Practical Machine Learning Course Project

Elena Fedorova

22 September 2018

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.

The data from accelerometers are preliminarily collected from the provided sources: The training data: https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv The test data: https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv

Ultimately, the prediction model will be used to predict 20 different test cases, each result of which will be written to the separate txt file. The outputs are used as answers to the final quiz.

Initial preparation

For the purposes of the project the following libraries in R are going to be used. The first step is to load them:

```
library(caret)
library(rpart)
library(rpart.plot)
library(randomForest)
library(corrplot)
library(dplyr)
```

Reading the data into R

```
train_data <- read.csv("./data/pml-training.csv")
test_data <- read.csv("./data/pml-testing.csv")
```

Exploring the data

```
glimpse(train_data)
```

```
## Observations: 19,622
## Variables: 160
```

```
## $ X                       <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12...
## $ user_name               <fct> carlitos, carlitos, carlitos, carlito...
## $ raw_timestamp_part_1    <int> 1323084231, 1323084231, 1323084231, 1...
## $ raw_timestamp_part_2    <int> 788290, 808298, 820366, 120339, 19632...
## $ cvtd_timestamp          <fct> 05/12/2011 11:23, 05/12/2011 11:23, 0...
## $ new_window              <fct> no, no, no, no, no, no, no, no, no, n...
## $ num_window              <int> 11, 11, 11, 12, 12, 12, 12, 12, 12, 1...
## $ roll_belt               <dbl> 1.41, 1.41, 1.42, 1.48, 1.48, 1.45, 1...
## $ pitch_belt              <dbl> 8.07, 8.07, 8.07, 8.05, 8.07, 8.06, 8...
## $ yaw_belt                <dbl> -94.4, -94.4, -94.4, -94.4, -94.4, -9...
## $ total_accel_belt        <int> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3...
## $ kurtosis_roll_belt      <fct> , , , , , , , , , , , , , , , , , , ,...
## $ kurtosis_picth_belt     <fct> , , , , , , , , , , , , , , , , , , ,...
## $ kurtosis_yaw_belt       <fct> , , , , , , , , , , , , , , , , , , ,...
## $ skewness_roll_belt      <fct> , , , , , , , , , , , , , , , , , , ,...
## $ skewness_roll_belt.1    <fct> , , , , , , , , , , , , , , , , , , ,...
## $ skewness_yaw_belt       <fct> , , , , , , , , , , , , , , , , , , ,...
## $ max_roll_belt           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_picth_belt          <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_yaw_belt            <fct> , , , , , , , , , , , , , , , , , , ,...
## $ min_roll_belt           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_pitch_belt          <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_yaw_belt            <fct> , , , , , , , , , , , , , , , , , , ,...
## $ amplitude_roll_belt     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_pitch_belt    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_yaw_belt      <fct> , , , , , , , , , , , , , , , , , , ,...
## $ var_total_accel_belt    <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_roll_belt           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_roll_belt        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_roll_belt           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_pitch_belt          <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_pitch_belt       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_pitch_belt          <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_yaw_belt            <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_yaw_belt         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_yaw_belt            <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ gyros_belt_x            <dbl> 0.00, 0.02, 0.00, 0.02, 0.02, 0.02, 0...
## $ gyros_belt_y            <dbl> 0.00, 0.00, 0.00, 0.00, 0.02, 0.00, 0...
## $ gyros_belt_z            <dbl> -0.02, -0.02, -0.02, -0.03, -0.02, -0...
## $ accel_belt_x            <int> -21, -22, -20, -22, -21, -21, -22, -2...
## $ accel_belt_y            <int> 4, 4, 5, 3, 2, 4, 3, 4, 2, 4, 2, 2, 4...
## $ accel_belt_z            <int> 22, 22, 23, 21, 24, 21, 21, 21, 24, 2...
## $ 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, 60...
## $ magnet_belt_z           <int> -313, -311, -305, -310, -302, -312, -...
## $ roll_arm                <dbl> -128, -128, -128, -128, -128, -128, -...
## $ pitch_arm               <dbl> 22.5, 22.5, 22.5, 22.1, 22.1, 22.0, 2...
## $ yaw_arm                 <dbl> -161, -161, -161, -161, -161, -161, -...
## $ total_accel_arm         <int> 34, 34, 34, 34, 34, 34, 34, 34, 34, 3...
## $ var_accel_arm           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_roll_arm            <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_roll_arm         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
```

```
## $ var_roll_arm             <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_pitch_arm            <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_pitch_arm         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_pitch_arm            <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_yaw_arm              <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_yaw_arm           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_yaw_arm              <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ gyros_arm_x              <dbl> 0.00, 0.02, 0.02, 0.02, 0.00, 0.02, 0...
## $ gyros_arm_y              <dbl> 0.00, -0.02, -0.02, -0.03, -0.03, -0....
## $ gyros_arm_z              <dbl> -0.02, -0.02, -0.02, 0.02, 0.00, 0.00...
## $ accel_arm_x              <int> -288, -290, -289, -289, -289, -289, -...
## $ accel_arm_y              <int> 109, 110, 110, 111, 111, 111, 111, 11...
## $ accel_arm_z              <int> -123, -125, -126, -123, -123, -122, -...
## $ magnet_arm_x             <int> -368, -369, -368, -372, -374, -369, -...
## $ magnet_arm_y             <int> 337, 337, 344, 344, 337, 342, 336, 33...
## $ magnet_arm_z             <int> 516, 513, 513, 512, 506, 513, 509, 51...
## $ kurtosis_roll_arm        <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ kurtosis_picth_arm       <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ kurtosis_yaw_arm         <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ skewness_roll_arm        <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ skewness_pitch_arm       <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ skewness_yaw_arm         <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ max_roll_arm             <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_picth_arm            <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_yaw_arm              <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_roll_arm             <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_pitch_arm            <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_yaw_arm              <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_roll_arm       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_pitch_arm      <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_yaw_arm        <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ roll_dumbbell            <dbl> 13.05217, 13.13074, 12.85075, 13.4312...
## $ pitch_dumbbell           <dbl> -70.49400, -70.63751, -70.27812, -70....
## $ yaw_dumbbell             <dbl> -84.87394, -84.71065, -85.14078, -84....
## $ kurtosis_roll_dumbbell   <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ kurtosis_picth_dumbbell  <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ kurtosis_yaw_dumbbell    <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ skewness_roll_dumbbell   <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ skewness_pitch_dumbbell  <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ skewness_yaw_dumbbell    <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ max_roll_dumbbell        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_picth_dumbbell       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_yaw_dumbbell         <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ min_roll_dumbbell        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_pitch_dumbbell       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_yaw_dumbbell         <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ amplitude_roll_dumbbell  <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_pitch_dumbbell <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_yaw_dumbbell   <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ total_accel_dumbbell     <int> 37, 37, 37, 37, 37, 37, 37, 37, 37, 3...
## $ var_accel_dumbbell       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_roll_dumbbell        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
```

```
## $ stddev_roll_dumbbell     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_roll_dumbbell        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_pitch_dumbbell       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_pitch_dumbbell    <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_pitch_dumbbell       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_yaw_dumbbell         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_yaw_dumbbell      <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_yaw_dumbbell         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ gyros_dumbbell_x         <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0...
## $ gyros_dumbbell_y         <dbl> -0.02, -0.02, -0.02, -0.02, -0.02, -0...
## $ gyros_dumbbell_z         <dbl> 0.00, 0.00, 0.00, -0.02, 0.00, 0.00, ...
## $ accel_dumbbell_x         <int> -234, -233, -232, -232, -233, -234, -...
## $ accel_dumbbell_y         <int> 47, 47, 46, 48, 48, 48, 47, 46, 47, 4...
## $ accel_dumbbell_z         <int> -271, -269, -270, -269, -270, -269, -...
## $ magnet_dumbbell_x        <int> -559, -555, -561, -552, -554, -558, -...
## $ magnet_dumbbell_y        <int> 293, 296, 298, 303, 292, 294, 295, 30...
## $ magnet_dumbbell_z        <dbl> -65, -64, -63, -60, -68, -66, -70, -7...
## $ roll_forearm             <dbl> 28.4, 28.3, 28.3, 28.1, 28.0, 27.9, 2...
## $ pitch_forearm            <dbl> -63.9, -63.9, -63.9, -63.9, -63.9, -6...
## $ yaw_forearm              <dbl> -153, -153, -152, -152, -152, -152, -...
## $ kurtosis_roll_forearm    <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ kurtosis_picth_forearm   <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ kurtosis_yaw_forearm     <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ skewness_roll_forearm    <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ skewness_pitch_forearm   <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ skewness_yaw_forearm     <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ max_roll_forearm         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_picth_forearm        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_yaw_forearm          <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ min_roll_forearm         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_pitch_forearm        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_yaw_forearm          <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ amplitude_roll_forearm   <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_pitch_forearm  <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_yaw_forearm    <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ total_accel_forearm      <int> 36, 36, 36, 36, 36, 36, 36, 36, 36, 3...
## $ var_accel_forearm        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_roll_forearm         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_roll_forearm      <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_roll_forearm         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_pitch_forearm        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_pitch_forearm     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_pitch_forearm        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_yaw_forearm          <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_yaw_forearm       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_yaw_forearm          <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ gyros_forearm_x          <dbl> 0.03, 0.02, 0.03, 0.02, 0.02, 0.02, 0...
## $ gyros_forearm_y          <dbl> 0.00, 0.00, -0.02, -0.02, 0.00, -0.02...
## $ gyros_forearm_z          <dbl> -0.02, -0.02, 0.00, 0.00, -0.02, -0.0...
## $ accel_forearm_x          <int> 192, 192, 196, 189, 189, 193, 195, 19...
## $ accel_forearm_y          <int> 203, 203, 204, 206, 206, 203, 205, 20...
## $ accel_forearm_z          <int> -215, -216, -213, -214, -214, -215, -...
```

```
## $ magnet_forearm_x         <int> -17, -18, -18, -16, -17, -9, -18, -9,...
## $ magnet_forearm_y         <dbl> 654, 661, 658, 658, 655, 660, 659, 66...
## $ magnet_forearm_z         <dbl> 476, 473, 469, 469, 473, 478, 470, 47...
## $ classe                   <fct> A, A, A, A, A, A, A, A, A, A, A, A, A...
```

```
glimpse(test_data)
```

```
## Observations: 20
## Variables: 160
## $ X                        <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12...
## $ user_name                <fct> pedro, jeremy, jeremy, adelmo, eurico...
## $ raw_timestamp_part_1     <int> 1323095002, 1322673067, 1322673075, 1...
## $ raw_timestamp_part_2     <int> 868349, 778725, 342967, 560311, 81477...
## $ cvtd_timestamp           <fct> 05/12/2011 14:23, 30/11/2011 17:11, 3...
## $ new_window               <fct> no, no, no, no, no, no, no, no, no, n...
## $ num_window               <int> 74, 431, 439, 194, 235, 504, 485, 440...
## $ roll_belt                <dbl> 123.00, 1.02, 0.87, 125.00, 1.35, -5....
## $ pitch_belt               <dbl> 27.00, 4.87, 1.82, -41.60, 3.33, 1.59...
## $ yaw_belt                 <dbl> -4.75, -88.90, -88.50, 162.00, -88.60...
## $ total_accel_belt         <int> 20, 4, 5, 17, 3, 4, 4, 4, 4, 18, 3, 5...
## $ kurtosis_roll_belt       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ kurtosis_picth_belt      <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ kurtosis_yaw_belt        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ skewness_roll_belt       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ skewness_roll_belt.1     <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ skewness_yaw_belt        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_roll_belt            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_picth_belt           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_yaw_belt             <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_roll_belt            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_pitch_belt           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_yaw_belt             <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_roll_belt      <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_pitch_belt     <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_yaw_belt       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_total_accel_belt     <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_roll_belt            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_roll_belt         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_roll_belt            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_pitch_belt           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_pitch_belt        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_pitch_belt           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_yaw_belt             <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_yaw_belt          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_yaw_belt             <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ gyros_belt_x             <dbl> -0.50, -0.06, 0.05, 0.11, 0.03, 0.10,...
## $ gyros_belt_y             <dbl> -0.02, -0.02, 0.02, 0.11, 0.02, 0.05,...
## $ gyros_belt_z             <dbl> -0.46, -0.07, 0.03, -0.16, 0.00, -0.1...
## $ accel_belt_x             <int> -38, -13, 1, 46, -8, -11, -14, -10, -...
## $ 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, 3...
## $ magnet_belt_x            <int> -13, 43, 29, 169, 33, 31, 50, 39, -6,...
## $ magnet_belt_y            <int> 581, 636, 631, 608, 566, 638, 622, 63...
```

```
## $ magnet_belt_z            <int> -382, -309, -312, -304, -418, -291, -...
## $ roll_arm                 <dbl> 40.70, 0.00, 0.00, -109.00, 76.10, 0....
## $ pitch_arm                <dbl> -27.80, 0.00, 0.00, 55.00, 2.76, 0.00...
## $ yaw_arm                  <dbl> 178.0, 0.0, 0.0, -142.0, 102.0, 0.0, ...
## $ total_accel_arm          <int> 10, 38, 44, 25, 29, 14, 15, 22, 34, 3...
## $ var_accel_arm            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_roll_arm             <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_roll_arm          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_roll_arm             <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_pitch_arm            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_pitch_arm         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_pitch_arm            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_yaw_arm              <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_yaw_arm           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_yaw_arm              <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ gyros_arm_x              <dbl> -1.65, -1.17, 2.10, 0.22, -1.96, 0.02...
## $ gyros_arm_y              <dbl> 0.48, 0.85, -1.36, -0.51, 0.79, 0.05,...
## $ gyros_arm_z              <dbl> -0.18, -0.43, 1.13, 0.92, -0.54, -0.0...
## $ accel_arm_x              <int> 16, -290, -341, -238, -197, -26, 99, ...
## $ accel_arm_y              <int> 38, 215, 245, -57, 200, 130, 79, 175,...
## $ accel_arm_z              <int> 93, -90, -87, 6, -30, -19, -67, -78, ...
## $ magnet_arm_x             <int> -326, -325, -264, -173, -170, 396, 70...
## $ magnet_arm_y             <int> 385, 447, 474, 257, 275, 176, 15, 215...
## $ magnet_arm_z             <int> 481, 434, 413, 633, 617, 516, 217, 38...
## $ kurtosis_roll_arm        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ kurtosis_picth_arm       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ kurtosis_yaw_arm         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ skewness_roll_arm        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ skewness_pitch_arm       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ skewness_yaw_arm         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_roll_arm             <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_picth_arm            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_yaw_arm              <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_roll_arm             <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_pitch_arm            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_yaw_arm              <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_roll_arm       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_pitch_arm      <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_yaw_arm        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ roll_dumbbell            <dbl> -17.737480, 54.477605, 57.070308, 43....
## $ pitch_dumbbell           <dbl> 24.96085, -53.69758, -51.37303, -30.0...
## $ yaw_dumbbell             <dbl> 126.235964, -75.514799, -75.202873, -...
## $ kurtosis_roll_dumbbell   <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ kurtosis_picth_dumbbell  <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ kurtosis_yaw_dumbbell    <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ skewness_roll_dumbbell   <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ skewness_pitch_dumbbell  <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ skewness_yaw_dumbbell    <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_roll_dumbbell        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_picth_dumbbell       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_yaw_dumbbell         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_roll_dumbbell        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
```

```
## $ min_pitch_dumbbell       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_yaw_dumbbell         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_roll_dumbbell  <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_pitch_dumbbell <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_yaw_dumbbell   <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ total_accel_dumbbell     <int> 9, 31, 29, 18, 4, 29, 29, 29, 3, 2, 1...
## $ var_accel_dumbbell       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_roll_dumbbell        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_roll_dumbbell     <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_roll_dumbbell        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_pitch_dumbbell       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_pitch_dumbbell    <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_pitch_dumbbell       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_yaw_dumbbell         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_yaw_dumbbell      <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_yaw_dumbbell         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ gyros_dumbbell_x         <dbl> 0.64, 0.34, 0.39, 0.10, 0.29, -0.59, ...
## $ gyros_dumbbell_y         <dbl> 0.06, 0.05, 0.14, -0.02, -0.47, 0.80,...
## $ gyros_dumbbell_z         <dbl> -0.61, -0.71, -0.34, 0.05, -0.46, 1.1...
## $ accel_dumbbell_x         <int> 21, -153, -141, -51, -18, -138, -145,...
## $ accel_dumbbell_y         <int> -15, 155, 155, 72, -30, 166, 150, 159...
## $ accel_dumbbell_z         <int> 81, -205, -196, -148, -5, -186, -190,...
## $ magnet_dumbbell_x        <int> 523, -502, -506, -576, -424, -543, -4...
## $ magnet_dumbbell_y        <int> -528, 388, 349, 238, 252, 262, 354, 3...
## $ magnet_dumbbell_z        <int> -56, -36, 41, 53, 312, 96, 97, 53, -3...
## $ roll_forearm             <dbl> 141.0, 109.0, 131.0, 0.0, -176.0, 150...
## $ pitch_forearm            <dbl> 49.30, -17.60, -32.60, 0.00, -2.16, 1...
## $ yaw_forearm              <dbl> 156.0, 106.0, 93.0, 0.0, -47.9, 89.7,...
## $ kurtosis_roll_forearm    <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ kurtosis_picth_forearm   <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ kurtosis_yaw_forearm     <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ skewness_roll_forearm    <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ skewness_pitch_forearm   <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ skewness_yaw_forearm     <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_roll_forearm         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_picth_forearm        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ max_yaw_forearm          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_roll_forearm         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_pitch_forearm        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ min_yaw_forearm          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_roll_forearm   <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_pitch_forearm  <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ amplitude_yaw_forearm    <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ total_accel_forearm      <int> 33, 39, 34, 43, 24, 43, 32, 47, 36, 2...
## $ var_accel_forearm        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_roll_forearm         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_roll_forearm      <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_roll_forearm         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_pitch_forearm        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ stddev_pitch_forearm     <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_pitch_forearm        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ avg_yaw_forearm          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
```

```
## $ stddev_yaw_forearm       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ var_yaw_forearm          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ gyros_forearm_x          <dbl> 0.74, 1.12, 0.18, 1.38, -0.75, -0.88,...
## $ gyros_forearm_y          <dbl> -3.34, -2.78, -0.79, 0.69, 3.10, 4.26...
## $ gyros_forearm_z          <dbl> -0.59, -0.18, 0.28, 1.80, 0.80, 1.35,...
## $ accel_forearm_x          <int> -110, 212, 154, -92, 131, 230, -192, ...
## $ accel_forearm_y          <int> 267, 297, 271, 406, -93, 322, 170, -3...
## $ accel_forearm_z          <int> -149, -118, -129, -39, 172, -144, -17...
## $ magnet_forearm_x         <int> -714, -237, -51, -233, 375, -300, -67...
## $ magnet_forearm_y         <int> 419, 791, 698, 783, -787, 800, 284, -...
## $ magnet_forearm_z         <int> 617, 873, 783, 521, 91, 884, 585, -32...
## $ problem_id               <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12...
```

Tidying the data

Removal of the missing values in both data sets:

```
sum(complete.cases(train_data))
```

```
## [1] 406
```

```
train_data <- train_data[, colSums(is.na(train_data)) == 0]
test_data <- test_data[, colSums(is.na(test_data)) == 0]
```

For the goal of the project (predict the manner in which participants did the exercise) the "classe" variable in the training set will be used. As the next step, removal of unnecessary variables from both data sets will be performed (i.e. column "time stamps" might a be good indicator of performance type, however, these data won't be useful for making prediction outside of the sample):

```
classe <- train_data$classe # Clean training data-set
train_remove <- grepl("^X|timestamp|window", names(train_data))
train_data <- train_data[, !train_remove]
train_tidy <- train_data[, sapply(train_data, is.numeric)]

train_tidy$classe <- classe # Clean testing data-set
test_remove <- grepl("^X|timestamp|window", names(test_data))
test_data <- test_data[, !test_remove]
test_tidy <- test_data[, sapply(test_data, is.numeric)]

dim(train_tidy)
```

```
## [1] 19622    53
```

```
dim(test_tidy)
```

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

As a result, there are two cleaned data sets ready for further tasks: training set now contains 19622 observations of 53 variables and testing set - 20 observations of 53 variables. These variable will be candiate predictors for the purpose of further analysis.

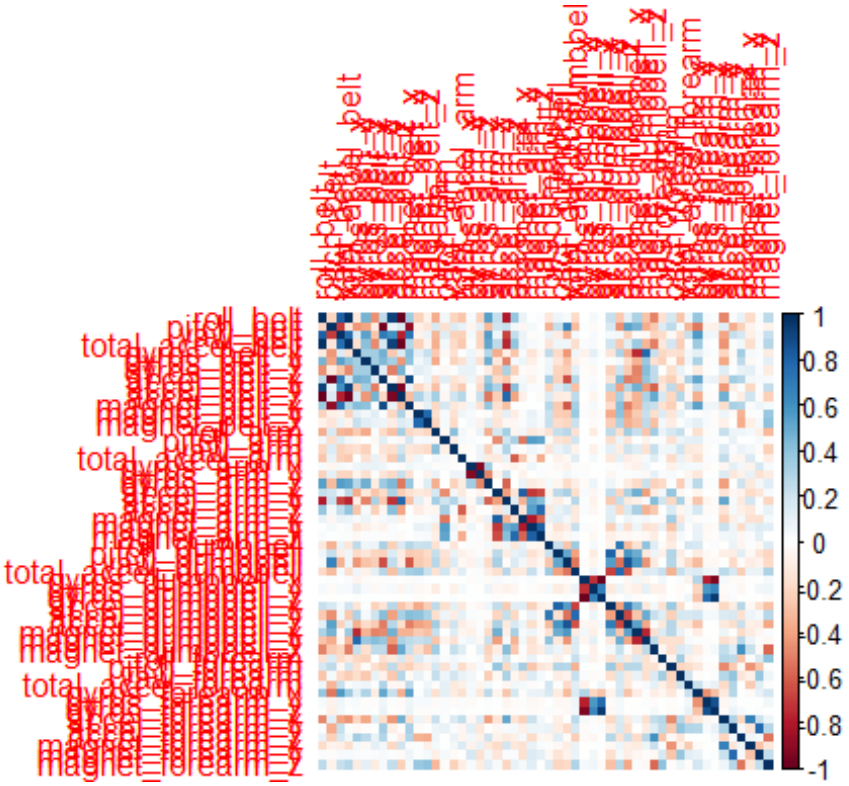
Data partition

One of the key steps of this exercise is to create data partition. The cleaned training data set will be split into training data set (70%) and a test data set (30%):

```
set.seed(23415)
inTrain <- createDataPartition(train_tidy$classe, p=0.70, list=F)
trainData <- train_tidy[inTrain, ]
testData <- train_tidy[-inTrain, ]
```

Correlation plot:

```
corrPlot <- cor(trainData[, -length(names(trainData))])
corrplot(corrPlot, method="color")
```



Prediction model development

To fit a prediction model for activity recognition, the Random Forest algorithm will be applied. This perediction algorithm is known as one of the most effective for classification problems: among the set of variables it automatically selects important ones. It is also robust to correlated covariates & outliers in general. Additionally, the 5-fold cross validation is going to be used when applying the algorithm to find the best parameters for the prediction model.

```
controlRf <- trainControl(method="cv", 5)
modelRf <- train(classe ~ .,
                 data=trainData,
                 method="rf",
```

```
                  trControl=controlRf,
                  ntree=250)
```

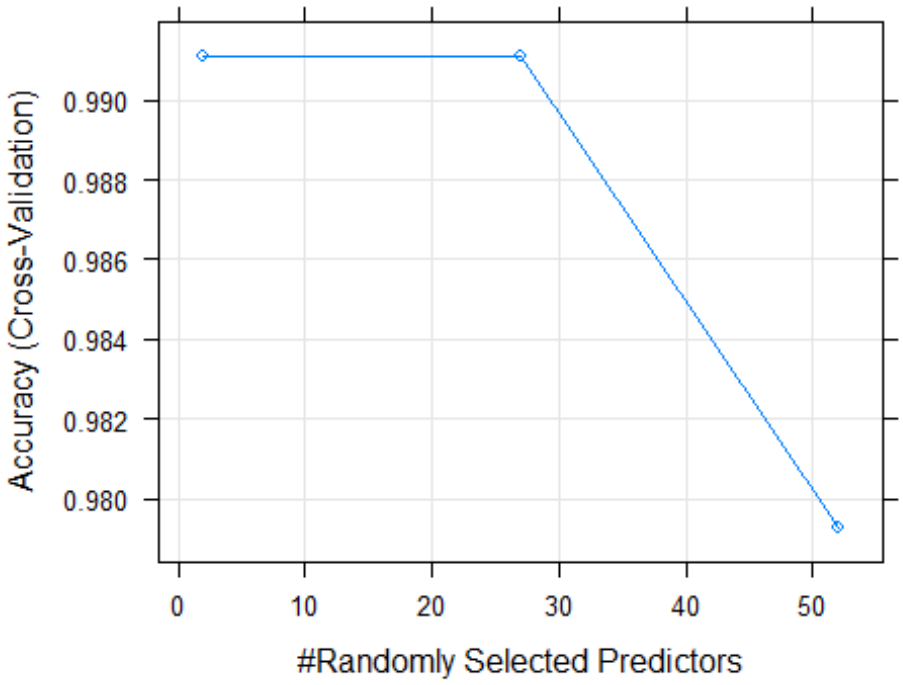
The final summary of the model output is the following:

```
modelRf
```

```
## Random Forest
##
## 13737 samples
##    52 predictor
##     5 classes: 'A', 'B', 'C', 'D', 'E'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 10989, 10990, 10990, 10990, 10989
## Resampling results across tuning parameters:
##
##   mtry  Accuracy   Kappa
##    2    0.9911187  0.9887641
##   27    0.9911189  0.9887655
##   52    0.9792532  0.9737565
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 27.
```

The estimated model accuracy accross different number of the randomly selected predictors can also be visualised in the following plot: of the values that were investigated for mtry, the best choice with the highest cross valitated accuracy is 27.

```
plot(modelRf)
```



Estimation of the performance of the model on the validation data set:

```
predictRf <- predict(modelRf, testData)
confusionMatrix(testData$classe, predictRf)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1673    0    1    0    0
##          B    7 1131    1    0    0
##          C    0    7 1015    4    0
##          D    1    1   13  949    0
##          E    0    0    0    2 1080
##
## Overall Statistics
##
##                Accuracy : 0.9937
##                  95% CI : (0.9913, 0.9956)
##     No Information Rate : 0.2856
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.992
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9952   0.9930   0.9854   0.9937   1.0000
```

```
## Specificity            0.9998   0.9983   0.9977   0.9970   0.9996
## Pos Pred Value         0.9994   0.9930   0.9893   0.9844   0.9982
## Neg Pred Value         0.9981   0.9983   0.9969   0.9988   1.0000
## Prevalence             0.2856   0.1935   0.1750   0.1623   0.1835
## Detection Rate         0.2843   0.1922   0.1725   0.1613   0.1835
## Detection Prevalence   0.2845   0.1935   0.1743   0.1638   0.1839
## Balanced Accuracy      0.9975   0.9956   0.9916   0.9953   0.9998
```

Ultimately, perform calculation of accuracy:

```
accuracy <- postResample(predictRf, testData$classe)
accuracy

##  Accuracy     Kappa
## 0.9937128 0.9920461

oose <- 1 - as.numeric(confusionMatrix(testData$classe, predictRf)$overall[1])
oose

## [1] 0.006287171
```

Result: the estimated accuracy of the model is 99.37% and the estimated out-of-sample error is 0.63%.

Perform prediction

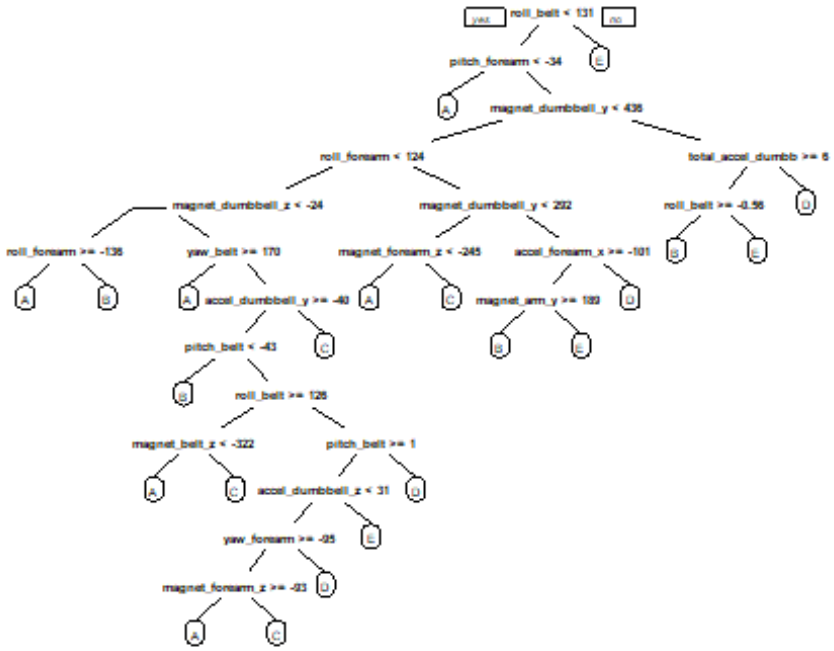
In the following step of the project, the developed model will be applied to the original testing data set downloaded from the data source:

```
result <- predict(modelRf, test_tidy[, -length(names(test_tidy))])
result

##  [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
```

Decision tree visualisation:

```
treeModel <- rpart(classe ~ ., data=trainData, method="class")
prp(treeModel)
```



Practical Machine Learning Course Project Quiz

To run and write into txt-files the prediction simulation for 20 cases the followwing chunk of R-code will be used:

```
quiz_result <- result
write_txt <- function(x){
  n = length(x)
  for(i in 1:n){
    filename = paste0("quiz/question_",i,".txt")
    write.table(x[i], file=filename, quote=FALSE,
                row.names=FALSE, col.names=FALSE)
  }
}
write_txt(quiz_result)
```