839
## Detection Rate         0.2845   0.1935   0.1743   0.1638   0.1839
## Detection Prevalence   0.2845   0.1935   0.1743   0.1638   0.1839
## Balanced Accuracy      1.0000   1.0000   1.0000   1.0000   1.0000
```

Finally, pass the quiz with the following predictions

```
predict(random_forest, test)
##  [1] B A B A A E D B A A B C B A E E A B B B
## Levels: A B C D E
```

Conclusions

In this project we have examined exercise data and were able to characterize the quality of the exercise using only a few variables. We have leveraged a random forest model to achieve a high classification on both in-sample and out-of-sample data. We conclude that ensemble models rock and that the government can recruit the future generation of secret service agents by hacking into the fitness watches of the daily users of such devices!

References

In preparing this report, I have consulted the following web-sites

https://stackoverflow.com/questions/2643939/remove-columns-from-dataframe-where-all-values-are-na

https://stackoverflow.com/questions/33282174/r-caret-package-error-createdataparition-no-observation

https://blog.exploratory.io/visualizing-a-decision-tree-using-r-packages-in-explortory-b26d4cb5e71f

https://www.gormanalysis.com/blog/decision-trees-in-r-using-rpart/

https://stackoverflow.com/questions/39620287/how-to-create-a-confusion-matrix-for-a-decision-tree-model