

Coursera - Practical Machine Learning Project

Kelvin Leung

6/20/2018

Background

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, your goal will be 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).

Load libraries

```
library(randomForest)

## randomForest 4.6-14
## Type rfNews() to see new features/changes/bug fixes.

library(caret)

## Loading required package: lattice
## Loading required package: ggplot2
##
## Attaching package: 'ggplot2'
## The following object is masked from 'package:randomForest':
##
##     margin

library(rattle)

## Rattle: A free graphical interface for data science with R.
## Version 5.1.0 Copyright (c) 2006-2017 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.
##
## Attaching package: 'rattle'
## The following object is masked from 'package:randomForest':
##
##     importance

library(rpart)
library(rpart.plot)
library(RColorBrewer)
```

Data load and cleansing

```
#train_data <- read.csv(url("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv"), he
train_data <- read.csv('data/pml-training.csv', sep=',', na.strings=c("NA","#DIV/0!",""))
dim(train_data)
```

```
## [1] 19622 160
```

```
#test_data <- read.csv(url("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv"), head
test_data <- read.csv('data/pml-testing.csv', sep = ',', na.strings=c("NA","#DIV/0!",""))
dim(test_data)
```

```
## [1] 20 160
```

```
str(train_data)
```

```
## 'data.frame': 19622 obs. of 160 variables:
## $ X : int 1 2 3 4 5 6 7 8 9 10 ...
## $ user_name : Factor w/ 6 levels "adelmo","carlitos",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ raw_timestamp_part_1 : int 1323084231 1323084231 1323084231 1323084232 1323084232 1323084232 ...
## $ raw_timestamp_part_2 : int 788290 808298 820366 120339 196328 304277 368296 440390 484323 484...
## $ cvtd_timestamp : Factor w/ 20 levels "02/12/2011 13:32",...: 9 9 9 9 9 9 9 9 9 9 ...
## $ new_window : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
## $ num_window : int 11 11 11 12 12 12 12 12 12 12 ...
## $ roll_belt : num 1.41 1.41 1.42 1.48 1.48 1.45 1.42 1.42 1.43 1.45 ...
## $ pitch_belt : num 8.07 8.07 8.07 8.05 8.07 8.06 8.09 8.13 8.16 8.17 ...
## $ yaw_belt : num -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 ...
## $ total_accel_belt : int 3 3 3 3 3 3 3 3 3 3 ...
## $ kurtosis_roll_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_pitch_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_yaw_belt : logi NA NA NA NA NA NA ...
## $ skewness_roll_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_roll_belt.1 : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_yaw_belt : logi NA NA NA NA NA NA ...
## $ max_roll_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ max_pitch_belt : int NA NA NA NA NA NA NA NA NA NA ...
## $ max_yaw_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_roll_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_pitch_belt : int NA NA NA NA NA NA NA NA NA NA ...
## $ min_yaw_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_roll_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_pitch_belt : int NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_yaw_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_total_accel_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_roll_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_roll_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_roll_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_pitch_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_pitch_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_pitch_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_yaw_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_yaw_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_yaw_belt : num NA NA NA NA NA NA NA NA NA NA ...
## $ gyros_belt_x : num 0 0.02 0 0.02 0.02 0.02 0.02 0.02 0.02 0.03 ...
## $ gyros_belt_y : num 0 0 0 0 0.02 0 0 0 0 0 ...
```

```

## $ gyros_belt_z      : num -0.02 -0.02 -0.02 -0.03 -0.02 -0.02 -0.02 -0.02 -0.02 0 ...
## $ accel_belt_x      : int -21 -22 -20 -22 -21 -21 -22 -22 -20 -21 ...
## $ accel_belt_y      : int 4 4 5 3 2 4 3 4 2 4 ...
## $ accel_belt_z      : int 22 22 23 21 24 21 21 21 24 22 ...
## $ magnet_belt_x     : int -3 -7 -2 -6 -6 0 -4 -2 1 -3 ...
## $ magnet_belt_y     : int 599 608 600 604 600 603 599 603 602 609 ...
## $ magnet_belt_z     : int -313 -311 -305 -310 -302 -312 -311 -313 -312 -308 ...
## $ roll_arm          : num -128 -128 -128 -128 -128 -128 -128 -128 -128 -128 ...
## $ pitch_arm         : num 22.5 22.5 22.5 22.1 22.1 22 21.9 21.8 21.7 21.6 ...
## $ yaw_arm           : num -161 -161 -161 -161 -161 -161 -161 -161 -161 -161 ...
## $ total_accel_arm   : int 34 34 34 34 34 34 34 34 34 34 ...
## $ var_accel_arm     : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_roll_arm      : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_roll_arm   : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_roll_arm      : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_pitch_arm     : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_pitch_arm  : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_pitch_arm     : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_yaw_arm       : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_yaw_arm    : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_yaw_arm       : num NA NA NA NA NA NA NA NA NA NA ...
## $ gyros_arm_x       : num 0 0.02 0.02 0.02 0 0.02 0 0.02 0.02 0.02 ...
## $ gyros_arm_y       : num 0 -0.02 -0.02 -0.03 -0.03 -0.03 -0.03 -0.02 -0.03 -0.03 ...
## $ gyros_arm_z       : num -0.02 -0.02 -0.02 0.02 0 0 0 0 -0.02 -0.02 ...
## $ accel_arm_x       : int -288 -290 -289 -289 -289 -289 -289 -289 -288 -288 ...
## $ accel_arm_y       : int 109 110 110 111 111 111 111 111 109 110 ...
## $ accel_arm_z       : int -123 -125 -126 -123 -123 -122 -125 -124 -122 -124 ...
## $ magnet_arm_x      : int -368 -369 -368 -372 -374 -369 -373 -372 -369 -376 ...
## $ magnet_arm_y      : int 337 337 344 344 337 342 336 338 341 334 ...
## $ magnet_arm_z      : int 516 513 513 512 506 513 509 510 518 516 ...
## $ kurtosis_roll_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_pitch_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_yaw_arm  : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_roll_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_pitch_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_yaw_arm  : num NA NA NA NA NA NA NA NA NA NA ...
## $ max_roll_arm      : num NA NA NA NA NA NA NA NA NA NA ...
## $ max_pitch_arm     : num NA NA NA NA NA NA NA NA NA NA ...
## $ max_yaw_arm       : int NA NA NA NA NA NA NA NA NA NA ...
## $ min_roll_arm      : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_pitch_arm     : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_yaw_arm       : int NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_roll_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_pitch_arm : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_yaw_arm  : int NA NA NA NA NA NA NA NA NA NA ...
## $ roll_dumbbell     : num 13.1 13.1 12.9 13.4 13.4 ...
## $ pitch_dumbbell    : num -70.5 -70.6 -70.3 -70.4 -70.4 ...
## $ yaw_dumbbell      : num -84.9 -84.7 -85.1 -84.9 -84.9 ...
## $ kurtosis_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_pitch_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_yaw_dumbbell : logi NA NA NA NA NA NA ...
## $ skewness_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_pitch_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_yaw_dumbbell : logi NA NA NA NA NA NA ...

```

```
## $ max_roll_dumbbell      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ max_pitch_dumbbell     : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ max_yaw_dumbbell       : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ min_roll_dumbbell      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ min_pitch_dumbbell     : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ min_yaw_dumbbell       : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_roll_dumbbell : num  NA NA NA NA NA NA NA NA NA NA NA ...
## [list output truncated]
```

The training set has 19622 observations and each observation has 160 columns. We notice that many columns have N/A values or blank values. So we will remove them because they will not produce any information. Also, the first seven columns give information about the people who did the test and the timestamps. We will remove these columns in our model. Note, the “classe” variable is in the last column of our training set.

The testing set has 20 cases. It will be used to test the accuracy of our models.

Cleansing procedure

Here is the R code to remove the columns that has N/A or “” values.

```
# removing columns having value of "N/A" or "" value that have at least 90% of the total number of rows
tidx_2remove <- which(colSums(is.na(train_data) | train_data=="") > 0.9* dim(train_data)[1])
# removing those columns
train_clean <- train_data[, -tidx_2remove]
# removing the first 7 columns that are irrelevant to the prediction model
train_clean <- train_data[, -(1:7)]
str(train_clean)
```

```
## 'data.frame':   19622 obs. of  153 variables:
## $ roll_belt      : num  1.41 1.41 1.42 1.48 1.48 1.45 1.42 1.42 1.43 1.45 ...
## $ pitch_belt     : num  8.07 8.07 8.07 8.05 8.07 8.06 8.09 8.13 8.16 8.17 ...
## $ yaw_belt       : num  -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 -94.4 ...
## $ total_accel_belt : int   3 3 3 3 3 3 3 3 3 3 ...
## $ kurtosis_roll_belt : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_pitch_belt : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_yaw_belt  : logi  NA NA NA NA NA NA ...
## $ skewness_roll_belt : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_roll_belt.1 : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_yaw_belt  : logi  NA NA NA NA NA NA ...
## $ max_roll_belt     : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ max_pitch_belt    : int   NA NA NA NA NA NA NA NA NA NA NA ...
## $ max_yaw_belt      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ min_roll_belt     : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ min_pitch_belt    : int   NA NA NA NA NA NA NA NA NA NA NA ...
## $ min_yaw_belt      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_roll_belt : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_pitch_belt : int   NA NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_yaw_belt : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ var_total_accel_belt : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ avg_roll_belt     : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_roll_belt  : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ var_roll_belt     : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ avg_pitch_belt    : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_pitch_belt : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ var_pitch_belt    : num  NA NA NA NA NA NA NA NA NA NA NA ...
```

```

## $ avg_yaw_belt      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_yaw_belt   : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ var_yaw_belt      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ gyros_belt_x      : num  0 0.02 0 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.03 ...
## $ gyros_belt_y      : num  0 0 0 0 0.02 0 0 0 0 0 ...
## $ gyros_belt_z      : num  -0.02 -0.02 -0.02 -0.03 -0.02 -0.02 -0.02 -0.02 -0.02 0 ...
## $ accel_belt_x      : int   -21 -22 -20 -22 -21 -21 -22 -22 -20 -21 ...
## $ accel_belt_y      : int    4 4 5 3 2 4 3 4 2 4 ...
## $ accel_belt_z      : int   22 22 23 21 24 21 21 21 24 22 ...
## $ magnet_belt_x     : int   -3 -7 -2 -6 -6 0 -4 -2 1 -3 ...
## $ magnet_belt_y     : int  599 608 600 604 600 603 599 603 602 609 ...
## $ magnet_belt_z     : int  -313 -311 -305 -310 -302 -312 -311 -313 -312 -308 ...
## $ roll_arm          : num  -128 -128 -128 -128 -128 -128 -128 -128 -128 -128 ...
## $ pitch_arm         : num  22.5 22.5 22.5 22.1 22.1 22 21.9 21.8 21.7 21.6 ...
## $ yaw_arm           : num  -161 -161 -161 -161 -161 -161 -161 -161 -161 -161 ...
## $ total_accel_arm   : int   34 34 34 34 34 34 34 34 34 34 ...
## $ var_accel_arm     : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ avg_roll_arm      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_roll_arm   : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ var_roll_arm      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ avg_pitch_arm     : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_pitch_arm  : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ var_pitch_arm     : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ avg_yaw_arm       : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_yaw_arm    : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ var_yaw_arm       : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ gyros_arm_x       : num  0 0.02 0.02 0.02 0 0.02 0 0.02 0.02 0.02 ...
## $ gyros_arm_y       : num  0 -0.02 -0.02 -0.03 -0.03 -0.03 -0.03 -0.02 -0.03 -0.03 ...
## $ gyros_arm_z       : num  -0.02 -0.02 -0.02 0.02 0 0 0 0 -0.02 -0.02 ...
## $ accel_arm_x       : int  -288 -290 -289 -289 -289 -289 -289 -289 -288 -288 ...
## $ accel_arm_y       : int  109 110 110 111 111 111 111 111 109 110 ...
## $ accel_arm_z       : int  -123 -125 -126 -123 -123 -122 -125 -124 -122 -124 ...
## $ magnet_arm_x      : int  -368 -369 -368 -372 -374 -369 -373 -372 -369 -376 ...
## $ magnet_arm_y      : int  337 337 344 344 337 342 336 338 341 334 ...
## $ magnet_arm_z      : int  516 513 513 512 506 513 509 510 518 516 ...
## $ kurtosis_roll_arm : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_pitch_arm : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_yaw_arm  : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_roll_arm : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_pitch_arm : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_yaw_arm  : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ max_roll_arm     : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ max_pitch_arm    : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ max_yaw_arm      : int   NA NA NA NA NA NA NA NA NA NA NA ...
## $ min_roll_arm     : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ min_pitch_arm    : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ min_yaw_arm      : int   NA NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_roll_arm : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_pitch_arm : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_yaw_arm : int   NA NA NA NA NA NA NA NA NA NA NA ...
## $ roll_dumbbell    : num  13.1 13.1 12.9 13.4 13.4 ...
## $ pitch_dumbbell   : num  -70.5 -70.6 -70.3 -70.4 -70.4 ...
## $ yaw_dumbbell     : num  -84.9 -84.7 -85.1 -84.9 -84.9 ...
## $ kurtosis_roll_dumbbell : num  NA NA NA NA NA NA NA NA NA NA NA ...

```

```
## $ kurtosis_pitch_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ kurtosis_yaw_dumbbell : logi NA NA NA NA NA NA ...
## $ skewness_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_pitch_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ skewness_yaw_dumbbell : logi NA NA NA NA NA NA ...
## $ max_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ max_pitch_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ max_yaw_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_pitch_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ min_yaw_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_pitch_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ amplitude_yaw_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ total_accel_dumbbell : int 37 37 37 37 37 37 37 37 37 37 ...
## $ var_accel_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ avg_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ stddev_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## $ var_roll_dumbbell : num NA NA NA NA NA NA NA NA NA NA ...
## [list output truncated]
```

```
# removing columns having value of "N/A" or "" value that have at least 90% of the total number of rows
tidx_2remove <- which(colSums(is.na(test_data) | test_data=="") > 0.9* dim(test_data)[1])
# removing those columns
test_clean <- test_data[, -tidx_2remove]
# removing the first 7 columns that are irrelevant to the prediction model
test_clean <- test_clean[, -(1:7)]
str(test_clean)
```

```
## 'data.frame': 20 obs. of 53 variables:
## $ roll_belt : num 123 1.02 0.87 125 1.35 -5.92 1.2 0.43 0.93 114 ...
## $ pitch_belt : num 27 4.87 1.82 -41.6 3.33 1.59 4.44 4.15 6.72 22.4 ...
## $ yaw_belt : num -4.75 -88.9 -88.5 162 -88.6 -87.7 -87.3 -88.5 -93.7 -13.1 ...
## $ total_accel_belt : int 20 4 5 17 3 4 4 4 4 18 ...
## $ gyros_belt_x : num -0.5 -0.06 0.05 0.11 0.03 0.1 -0.06 -0.18 0.1 0.14 ...
## $ gyros_belt_y : num -0.02 -0.02 0.02 0.11 0.02 0.05 0 -0.02 0 0.11 ...
## $ gyros_belt_z : num -0.46 -0.07 0.03 -0.16 0 -0.13 0 -0.03 -0.02 -0.16 ...
## $ accel_belt_x : int -38 -13 1 46 -8 -11 -14 -10 -15 -25 ...
## $ accel_belt_y : int 69 11 -1 45 4 -16 2 -2 1 63 ...
## $ accel_belt_z : int -179 39 49 -156 27 38 35 42 32 -158 ...
## $ magnet_belt_x : int -13 43 29 169 33 31 50 39 -6 10 ...
## $ magnet_belt_y : int 581 636 631 608 566 638 622 635 600 601 ...
## $ magnet_belt_z : int -382 -309 -312 -304 -418 -291 -315 -305 -302 -330 ...
## $ roll_arm : num 40.7 0 0 -109 76.1 0 0 0 -137 -82.4 ...
## $ pitch_arm : num -27.8 0 0 55 2.76 0 0 0 11.2 -63.8 ...
## $ yaw_arm : num 178 0 0 -142 102 0 0 0 -167 -75.3 ...
## $ total_accel_arm : int 10 38 44 25 29 14 15 22 34 32 ...
## $ gyros_arm_x : num -1.65 -1.17 2.1 0.22 -1.96 0.02 2.36 -3.71 0.03 0.26 ...
## $ gyros_arm_y : num 0.48 0.85 -1.36 -0.51 0.79 0.05 -1.01 1.85 -0.02 -0.5 ...
## $ gyros_arm_z : num -0.18 -0.43 1.13 0.92 -0.54 -0.07 0.89 -0.69 -0.02 0.79 ...
## $ accel_arm_x : int 16 -290 -341 -238 -197 -26 99 -98 -287 -301 ...
## $ accel_arm_y : int 38 215 245 -57 200 130 79 175 111 -42 ...
## $ accel_arm_z : int 93 -90 -87 6 -30 -19 -67 -78 -122 -80 ...
## $ magnet_arm_x : int -326 -325 -264 -173 -170 396 702 535 -367 -420 ...
## $ magnet_arm_y : int 385 447 474 257 275 176 15 215 335 294 ...
```

```

## $ magnet_arm_z      : int  481 434 413 633 617 516 217 385 520 493 ...
## $ roll_dumbbell     : num -17.7 54.5 57.1 43.1 -101.4 ...
## $ pitch_dumbbell    : num  25 -53.7 -51.4 -30 -53.4 ...
## $ yaw_dumbbell      : num 126.2 -75.5 -75.2 -103.3 -14.2 ...
## $ total_accel_dumbbell: int  9 31 29 18 4 29 29 29 3 2 ...
## $ gyros_dumbbell_x   : num  0.64 0.34 0.39 0.1 0.29 -0.59 0.34 0.37 0.03 0.42 ...
## $ gyros_dumbbell_y   : num  0.06 0.05 0.14 -0.02 -0.47 0.8 0.16 0.14 -0.21 0.51 ...
## $ gyros_dumbbell_z   : num -0.61 -0.71 -0.34 0.05 -0.46 1.1 -0.23 -0.39 -0.21 -0.03 ...
## $ accel_dumbbell_x   : int  21 -153 -141 -51 -18 -138 -145 -140 0 -7 ...
## $ accel_dumbbell_y   : int -15 155 155 72 -30 166 150 159 25 -20 ...
## $ accel_dumbbell_z   : int  81 -205 -196 -148 -5 -186 -190 -191 9 7 ...
## $ magnet_dumbbell_x  : int  523 -502 -506 -576 -424 -543 -484 -515 -519 -531 ...
## $ magnet_dumbbell_y  : int -528 388 349 238 252 262 354 350 348 321 ...
## $ magnet_dumbbell_z  : int -56 -36 41 53 312 96 97 53 -32 -164 ...
## $ roll_forearm       : num  141 109 131 0 -176 150 155 -161 15.5 13.2 ...
## $ pitch_forearm      : num  49.3 -17.6 -32.6 0 -2.16 1.46 34.5 43.6 -63.5 19.4 ...
## $ yaw_forearm        : num  156 106 93 0 -47.9 89.7 152 -89.5 -139 -105 ...
## $ total_accel_forearm : int  33 39 34 43 24 43 32 47 36 24 ...
## $ gyros_forearm_x    : num  0.74 1.12 0.18 1.38 -0.75 -0.88 -0.53 0.63 0.03 0.02 ...
## $ gyros_forearm_y    : num -3.34 -2.78 -0.79 0.69 3.1 4.26 1.8 -0.74 0.02 0.13 ...
## $ gyros_forearm_z    : num -0.59 -0.18 0.28 1.8 0.8 1.35 0.75 0.49 -0.02 -0.07 ...
## $ accel_forearm_x    : int -110 212 154 -92 131 230 -192 -151 195 -212 ...
## $ accel_forearm_y    : int  267 297 271 406 -93 322 170 -331 204 98 ...
## $ accel_forearm_z    : int -149 -118 -129 -39 172 -144 -175 -282 -217 -7 ...
## $ magnet_forearm_x   : int -714 -237 -51 -233 375 -300 -678 -109 0 -403 ...
## $ magnet_forearm_y   : int  419 791 698 783 -787 800 284 -619 652 723 ...
## $ magnet_forearm_z   : int  617 873 783 521 91 884 585 -32 469 512 ...
## $ problem_id         : int  1 2 3 4 5 6 7 8 9 10 ...

```

```
names(train_clean)
```

```

## [1] "roll_belt"           "pitch_belt"
## [3] "yaw_belt"            "total_accel_belt"
## [5] "kurtosis_roll_belt"  "kurtosis_pitch_belt"
## [7] "kurtosis_yaw_belt"   "skewness_roll_belt"
## [9] "skewness_roll_belt.1" "skewness_yaw_belt"
## [11] "max_roll_belt"       "max_pitch_belt"
## [13] "max_yaw_belt"        "min_roll_belt"
## [15] "min_pitch_belt"      "min_yaw_belt"
## [17] "amplitude_roll_belt" "amplitude_pitch_belt"
## [19] "amplitude_yaw_belt"  "var_total_accel_belt"
## [21] "avg_roll_belt"       "stddev_roll_belt"
## [23] "var_roll_belt"       "avg_pitch_belt"
## [25] "stddev_pitch_belt"   "var_pitch_belt"
## [27] "avg_yaw_belt"        "stddev_yaw_belt"
## [29] "var_yaw_belt"        "gyros_belt_x"
## [31] "gyros_belt_y"        "gyros_belt_z"
## [33] "accel_belt_x"        "accel_belt_y"
## [35] "accel_belt_z"        "magnet_belt_x"
## [37] "magnet_belt_y"       "magnet_belt_z"
## [39] "roll_arm"           "pitch_arm"
## [41] "yaw_arm"            "total_accel_arm"
## [43] "var_accel_arm"       "avg_roll_arm"
## [45] "stddev_roll_arm"     "var_roll_arm"
## [47] "avg_pitch_arm"       "stddev_pitch_arm"

```

## [49]	"var_pitch_arm"	"avg_yaw_arm"
## [51]	"stddev_yaw_arm"	"var_yaw_arm"
## [53]	"gyros_arm_x"	"gyros_arm_y"
## [55]	"gyros_arm_z"	"accel_arm_x"
## [57]	"accel_arm_y"	"accel_arm_z"
## [59]	"magnet_arm_x"	"magnet_arm_y"
## [61]	"magnet_arm_z"	"kurtosis_roll_arm"
## [63]	"kurtosis_pitch_arm"	"kurtosis_yaw_arm"
## [65]	"skewness_roll_arm"	"skewness_pitch_arm"
## [67]	"skewness_yaw_arm"	"max_roll_arm"
## [69]	"max_pitch_arm"	"max_yaw_arm"
## [71]	"min_roll_arm"	"min_pitch_arm"
## [73]	"min_yaw_arm"	"amplitude_roll_arm"
## [75]	"amplitude_pitch_arm"	"amplitude_yaw_arm"
## [77]	"roll_dumbbell"	"pitch_dumbbell"
## [79]	"yaw_dumbbell"	"kurtosis_roll_dumbbell"
## [81]	"kurtosis_pitch_dumbbell"	"kurtosis_yaw_dumbbell"
## [83]	"skewness_roll_dumbbell"	"skewness_pitch_dumbbell"
## [85]	"skewness_yaw_dumbbell"	"max_roll_dumbbell"
## [87]	"max_pitch_dumbbell"	"max_yaw_dumbbell"
## [89]	"min_roll_dumbbell"	"min_pitch_dumbbell"
## [91]	"min_yaw_dumbbell"	"amplitude_roll_dumbbell"
## [93]	"amplitude_pitch_dumbbell"	"amplitude_yaw_dumbbell"
## [95]	"total_accel_dumbbell"	"var_accel_dumbbell"
## [97]	"avg_roll_dumbbell"	"stddev_roll_dumbbell"
## [99]	"var_roll_dumbbell"	"avg_pitch_dumbbell"
## [101]	"stddev_pitch_dumbbell"	"var_pitch_dumbbell"
## [103]	"avg_yaw_dumbbell"	"stddev_yaw_dumbbell"
## [105]	"var_yaw_dumbbell"	"gyros_dumbbell_x"
## [107]	"gyros_dumbbell_y"	"gyros_dumbbell_z"
## [109]	"accel_dumbbell_x"	"accel_dumbbell_y"
## [111]	"accel_dumbbell_z"	"magnet_dumbbell_x"
## [113]	"magnet_dumbbell_y"	"magnet_dumbbell_z"
## [115]	"roll_forearm"	"pitch_forearm"
## [117]	"yaw_forearm"	"kurtosis_roll_forearm"
## [119]	"kurtosis_pitch_forearm"	"kurtosis_yaw_forearm"
## [121]	"skewness_roll_forearm"	"skewness_pitch_forearm"
## [123]	"skewness_yaw_forearm"	"max_roll_forearm"
## [125]	"max_pitch_forearm"	"max_yaw_forearm"
## [127]	"min_roll_forearm"	"min_pitch_forearm"
## [129]	"min_yaw_forearm"	"amplitude_roll_forearm"
## [131]	"amplitude_pitch_forearm"	"amplitude_yaw_forearm"
## [133]	"total_accel_forearm"	"var_accel_forearm"
## [135]	"avg_roll_forearm"	"stddev_roll_forearm"
## [137]	"var_roll_forearm"	"avg_pitch_forearm"
## [139]	"stddev_pitch_forearm"	"var_pitch_forearm"
## [141]	"avg_yaw_forearm"	"stddev_yaw_forearm"
## [143]	"var_yaw_forearm"	"gyros_forearm_x"
## [145]	"gyros_forearm_y"	"gyros_forearm_z"
## [147]	"accel_forearm_x"	"accel_forearm_y"
## [149]	"accel_forearm_z"	"magnet_forearm_x"
## [151]	"magnet_forearm_y"	"magnet_forearm_z"
## [153]	"classe"	


```
rm(train_data)
```

```
names(test_clean)
```

```
## [1] "roll_belt"          "pitch_belt"          "yaw_belt"
## [4] "total_accel_belt"   "gyros_belt_x"        "gyros_belt_y"
## [7] "gyros_belt_z"      "accel_belt_x"        "accel_belt_y"
## [10] "accel_belt_z"      "magnet_belt_x"       "magnet_belt_y"
## [13] "magnet_belt_z"     "roll_arm"           "pitch_arm"
## [16] "yaw_arm"           "total_accel_arm"     "gyros_arm_x"
## [19] "gyros_arm_y"       "gyros_arm_z"        "accel_arm_x"
## [22] "accel_arm_y"       "accel_arm_z"        "magnet_arm_x"
## [25] "magnet_arm_y"      "magnet_arm_z"       "roll_dumbbell"
## [28] "pitch_dumbbell"    "yaw_dumbbell"       "total_accel_dumbbell"
## [31] "gyros_dumbbell_x"  "gyros_dumbbell_y"   "gyros_dumbbell_z"
## [34] "accel_dumbbell_x"  "accel_dumbbell_y"   "accel_dumbbell_z"
## [37] "magnet_dumbbell_x" "magnet_dumbbell_y"  "magnet_dumbbell_z"
## [40] "roll_forearm"      "pitch_forearm"      "yaw_forearm"
## [43] "total_accel_forearm" "gyros_forearm_x"    "gyros_forearm_y"
## [46] "gyros_forearm_z"   "accel_forearm_x"     "accel_forearm_y"
## [49] "accel_forearm_z"   "magnet_forearm_x"    "magnet_forearm_y"
## [52] "magnet_forearm_z"  "problem_id"
```

```
rm(test_data)
```

From the above, the columns of “train_clean” match with the columns of “test_clean”, except for the last column, “problem_id”. But we do not need to care about the “problem_id” column at this time.

We use “decision tree” and “random forest” to build classification model.

First, we partition the training data into 2 parts and use cross validation method to validate the model we build.

To ensure the reproductivity of this experiment, we initial the seed to 12345.

```
set.seed(12345)
tr1 <- createDataPartition(train_clean$classe, p=0.6, list=FALSE)
training <- train_clean[tr1,]
testing <- train_clean[-tr1,]
```

```
dim(training)
```

```
## [1] 11776 153
```

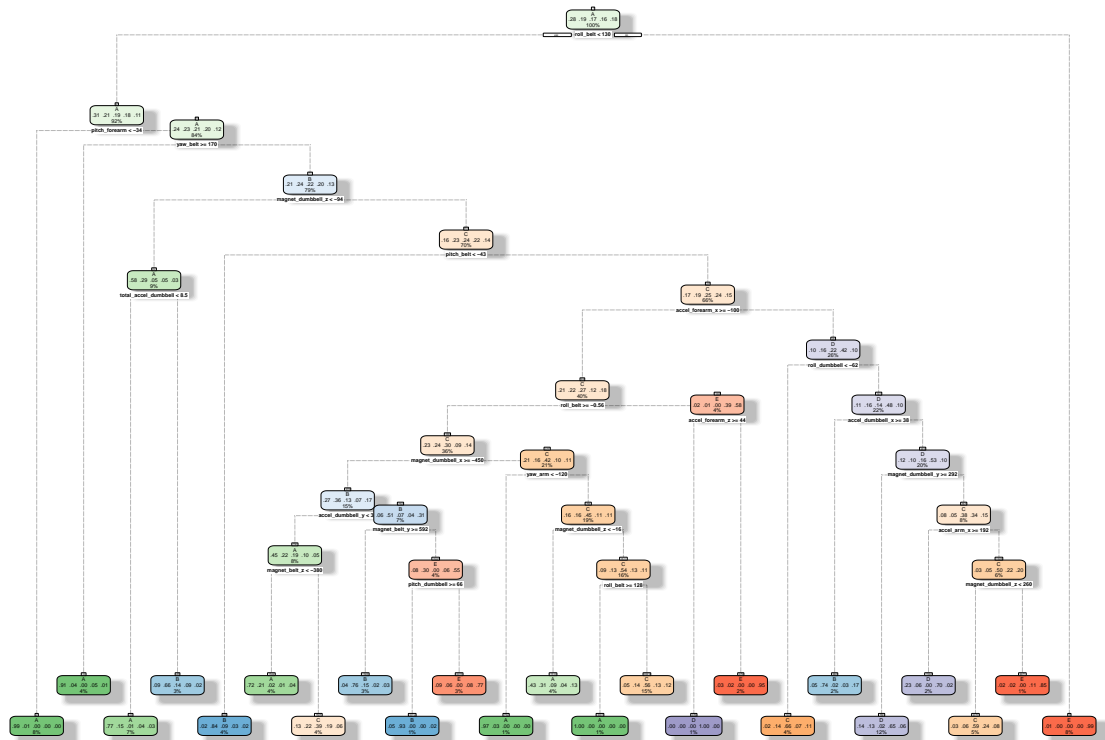
```
dim(testing)
```

```
## [1] 7846 153
```

Train with decision tree

```
mDT <- rpart(classe~., data=training, method='class')
fancyRpartPlot(mDT)
```

```
## Warning: labs do not fit even at cex 0.15, there may be some overplotting
```



Rattle 2018-Jun-26 17:34:49 kelvin

```
ctree_pred <- predict(mDT, newdata=testing, type='class')
cm_ctree <- confusionMatrix(ctree_pred, testing$classe)
cm_ctree$table
```

```
##           Reference
## Prediction   A    B    C    D    E
##           A 1879  260   30   69   66
##           B   56  759   88   34   54
##           C  105  340 1226  354  234
##           D  155  132   23  807   57
##           E   37   27    1   22 1031
```