

# Practical Machine Learning - Week 4

*PGM*

18/11/2017

## 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://web.archive.org/web/20161224072740/http://groupware.les.inf.puc-rio.br/har>

(<http://web.archive.org/web/20161224072740/http://groupware.les.inf.puc-rio.br/har>) (see the section on the Weight Lifting Exercise Dataset).

### Data

The training data for this project are available here:

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

(<https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv>)

The test data are available here:

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

(<https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv>)

The data for this project come from this source:

<http://web.archive.org/web/20161224072740/http://groupware.les.inf.puc-rio.br/har>

(<http://web.archive.org/web/20161224072740/http://groupware.les.inf.puc-rio.br/har>). If you use the document you create for this class for any purpose please cite them as they have been very generous in allowing their data to be used for this kind of assignment.

## Cleaning data

```
## Download the CSV files with the data
fileURL <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv"
download.file(fileURL, destfile = "./pml-training.csv", method = "curl")
dateDownloaded <- date()

fileURL <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv"
download.file(fileURL, destfile = "./pml-testing.csv", method = "curl")
dateDownloaded <- date()

## Loading the data from the CSV files downloaded
pml_training <- read.csv("./pml-training.csv", sep = ",", header = TRUE)
head(pml_training)
```

```
##      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_pitch_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_pitch_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
##	2	NA	NA	NA
##	3	NA	NA	NA
##	4	NA	NA	NA
##	5	NA	NA	NA
##	6	NA	NA	NA
##	gyros_dumbbell_y gyros_dumbbell_z accel_dumbbell_x accel_dumbbell_y			
##	1	-0.02	0.00	-234
##	2	-0.02	0.00	-233
##	3	-0.02	0.00	-232
##	4	-0.02	-0.02	-232
##	5	-0.02	0.00	-233
##	6	-0.02	0.00	-234
##	accel_dumbbell_z magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z			
##	1	-271	-559	293
##	2	-269	-555	296
##	3	-270	-561	298
##	4	-269	-552	303
##	5	-270	-554	292

##	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
```

```
pml_testing <- read.csv("./pml-testing.csv", sep = ",", header = TRUE)
head(pml_testing)
```

```
##  X user_name raw_timestamp_part_1 raw_timestamp_part_2   cvtd_timestamp
## 1 1   pedro          1323095002          868349 05/12/2011 14:23
## 2 2   jeremy          1322673067          778725 30/11/2011 17:11
## 3 3   jeremy          1322673075          342967 30/11/2011 17:11
## 4 4   adelmo          1322832789          560311 02/12/2011 13:33
## 5 5   eurico          1322489635          814776 28/11/2011 14:13
## 6 6   jeremy          1322673149          510661 30/11/2011 17:12
##  new_window num_window roll_belt pitch_belt yaw_belt total_accel_belt
## 1         no          74    123.00    27.00   -4.75          20
## 2         no         431     1.02     4.87  -88.90           4
## 3         no         439     0.87     1.82  -88.50           5
## 4         no         194    125.00   -41.60  162.00          17
## 5         no         235     1.35     3.33  -88.60           3
## 6         no         504    -5.92     1.59  -87.70           4
##  kurtosis_roll_belt kurtosis_pitch_belt kurtosis_yaw_belt
## 1          NA          NA          NA
## 2          NA          NA          NA
## 3          NA          NA          NA
## 4          NA          NA          NA
## 5          NA          NA          NA
## 6          NA          NA          NA
##  skewness_roll_belt skewness_roll_belt.1 skewness_yaw_belt max_roll_belt
## 1          NA          NA          NA          NA
## 2          NA          NA          NA          NA
```

##	3	NA	NA	NA	NA
##	4	NA	NA	NA	NA
##	5	NA	NA	NA	NA
##	6	NA	NA	NA	NA
##		max_pitch_belt	max_yaw_belt	min_roll_belt	min_pitch_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
##		amplitude_roll_belt	amplitude_pitch_belt	amplitude_yaw_belt	
##	1	NA	NA	NA	
##	2	NA	NA	NA	
##	3	NA	NA	NA	
##	4	NA	NA	NA	
##	5	NA	NA	NA	
##	6	NA	NA	NA	
##		var_total_accel_belt	avg_roll_belt	stddev_roll_belt	var_roll_belt
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##	3	NA	NA	NA	NA
##	4	NA	NA	NA	NA
##	5	NA	NA	NA	NA
##	6	NA	NA	NA	NA
##		avg_pitch_belt	stddev_pitch_belt	var_pitch_belt	avg_yaw_belt
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##	3	NA	NA	NA	NA
##	4	NA	NA	NA	NA
##	5	NA	NA	NA	NA
##	6	NA	NA	NA	NA
##		stddev_yaw_belt	var_yaw_belt	gyros_belt_x	gyros_belt_y
##	1	NA	NA	-0.50	-0.02
##	2	NA	NA	-0.06	-0.02
##	3	NA	NA	0.05	0.02
##	4	NA	NA	0.11	0.11
##	5	NA	NA	0.03	0.02
##	6	NA	NA	0.10	0.05
##		gyros_belt_z	magnet_belt_x	magnet_belt_y	
##	1	-0.46	-13	581	
##	2	-0.07	43	636	
##	3	0.03	29	631	
##	4	-0.16	169	608	
##	5	0.00	33	566	
##	6	-0.13	31	638	
##		accel_belt_x	accel_belt_y	accel_belt_z	magnet_belt_z
##	1	-38	69	-179	-382
##	2	-13	11	39	-309
##	3	1	-1	49	-312
##	4	46	45	-156	-304
##	5	-8	4	27	-418
##	6	-11	-16	38	-291
##		roll_arm	pitch_arm	yaw_arm	total_accel_arm
##	1	40.7	-27.80	178	10
##	2	0.0	0.00	0	38
##	3	0.0	0.00	0	44
##	4	-109.0	55.00	-142	25
##	5	76.1	2.76	102	29
##	6	0.0	0.00	0	14
##		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	-1.65
##	2	NA	NA	NA	NA	-1.17
##	3	NA	NA	NA	NA	2.10
##	4	NA	NA	NA	NA	0.22
##	5	NA	NA	NA	NA	-1.96
##	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	0.48	-0.18	16	38	93 -326
##	2	0.85	-0.43	-290	215	-90 -325
##	3	-1.36	1.13	-341	245	-87 -264
##	4	-0.51	0.92	-238	-57	6 -173
##	5	0.79	-0.54	-197	200	-30 -170
##	6	0.05	-0.07	-26	130	-19 396
##	magnet_arm_y magnet_arm_z kurtosis_roll_arm kurtosis_pitch_arm					
##	1	385	481	NA	NA	
##	2	447	434	NA	NA	
##	3	474	413	NA	NA	
##	4	257	633	NA	NA	
##	5	275	617	NA	NA	
##	6	176	516	NA	NA	
##	kurtosis_yaw_arm skewness_roll_arm skewness_pitch_arm skewness_yaw_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
##	max_roll_arm max_pitch_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	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
##	roll_dumbbell pitch_dumbbell yaw_dumbbell kurtosis_roll_dumbbell					
##	1	-17.73748	24.96085	126.23596	NA	
##	2	54.47761	-53.69758	-75.51480	NA	
##	3	57.07031	-51.37303	-75.20287	NA	
##	4	43.10927	-30.04885	-103.32003	NA	
##	5	-101.38396	-53.43952	-14.19542	NA	

## 6	62.18750	-50.55595	-71.12063	NA
##	kurtosis_picth_dumbbell	kurtosis_yaw_dumbbell	skewness_roll_dumbbell	
## 1	NA	NA	NA	NA
## 2	NA	NA	NA	NA
## 3	NA	NA	NA	NA
## 4	NA	NA	NA	NA
## 5	NA	NA	NA	NA
## 6	NA	NA	NA	NA
##	skewness_pitch_dumbbell	skewness_yaw_dumbbell	max_roll_dumbbell	
## 1	NA	NA	NA	NA
## 2	NA	NA	NA	NA
## 3	NA	NA	NA	NA
## 4	NA	NA	NA	NA
## 5	NA	NA	NA	NA
## 6	NA	NA	NA	NA
##	max_picth_dumbbell	max_yaw_dumbbell	min_roll_dumbbell	min_pitch_dumbbell
## 1	NA	NA	NA	NA
## 2	NA	NA	NA	NA
## 3	NA	NA	NA	NA
## 4	NA	NA	NA	NA
## 5	NA	NA	NA	NA
## 6	NA	NA	NA	NA
##	min_yaw_dumbbell	amplitude_roll_dumbbell	amplitude_pitch_dumbbell	
## 1	NA	NA	NA	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
##	amplitude_yaw_dumbbell	total_accel_dumbbell	var_accel_dumbbell	
## 1	NA	9	NA	NA
## 2	NA	31	NA	NA
## 3	NA	29	NA	NA
## 4	NA	18	NA	NA
## 5	NA	4	NA	NA
## 6	NA	29	NA	NA
##	avg_roll_dumbbell	stddev_roll_dumbbell	var_roll_dumbbell	
## 1	NA	NA	NA	NA
## 2	NA	NA	NA	NA
## 3	NA	NA	NA	NA
## 4	NA	NA	NA	NA
## 5	NA	NA	NA	NA
## 6	NA	NA	NA	NA
##	avg_pitch_dumbbell	stddev_pitch_dumbbell	var_pitch_dumbbell	
## 1	NA	NA	NA	NA
## 2	NA	NA	NA	NA
## 3	NA	NA	NA	NA
## 4	NA	NA	NA	NA
## 5	NA	NA	NA	NA
## 6	NA	NA	NA	NA
##	avg_yaw_dumbbell	stddev_yaw_dumbbell	var_yaw_dumbbell	gyros_dumbbell_x
## 1	NA	NA	NA	0.64
## 2	NA	NA	NA	0.34
## 3	NA	NA	NA	0.39

##	4	NA	NA	NA	0.10
##	5	NA	NA	NA	0.29
##	6	NA	NA	NA	-0.59
##	gyros_dumbbell_y gyros_dumbbell_z accel_dumbbell_x accel_dumbbell_y				
##	1	0.06	-0.61	21	-15
##	2	0.05	-0.71	-153	155
##	3	0.14	-0.34	-141	155
##	4	-0.02	0.05	-51	72
##	5	-0.47	-0.46	-18	-30
##	6	0.80	1.10	-138	166
##	accel_dumbbell_z magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z				
##	1	81	523	-528	-56
##	2	-205	-502	388	-36
##	3	-196	-506	349	41
##	4	-148	-576	238	53
##	5	-5	-424	252	312
##	6	-186	-543	262	96
##	roll_forearm pitch_forearm yaw_forearm kurtosis_roll_forearm				
##	1	141	49.30	156.0	NA
##	2	109	-17.60	106.0	NA
##	3	131	-32.60	93.0	NA
##	4	0	0.00	0.0	NA
##	5	-176	-2.16	-47.9	NA
##	6	150	1.46	89.7	NA
##	kurtosis_pitch_forearm kurtosis_yaw_forearm skewness_roll_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
##	skewness_pitch_forearm skewness_yaw_forearm max_roll_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
##	max_pitch_forearm max_yaw_forearm min_roll_forearm min_pitch_forearm				
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##	3	NA	NA	NA	NA
##	4	NA	NA	NA	NA
##	5	NA	NA	NA	NA
##	6	NA	NA	NA	NA
##	min_yaw_forearm amplitude_roll_forearm amplitude_pitch_forearm				
##	1	NA	NA	NA	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
##	amplitude_yaw_forearm total_accel_forearm var_accel_forearm				
##	1	NA	33	NA	

```
## 2          NA          39          NA
## 3          NA          34          NA
## 4          NA          43          NA
## 5          NA          24          NA
## 6          NA          43          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          0.74          -3.34
## 2          NA          NA          1.12          -2.78
## 3          NA          NA          0.18          -0.79
## 4          NA          NA          1.38          0.69
## 5          NA          NA          -0.75          3.10
## 6          NA          NA          -0.88          4.26
##  gyros_forearm_z accel_forearm_x accel_forearm_y accel_forearm_z
## 1          -0.59          -110          267          -149
## 2          -0.18          212          297          -118
## 3           0.28          154          271          -129
## 4           1.80          -92          406          -39
## 5           0.80          131          -93          172
## 6           1.35          230          322          -144
##  magnet_forearm_x magnet_forearm_y magnet_forearm_z problem_id
## 1          -714          419          617          1
## 2          -237          791          873          2
## 3          -51          698          783          3
## 4          -233          783          521          4
## 5           375          -787          91          5
## 6          -300          800          884          6
```

Search for the strings regarding the accelerometers on 4 locations (belt, forearm, arm, dumbbell)

```
choices <- c("accel_belt", "accel_forearm", "accel_arm", "accel_dumbbell")
columns <- grepl(paste(choices, collapse = "|"), colnames(pml_training))
```

Manually making that the first and last elements true to select those extra columns:

```
# Extracting the data in those columns only
specificTraining <- pml_training[, columns]

## Removing choices and columns
rm(choices, columns)
```

Looking briefly at the resulting data:

```
summary(specificTraining)
```

```

## total_accel_belt var_total_accel_belt accel_belt_x accel_belt_y
## Min. : 0.00 Min. : 0.000 Min. : -120.000 Min. : -69.00
## 1st Qu.: 3.00 1st Qu.: 0.100 1st Qu.: -21.000 1st Qu.: 3.00
## Median :17.00 Median : 0.200 Median : -15.000 Median : 35.00
## Mean :11.31 Mean : 0.926 Mean : -5.595 Mean : 30.15
## 3rd Qu.:18.00 3rd Qu.: 0.300 3rd Qu.: -5.000 3rd Qu.: 61.00
## Max. :29.00 Max. :16.500 Max. : 85.000 Max. :164.00
## NA's :19216
## accel_belt_z total_accel_arm var_accel_arm accel_arm_x
## Min. : -275.00 Min. : 1.00 Min. : 0.00 Min. : -404.00
## 1st Qu.: -162.00 1st Qu.:17.00 1st Qu.: 9.03 1st Qu.: -242.00
## Median : -152.00 Median :27.00 Median : 40.61 Median : -44.00
## Mean : -72.59 Mean :25.51 Mean : 53.23 Mean : -60.24
## 3rd Qu.: 27.00 3rd Qu.:33.00 3rd Qu.: 75.62 3rd Qu.: 84.00
## Max. : 105.00 Max. :66.00 Max. :331.70 Max. : 437.00
## NA's :19216
## accel_arm_y accel_arm_z total_accel_dumbbell
## Min. : -318.0 Min. : -636.00 Min. : 0.00
## 1st Qu.: -54.0 1st Qu.: -143.00 1st Qu.: 4.00
## Median : 14.0 Median : -47.00 Median :10.00
## Mean : 32.6 Mean : -71.25 Mean :13.72
## 3rd Qu.:139.0 3rd Qu.: 23.00 3rd Qu.:19.00
## Max. : 308.0 Max. : 292.00 Max. :58.00
##
## var_accel_dumbbell accel_dumbbell_x accel_dumbbell_y accel_dumbbell_z
## Min. : 0.000 Min. : -419.00 Min. : -189.00 Min. : -334.00
## 1st Qu.: 0.378 1st Qu.: -50.00 1st Qu.: -8.00 1st Qu.: -142.00
## Median : 1.000 Median : -8.00 Median : 41.50 Median : -1.00
## Mean : 4.388 Mean : -28.62 Mean : 52.63 Mean : -38.32
## 3rd Qu.: 3.434 3rd Qu.: 11.00 3rd Qu.:111.00 3rd Qu.: 38.00
## Max. :230.428 Max. : 235.00 Max. : 315.00 Max. : 318.00
## NA's :19216
## total_accel_forearm var_accel_forearm accel_forearm_x accel_forearm_y
## Min. : 0.00 Min. : 0.000 Min. : -498.00 Min. : -632.0
## 1st Qu.: 29.00 1st Qu.: 6.759 1st Qu.: -178.00 1st Qu.: 57.0
## Median : 36.00 Median : 21.165 Median : -57.00 Median : 201.0
## Mean : 34.72 Mean : 33.502 Mean : -61.65 Mean : 163.7
## 3rd Qu.: 41.00 3rd Qu.:51.240 3rd Qu.: 76.00 3rd Qu.: 312.0
## Max. :108.00 Max. :172.606 Max. : 477.00 Max. : 923.0
## NA's :19216
## accel_forearm_z
## Min. : -446.00
## 1st Qu.: -182.00
## Median : -39.00
## Mean : -55.29
## 3rd Qu.: 26.00
## Max. : 291.00
##

```

The amount of NA's is so high on the variables 'var\_accel' & 'var\_total\_accel' from the 4 locations that it those columns will be removed from the pool of independent variables.

```

choices <- c("var_accel", "var_total_accel")
columns <- grepl(paste(choices, collapse = "|"), colnames(specificTraining))

# Extracting the data in those columns only
specificTraining <- specificTraining[, !columns]

## Removing choices and columns
rm(choices, columns)

```

Adding the “classe” variable to the specificTraining dataset:

```

# Adding classe
specificTraining <- cbind(specificTraining, classe = pml_training$classe)

```

## Data partition to evaluate out-of-sample errors

As the model cannot be tested with `pml_testing` without knowing how it performs, it is necessary to split the training data in 2 groups:

```

library(caret)

```

```

## Loading required package: lattice

```

```

## Loading required package: ggplot2

```

```

set.seed(28322)

# Generating a possible partition (80% / 20%)
inTrain <- createDataPartition(y = specificTraining$classe, p = 0.80, list = FALSE)

# Assigning most of the data to the subtraining group and the rest to the subtesting group
subtraining <- specificTraining[inTrain, ]
subtesting <- specificTraining[-inTrain, ]

```

## Model development

Building a random forest model:

```

library(randomForest)

```

```

## randomForest 4.6-12

```

```

## Type rfNews() to see new features/changes/bug fixes.

```

```
##  
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':  
##  
##     margin
```

```
set.seed(28322)  
  
# Building a random forest model  
modFit <- randomForest(classe ~ ., data = subtraining)  
modFit
```

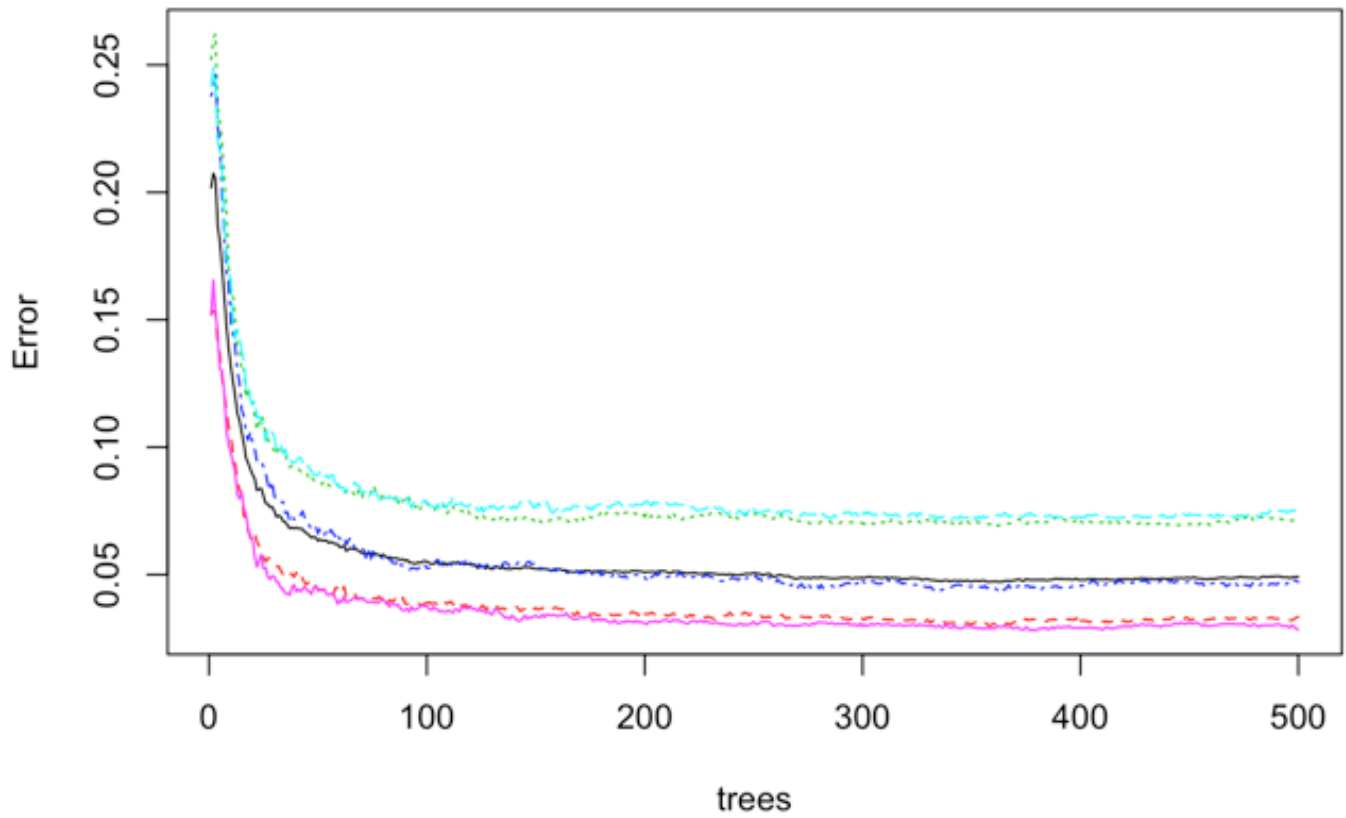
```
##  
## Call:  
## randomForest(formula = classe ~ ., data = subtraining)  
##           Type of random forest: classification  
##           Number of trees: 500  
## No. of variables tried at each split: 4  
##  
##           OOB estimate of  error rate: 4.92%  
## Confusion matrix:  
##           A      B      C      D      E class.error  
## A 4315    28    60    58     3 0.03337814  
## B  101 2819    74    20    24 0.07208690  
## C   41   62 2609    21     5 0.04711468  
## D   65   13  101 2379    15 0.07539837  
## E    5   32   21   24 2804 0.02841303
```

Showing the results visually for the iterations made during the model creation:

```
plot(modFit)
```



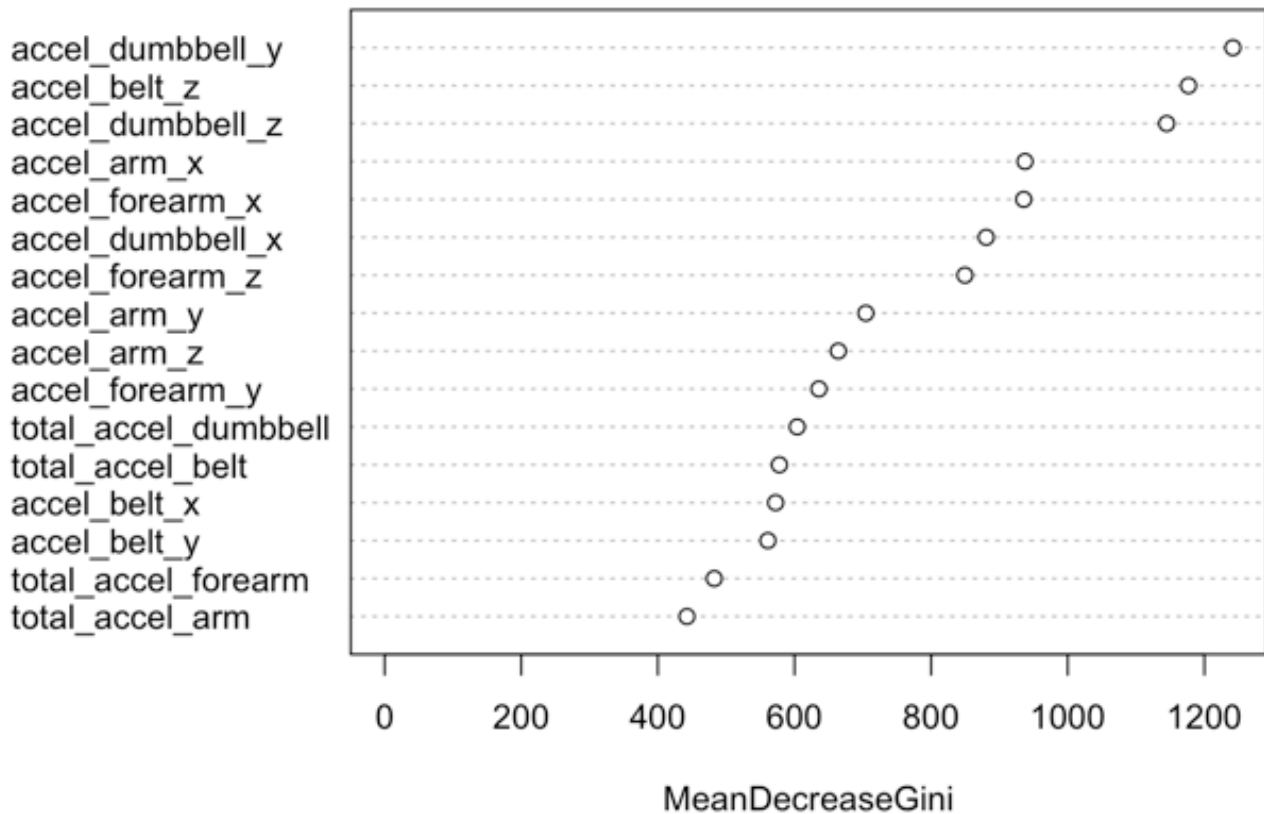
## modFit



Dotchart of variable importance as measured by the random forest model just created:

```
varImpPlot(modFit)
```

## modFit



## Application of model to subtesting dataset

Predicting the classe values for subtesting:

```
# A variable classe is added to the testing dataset with the predictions
prediction <- predict(modFit, newdata = subtesting)

# Creating the confusion matrix for the results achieved
confusion <- confusionMatrix(prediction, subtesting$classe)
confusion
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   A    B    C    D    E
##           A 1071   32   10   10   1
##           B   11  701   19    2   5
##           C   15   22  640   27   5
##           D   18    1   11  603   5
##           E    1    3    4    1  705
##
## Overall Statistics
##
##           Accuracy : 0.9483
##           95% CI : (0.9409, 0.955)
##           No Information Rate : 0.2845
##           P-Value [Acc > NIR] : < 2.2e-16
##
##           Kappa : 0.9345
##           McNemar's Test P-Value : 0.007316
##
## Statistics by Class:
##
##           Class: A Class: B Class: C Class: D Class: E
## Sensitivity      0.9597   0.9236   0.9357   0.9378   0.9778
## Specificity      0.9811   0.9883   0.9787   0.9893   0.9972
## Pos Pred Value   0.9528   0.9499   0.9027   0.9451   0.9874
## Neg Pred Value   0.9839   0.9818   0.9863   0.9878   0.9950
## Prevalence       0.2845   0.1935   0.1744   0.1639   0.1838
## Detection Rate   0.2730   0.1787   0.1631   0.1537   0.1797
## Detection Prevalence 0.2865   0.1881   0.1807   0.1626   0.1820
## Balanced Accuracy 0.9704   0.9559   0.9572   0.9636   0.9875
```

Accuracy of 94.83%, so that seems interesting.

From this result, it is possible to have a reference of the performance from the calculated model:

```
out_of_sample_error <- 1 - confusion$overall[1]
names(out_of_sample_error) <- "Error"
out_of_sample_error
```

```
##           Error
## 0.05174611
```

## Application of model to global testing dataset

Predicting the classe values for pml\_testing:

```
# A variable classe is added to the testing dataset with the predictions
pml_testing$classe <- predict(modFit, newdata = pml_testing)

# The predictions to be provider to the automatic checker
pml_testing$classe
```

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