Practical Machine Learning Project

Patrick Sciamma

November 4, 2017

Summary

The aim of the study is to predict if a physical activity is performed correctly. We 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. We identify primary and statistical measures in our dataset and restrict our study to primary measures. Using 5-fold cross-validation we test three models: k-nearest neighbours, naive Bayes and random Forest. Based on an estimated accuracy of 99.29% we select the random forest model classifier, confirm its accuracy on our test set at above 99.29% and apply it to predict our final results.

Source for the dataset

Velloso, E.; Bulling, A.; Gellersen, H.; Ugulino, W.; Fuks, H. Qualitative Activity Recognition of Weight Lifting Exercises. Proceedings of 4th International Conference in Cooperation with SIGCHI (Augmented Human '13) . Stuttgart, Germany: ACM SIGCHI, 2013.

Read more: http://groupware.les.inf.puc-rio.br/har#wle_paper_section#ixzz4y8Ou0ILz

Data analysis

load data and libraries

```
setwd("C:/Users/Patrick/Documents/R/MachineLearning/Project")
getwd()
library(caret)
library(dplyr)
library(magrittr)
library(Hmisc)
train <- read.csv("./data/pml-training.csv")
test <- read.csv("./data/pml-testing.csv")
```

field selection

The training test contains 19622 observations of 160 variables. The first fields are identifier, user name (6 users), time stamps, window identifier and indicator for new window

```
dim(train)
```

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

```
dim(test)
```

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

```
str(train[,1:7])
```

```
## 'data.frame':    19622 obs. of  7 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 1323
##  $ raw_timestamp_part_2: int  788290 808298 820366 120339 196328 304277 368296 440390 484323 484434
##  $ 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 ...
```

From the data description at "http://groupware.les.inf.puc-rio.br/har" we know the dataset contains measures for 4 sensors: arm, forearm, belt and dumbbell

```
names(train)[grep("belt", names(train))]
```

```
##  [1] "roll_belt"            "pitch_belt"           "yaw_belt"
##  [4] "total_accel_belt"     "kurtosis_roll_belt"   "kurtosis_picth_belt"
##  [7] "kurtosis_yaw_belt"    "skewness_roll_belt"   "skewness_roll_belt.1"
## [10] "skewness_yaw_belt"    "max_roll_belt"        "max_picth_belt"
## [13] "max_yaw_belt"         "min_roll_belt"        "min_pitch_belt"
## [16] "min_yaw_belt"         "amplitude_roll_belt"  "amplitude_pitch_belt"
## [19] "amplitude_yaw_belt"   "var_total_accel_belt" "avg_roll_belt"
## [22] "stddev_roll_belt"     "var_roll_belt"        "avg_pitch_belt"
## [25] "stddev_pitch_belt"    "var_pitch_belt"       "avg_yaw_belt"
## [28] "stddev_yaw_belt"      "var_yaw_belt"         "gyros_belt_x"
## [31] "gyros_belt_y"         "gyros_belt_z"         "accel_belt_x"
## [34] "accel_belt_y"         "accel_belt_z"         "magnet_belt_x"
## [37] "magnet_belt_y"        "magnet_belt_z"
```

```
names(train)[grep("_arm", names(train))]
```

```
##  [1] "roll_arm"            "pitch_arm"           "yaw_arm"
##  [4] "total_accel_arm"     "var_accel_arm"       "avg_roll_arm"
##  [7] "stddev_roll_arm"     "var_roll_arm"        "avg_pitch_arm"
## [10] "stddev_pitch_arm"    "var_pitch_arm"       "avg_yaw_arm"
## [13] "stddev_yaw_arm"      "var_yaw_arm"         "gyros_arm_x"
## [16] "gyros_arm_y"         "gyros_arm_z"         "accel_arm_x"
## [19] "accel_arm_y"         "accel_arm_z"         "magnet_arm_x"
## [22] "magnet_arm_y"        "magnet_arm_z"        "kurtosis_roll_arm"
## [25] "kurtosis_picth_arm"  "kurtosis_yaw_arm"    "skewness_roll_arm"
## [28] "skewness_pitch_arm"  "skewness_yaw_arm"    "max_roll_arm"
## [31] "max_picth_arm"       "max_yaw_arm"         "min_roll_arm"
## [34] "min_pitch_arm"       "min_yaw_arm"         "amplitude_roll_arm"
## [37] "amplitude_pitch_arm" "amplitude_yaw_arm"
```

```
names(train)[grep("forearm", names(train))]
```

```
##  [1] "roll_forearm"            "pitch_forearm"
##  [3] "yaw_forearm"             "kurtosis_roll_forearm"
##  [5] "kurtosis_picth_forearm"  "kurtosis_yaw_forearm"
##  [7] "skewness_roll_forearm"   "skewness_pitch_forearm"
##  [9] "skewness_yaw_forearm"    "max_roll_forearm"
## [11] "max_picth_forearm"       "max_yaw_forearm"
## [13] "min_roll_forearm"        "min_pitch_forearm"
## [15] "min_yaw_forearm"         "amplitude_roll_forearm"
## [17] "amplitude_pitch_forearm" "amplitude_yaw_forearm"
## [19] "total_accel_forearm"     "var_accel_forearm"
## [21] "avg_roll_forearm"        "stddev_roll_forearm"
```

```
## [23] "var_roll_forearm"        "avg_pitch_forearm"
## [25] "stddev_pitch_forearm"    "var_pitch_forearm"
## [27] "avg_yaw_forearm"         "stddev_yaw_forearm"
## [29] "var_yaw_forearm"         "gyros_forearm_x"
## [31] "gyros_forearm_y"         "gyros_forearm_z"
## [33] "accel_forearm_x"         "accel_forearm_y"
## [35] "accel_forearm_z"         "magnet_forearm_x"
## [37] "magnet_forearm_y"        "magnet_forearm_z"
```

```
names(train)[grep("dumbbell", names(train))]
```

```
##  [1] "roll_dumbbell"             "pitch_dumbbell"
##  [3] "yaw_dumbbell"              "kurtosis_roll_dumbbell"
##  [5] "kurtosis_picth_dumbbell"   "kurtosis_yaw_dumbbell"
##  [7] "skewness_roll_dumbbell"    "skewness_pitch_dumbbell"
##  [9] "skewness_yaw_dumbbell"     "max_roll_dumbbell"
## [11] "max_picth_dumbbell"        "max_yaw_dumbbell"
## [13] "min_roll_dumbbell"         "min_pitch_dumbbell"
## [15] "min_yaw_dumbbell"          "amplitude_roll_dumbbell"
## [17] "amplitude_pitch_dumbbell" "amplitude_yaw_dumbbell"
## [19] "total_accel_dumbbell"      "var_accel_dumbbell"
## [21] "avg_roll_dumbbell"         "stddev_roll_dumbbell"
## [23] "var_roll_dumbbell"         "avg_pitch_dumbbell"
## [25] "stddev_pitch_dumbbell"     "var_pitch_dumbbell"
## [27] "avg_yaw_dumbbell"          "stddev_yaw_dumbbell"
## [29] "var_yaw_dumbbell"          "gyros_dumbbell_x"
## [31] "gyros_dumbbell_y"          "gyros_dumbbell_z"
## [33] "accel_dumbbell_x"          "accel_dumbbell_y"
## [35] "accel_dumbbell_z"          "magnet_dumbbell_x"
## [37] "magnet_dumbbell_y"         "magnet_dumbbell_z"
```

There are 13 primary measures for the belt: roll, pitch, yaw, total acceleration, 'gyros', 'accel' and 'magnet' measures on 3 axes

```
beltm <- names(train)[grep("belt", names(train))]
i <- grep("(^[a-z]+_belt)|(^total_[a-z]+_belt)", beltm)
beltm[i]
```

```
##  [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"
```

```
anyNA(train[,beltm[i]])
```

```
## [1] FALSE
```

```
apply(train[,beltm[i]],2,function(x) sum(x == ''))
```

```
##        roll_belt       pitch_belt         yaw_belt total_accel_belt
##                0                0                0                0
##     gyros_belt_x     gyros_belt_y     gyros_belt_z     accel_belt_x
##                0                0                0                0
##     accel_belt_y     accel_belt_z    magnet_belt_x    magnet_belt_y
##                0                0                0                0
##    magnet_belt_z
```

```
##                0
```

There are 25 statistics on those measures, with mostly NAs or errors when the new window indicator is False. The reasonable explanation is that for every new window, statistics are given for all measures in that window.

```
beltm[-i]
```

```
##  [1] "kurtosis_roll_belt"   "kurtosis_picth_belt"  "kurtosis_yaw_belt"
##  [4] "skewness_roll_belt"   "skewness_roll_belt.1" "skewness_yaw_belt"
##  [7] "max_roll_belt"        "max_picth_belt"       "max_yaw_belt"
## [10] "min_roll_belt"        "min_pitch_belt"       "min_yaw_belt"
## [13] "amplitude_roll_belt"  "amplitude_pitch_belt" "amplitude_yaw_belt"
## [16] "var_total_accel_belt" "avg_roll_belt"        "stddev_roll_belt"
## [19] "var_roll_belt"        "avg_pitch_belt"       "stddev_pitch_belt"
## [22] "var_pitch_belt"       "avg_yaw_belt"         "stddev_yaw_belt"
## [25] "var_yaw_belt"
```

```
t <- train[train$new_window=='no',beltm[-i]]
str(t)
```

```
## 'data.frame':    19216 obs. of  25 variables:
##  $ kurtosis_roll_belt  : Factor w/ 397 levels "","-0.016850",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ kurtosis_picth_belt : Factor w/ 317 levels "","-0.021887",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ kurtosis_yaw_belt   : Factor w/ 2 levels "","#DIV/0!": 1 1 1 1 1 1 1 1 1 1 ...
##  $ skewness_roll_belt  : Factor w/ 395 levels "","-0.003095",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ skewness_roll_belt.1: Factor w/ 338 levels "","-0.005928",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ skewness_yaw_belt   : Factor w/ 2 levels "","#DIV/0!": 1 1 1 1 1 1 1 1 1 1 ...
##  $ max_roll_belt       : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_picth_belt      : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ max_yaw_belt        : Factor w/ 68 levels "","-0.1","-0.2",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ 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        : Factor w/ 68 levels "","-0.1","-0.2",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ 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  : Factor w/ 4 levels "","#DIV/0!","0.00",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ 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 ...
```

For example for window 2, avg_roll_belt contains average of all measures in that window.

```
mean(train[train$num_window==2,'roll_belt'])
```

```
## [1] -27.43333
```

```
train[train$num_window==2 & train$new_window =='yes',]$avg_roll_belt
```

```
## [1] -27.4
```

We will leave the statistics out as they are mostly redundant with the raw measure. We will revisit this choice if we find we don't have enough data after all

```
train2 <- train %>%
  select(-starts_with("kurtosis"),
         -starts_with("skewness"),
         -starts_with("max"),
         -starts_with("min"),
         -starts_with("amplitude"),
         -starts_with("avg"),
         -starts_with("var"),
         -starts_with("stddev"),
         -contains("timestamp"),
         -X,-new_window,-num_window
         )

test2 <- test %>%
  select(-starts_with("kurtosis"),
         -starts_with("skewness"),
         -starts_with("max"),
         -starts_with("min"),
         -starts_with("amplitude"),
         -starts_with("avg"),
         -starts_with("var"),
         -starts_with("stddev"),
         -contains("timestamp"),
         -X,-new_window,-num_window,-problem_id
  )
```

The variable names are now consistent between train and test set, there are no NAs in our filtered dataset and a quick summary shows nothing untoward.

```
data.frame(names(train2[,1:53]), names(test2))
```

```
##    names.train2...1.53..         names.test2.
## 1              user_name            user_name
## 2              roll_belt            roll_belt
## 3             pitch_belt           pitch_belt
## 4               yaw_belt             yaw_belt
## 5       total_accel_belt     total_accel_belt
## 6           gyros_belt_x         gyros_belt_x
## 7           gyros_belt_y         gyros_belt_y
## 8           gyros_belt_z         gyros_belt_z
## 9           accel_belt_x         accel_belt_x
## 10          accel_belt_y         accel_belt_y
## 11          accel_belt_z         accel_belt_z
## 12        magnet_belt_x        magnet_belt_x
## 13        magnet_belt_y        magnet_belt_y
## 14        magnet_belt_z        magnet_belt_z
## 15             roll_arm             roll_arm
## 16            pitch_arm            pitch_arm
## 17              yaw_arm              yaw_arm
## 18       total_accel_arm      total_accel_arm
## 19          gyros_arm_x          gyros_arm_x
## 20          gyros_arm_y          gyros_arm_y
## 21          gyros_arm_z          gyros_arm_z
## 22          accel_arm_x          accel_arm_x
## 23          accel_arm_y          accel_arm_y
```

```
## 24           accel_arm_z          accel_arm_z
## 25          magnet_arm_x         magnet_arm_x
## 26          magnet_arm_y         magnet_arm_y
## 27          magnet_arm_z         magnet_arm_z
## 28         roll_dumbbell        roll_dumbbell
## 29        pitch_dumbbell       pitch_dumbbell
## 30          yaw_dumbbell         yaw_dumbbell
## 31  total_accel_dumbbell total_accel_dumbbell
## 32      gyros_dumbbell_x     gyros_dumbbell_x
## 33      gyros_dumbbell_y     gyros_dumbbell_y
## 34      gyros_dumbbell_z     gyros_dumbbell_z
## 35      accel_dumbbell_x     accel_dumbbell_x
## 36      accel_dumbbell_y     accel_dumbbell_y
## 37      accel_dumbbell_z     accel_dumbbell_z
## 38     magnet_dumbbell_x    magnet_dumbbell_x
## 39     magnet_dumbbell_y    magnet_dumbbell_y
## 40     magnet_dumbbell_z    magnet_dumbbell_z
## 41          roll_forearm         roll_forearm
## 42         pitch_forearm        pitch_forearm
## 43           yaw_forearm          yaw_forearm
## 44   total_accel_forearm  total_accel_forearm
## 45       gyros_forearm_x      gyros_forearm_x
## 46       gyros_forearm_y      gyros_forearm_y
## 47       gyros_forearm_z      gyros_forearm_z
## 48       accel_forearm_x      accel_forearm_x
## 49       accel_forearm_y      accel_forearm_y
## 50       accel_forearm_z      accel_forearm_z
## 51      magnet_forearm_x     magnet_forearm_x
## 52      magnet_forearm_y     magnet_forearm_y
## 53      magnet_forearm_z     magnet_forearm_z
```

```
anyNA(train2)
```

```
## [1] FALSE
```

```
anyNA(test2)
```

```
## [1] FALSE
```

```
summary(train2)
```

```
##     user_name      roll_belt        pitch_belt          yaw_belt
##  adelmo  :3892   Min.   :-28.90   Min.   :-55.8000   Min.   :-180.00
##  carlitos:3112   1st Qu.:  1.10   1st Qu.:  1.7600   1st Qu.: -88.30
##  charles :3536   Median :113.00   Median :  5.2800   Median : -13.00
##  eurico  :3070   Mean   : 64.41   Mean   :  0.3053   Mean   : -11.21
##  jeremy  :3402   3rd Qu.:123.00   3rd Qu.: 14.9000   3rd Qu.:  12.90
##  pedro   :2610   Max.   :162.00   Max.   : 60.3000   Max.   : 179.00
##  total_accel_belt  gyros_belt_x        gyros_belt_y       gyros_belt_z
##  Min.   : 0.00    Min.   :-1.040000   Min.   :-0.64000   Min.   :-1.4600
##  1st Qu.: 3.00    1st Qu.:-0.030000   1st Qu.: 0.00000   1st Qu.:-0.2000
##  Median :17.00    Median : 0.030000   Median : 0.02000   Median :-0.1000
##  Mean   :11.31    Mean   :-0.005592   Mean   : 0.03959   Mean   :-0.1305
##  3rd Qu.:18.00    3rd Qu.: 0.110000   3rd Qu.: 0.11000   3rd Qu.:-0.0200
##  Max.   :29.00    Max.   : 2.220000   Max.   : 0.64000   Max.   : 1.6200
##   accel_belt_x       accel_belt_y     accel_belt_z     magnet_belt_x
```

```
##  Min.   :-120.000   Min.   :-69.00   Min.   :-275.00   Min.   :-52.0
##  1st Qu.: -21.000   1st Qu.:  3.00   1st Qu.:-162.00   1st Qu.:  9.0
##  Median : -15.000   Median : 35.00   Median :-152.00   Median : 35.0
##  Mean   :  -5.595   Mean   : 30.15   Mean   : -72.59   Mean   : 55.6
##  3rd Qu.:  -5.000   3rd Qu.: 61.00   3rd Qu.:  27.00   3rd Qu.: 59.0
##  Max.   :  85.000   Max.   :164.00   Max.   : 105.00   Max.   :485.0
##  magnet_belt_y   magnet_belt_z       roll_arm         pitch_arm
##  Min.   :354.0   Min.   :-623.0   Min.   :-180.00   Min.   :-88.800
##  1st Qu.:581.0   1st Qu.:-375.0   1st Qu.: -31.77   1st Qu.:-25.900
##  Median :601.0   Median :-320.0   Median :   0.00   Median :  0.000
##  Mean   :593.7   Mean   :-345.5   Mean   :  17.83   Mean   : -4.612
##  3rd Qu.:610.0   3rd Qu.:-306.0   3rd Qu.:  77.30   3rd Qu.: 11.200
##  Max.   :673.0   Max.   : 293.0   Max.   : 180.00   Max.   : 88.500
##     yaw_arm          total_accel_arm  gyros_arm_x        gyros_arm_y
##  Min.   :-180.0000   Min.   : 1.00   Min.   :-6.37000   Min.   :-3.4400
##  1st Qu.: -43.1000   1st Qu.:17.00   1st Qu.:-1.33000   1st Qu.:-0.8000
##  Median :   0.0000   Median :27.00   Median : 0.08000   Median :-0.2400
##  Mean   :  -0.6188   Mean   :25.51   Mean   : 0.04277   Mean   :-0.2571
##  3rd Qu.:  45.8750   3rd Qu.:33.00   3rd Qu.: 1.57000   3rd Qu.: 0.1400
##  Max.   : 180.0000   Max.   :66.00   Max.   : 4.87000   Max.   : 2.8400
##   gyros_arm_z       accel_arm_x       accel_arm_y      accel_arm_z
##  Min.   :-2.3300   Min.   :-404.00   Min.   :-318.0   Min.   :-636.00
##  1st Qu.:-0.0700   1st Qu.:-242.00   1st Qu.: -54.0   1st Qu.:-143.00
##  Median : 0.2300   Median : -44.00   Median :  14.0   Median : -47.00
##  Mean   : 0.2695   Mean   : -60.24   Mean   :  32.6   Mean   : -71.25
##  3rd Qu.: 0.7200   3rd Qu.:  84.00   3rd Qu.: 139.0   3rd Qu.:  23.00
##  Max.   : 3.0200   Max.   : 437.00   Max.   : 308.0   Max.   : 292.00
##   magnet_arm_x     magnet_arm_y     magnet_arm_z    roll_dumbbell
##  Min.   :-584.0   Min.   :-392.0   Min.   :-597.0   Min.   :-153.71
##  1st Qu.:-300.0   1st Qu.:  -9.0   1st Qu.: 131.2   1st Qu.: -18.49
##  Median : 289.0   Median : 202.0   Median : 444.0   Median :  48.17
##  Mean   : 191.7   Mean   : 156.6   Mean   : 306.5   Mean   :  23.84
##  3rd Qu.: 637.0   3rd Qu.: 323.0   3rd Qu.: 545.0   3rd Qu.:  67.61
##  Max.   : 782.0   Max.   : 583.0   Max.   : 694.0   Max.   : 153.55
##  pitch_dumbbell     yaw_dumbbell      total_accel_dumbbell
##  Min.   :-149.59   Min.   :-150.871   Min.   : 0.00
##  1st Qu.: -40.89   1st Qu.: -77.644   1st Qu.: 4.00
##  Median : -20.96   Median :  -3.324   Median :10.00
##  Mean   : -10.78   Mean   :   1.674   Mean   :13.72
##  3rd Qu.:  17.50   3rd Qu.:  79.643   3rd Qu.:19.00
##  Max.   : 149.40   Max.   : 154.952   Max.   :58.00
##  gyros_dumbbell_x    gyros_dumbbell_y   gyros_dumbbell_z
##  Min.   :-204.0000   Min.   :-2.10000   Min.   : -2.380
##  1st Qu.:  -0.0300   1st Qu.:-0.14000   1st Qu.: -0.310
##  Median :   0.1300   Median : 0.03000   Median : -0.130
##  Mean   :   0.1611   Mean   : 0.04606   Mean   : -0.129
##  3rd Qu.:   0.3500   3rd Qu.: 0.21000   3rd Qu.:  0.030
##  Max.   :   2.2200   Max.   :52.00000   Max.   :317.000
##  accel_dumbbell_x  accel_dumbbell_y  accel_dumbbell_z  magnet_dumbbell_x
##  Min.   :-419.00   Min.   :-189.00   Min.   :-334.00   Min.   :-643.0
##  1st Qu.: -50.00   1st Qu.:  -8.00   1st Qu.:-142.00   1st Qu.:-535.0
##  Median :  -8.00   Median :  41.50   Median :  -1.00   Median :-479.0
##  Mean   : -28.62   Mean   :  52.63   Mean   : -38.32   Mean   :-328.5
##  3rd Qu.:  11.00   3rd Qu.: 111.00   3rd Qu.:  38.00   3rd Qu.:-304.0
```

```
##  Max.   : 235.00   Max.   : 315.00   Max.   : 318.00   Max.   : 592.0
##  magnet_dumbbell_y magnet_dumbbell_z  roll_forearm       pitch_forearm
##  Min.   :-3600     Min.   :-262.00   Min.   :-180.0000   Min.   :-72.50
##  1st Qu.:  231     1st Qu.: -45.00   1st Qu.:  -0.7375   1st Qu.:  0.00
##  Median :  311     Median :  13.00   Median :  21.7000   Median :  9.24
##  Mean   :  221     Mean   :  46.05   Mean   :  33.8265   Mean   : 10.71
##  3rd Qu.:  390     3rd Qu.:  95.00   3rd Qu.: 140.0000   3rd Qu.: 28.40
##  Max.   :  633     Max.   : 452.00   Max.   : 180.0000   Max.   : 89.80
##   yaw_forearm      total_accel_forearm gyros_forearm_x
##  Min.   :-180.00   Min.   :  0.00      Min.   :-22.000
##  1st Qu.: -68.60   1st Qu.: 29.00      1st Qu.: -0.220
##  Median :   0.00   Median : 36.00      Median :  0.050
##  Mean   :  19.21   Mean   : 34.72      Mean   :  0.158
##  3rd Qu.: 110.00   3rd Qu.: 41.00      3rd Qu.:  0.560
##  Max.   : 180.00   Max.   :108.00      Max.   :  3.970
##  gyros_forearm_y     gyros_forearm_z    accel_forearm_x   accel_forearm_y
##  Min.   : -7.02000   Min.   : -8.0900   Min.   :-498.00   Min.   :-632.0
##  1st Qu.: -1.46000   1st Qu.: -0.1800   1st Qu.:-178.00   1st Qu.:  57.0
##  Median :  0.03000   Median :  0.0800   Median : -57.00   Median : 201.0
##  Mean   :  0.07517   Mean   :  0.1512   Mean   : -61.65   Mean   : 163.7
##  3rd Qu.:  1.62000   3rd Qu.:  0.4900   3rd Qu.:  76.00   3rd Qu.: 312.0
##  Max.   :311.00000   Max.   :231.0000   Max.   : 477.00   Max.   : 923.0
##  accel_forearm_z   magnet_forearm_x  magnet_forearm_y magnet_forearm_z
##  Min.   :-446.00   Min.   :-1280.0   Min.   :-896.0   Min.   :-973.0
##  1st Qu.:-182.00   1st Qu.: -616.0   1st Qu.:   2.0   1st Qu.: 191.0
##  Median : -39.00   Median : -378.0   Median : 591.0   Median : 511.0
##  Mean   : -55.29   Mean   : -312.6   Mean   : 380.1   Mean   : 393.6
##  3rd Qu.:  26.00   3rd Qu.:  -73.0   3rd Qu.: 737.0   3rd Qu.: 653.0
##  Max.   : 291.00   Max.   :  672.0   Max.   :1480.0   Max.   :1090.0
##  classe
##  A:5580
##  B:3797
##  C:3422
##  D:3216
##  E:3607
##
```

```
summary(test2)
```

```
##     user_name   roll_belt          pitch_belt         yaw_belt
##  adelmo  :1   Min.   : -5.9200   Min.   :-41.600   Min.   :-93.70
##  carlitos:3   1st Qu.:  0.9075   1st Qu.:  3.013   1st Qu.:-88.62
##  charles :1   Median :  1.1100   Median :  4.655   Median :-87.85
##  eurico  :4   Mean   : 31.3055   Mean   :  5.824   Mean   :-59.30
##  jeremy  :8   3rd Qu.: 32.5050   3rd Qu.:  6.135   3rd Qu.:-63.50
##  pedro   :3   Max.   :129.0000   Max.   : 27.800   Max.   :162.00
##  total_accel_belt  gyros_belt_x     gyros_belt_y     gyros_belt_z
##  Min.   : 2.00    Min.   :-0.500   Min.   :-0.050   Min.   :-0.4800
##  1st Qu.: 3.00    1st Qu.:-0.070   1st Qu.:-0.005   1st Qu.:-0.1375
##  Median : 4.00    Median : 0.020   Median : 0.000   Median :-0.0250
##  Mean   : 7.55    Mean   :-0.045   Mean   : 0.010   Mean   :-0.1005
##  3rd Qu.: 8.00    3rd Qu.: 0.070   3rd Qu.: 0.020   3rd Qu.: 0.0000
##  Max.   :21.00    Max.   : 0.240   Max.   : 0.110   Max.   : 0.0500
##   accel_belt_x     accel_belt_y     accel_belt_z     magnet_belt_x
##  Min.   :-48.00   Min.   :-16.00   Min.   :-187.00   Min.   :-13.00
```

```
##  1st Qu.:-19.00   1st Qu.:  2.00   1st Qu.: -24.00   1st Qu.:  5.50
##  Median :-13.00   Median :  4.50   Median :  27.00   Median : 33.50
##  Mean   :-13.50   Mean   : 18.35   Mean   : -17.60   Mean   : 35.15
##  3rd Qu.: -8.75   3rd Qu.: 25.50   3rd Qu.:  38.25   3rd Qu.: 46.25
##  Max.   : 46.00   Max.   : 72.00   Max.   :  49.00   Max.   :169.00
##  magnet_belt_y   magnet_belt_z       roll_arm         pitch_arm
##  Min.   :566.0   Min.   :-426.0   Min.   :-137.00   Min.   :-63.800
##  1st Qu.:578.5   1st Qu.:-398.5   1st Qu.:   0.00   1st Qu.: -9.188
##  Median :600.5   Median :-313.5   Median :   0.00   Median :  0.000
##  Mean   :601.5   Mean   :-346.9   Mean   :  16.42   Mean   : -3.950
##  3rd Qu.:631.2   3rd Qu.:-305.0   3rd Qu.:  71.53   3rd Qu.:  3.465
##  Max.   :638.0   Max.   :-291.0   Max.   : 152.00   Max.   : 55.000
##     yaw_arm        total_accel_arm  gyros_arm_x      gyros_arm_y
##  Min.   :-167.00   Min.   : 3.00   Min.   :-3.710   Min.   :-2.0900
##  1st Qu.: -60.15   1st Qu.:20.25   1st Qu.:-0.645   1st Qu.:-0.6350
##  Median :   0.00   Median :29.50   Median : 0.020   Median :-0.0400
##  Mean   :  -2.80   Mean   :26.40   Mean   : 0.077   Mean   :-0.1595
##  3rd Qu.:  25.50   3rd Qu.:33.25   3rd Qu.: 1.248   3rd Qu.: 0.2175
##  Max.   : 178.00   Max.   :44.00   Max.   : 3.660   Max.   : 1.8500
##   gyros_arm_z       accel_arm_x      accel_arm_y      accel_arm_z
##  Min.   :-0.6900   Min.   :-341.0   Min.   :-65.00   Min.   :-404.00
##  1st Qu.:-0.1800   1st Qu.:-277.0   1st Qu.: 52.25   1st Qu.:-128.50
##  Median :-0.0250   Median :-194.5   Median :112.00   Median : -83.50
##  Mean   : 0.1205   Mean   :-134.6   Mean   :103.10   Mean   : -87.85
##  3rd Qu.: 0.5650   3rd Qu.:   5.5   3rd Qu.:168.25   3rd Qu.: -27.25
##  Max.   : 1.1300   Max.   : 106.0   Max.   :245.00   Max.   :  93.00
##   magnet_arm_x      magnet_arm_y     magnet_arm_z    roll_dumbbell
##  Min.   :-428.00   Min.   :-307.0   Min.   :-499.0   Min.   :-111.118
##  1st Qu.:-373.75   1st Qu.: 205.2   1st Qu.: 403.0   1st Qu.:   7.494
##  Median :-265.00   Median : 291.0   Median : 476.5   Median :  50.403
##  Mean   : -38.95   Mean   : 239.4   Mean   : 369.8   Mean   :  33.760
##  3rd Qu.: 250.50   3rd Qu.: 358.8   3rd Qu.: 517.0   3rd Qu.:  58.129
##  Max.   : 750.00   Max.   : 474.0   Max.   : 633.0   Max.   : 123.984
##  pitch_dumbbell    yaw_dumbbell       total_accel_dumbbell
##  Min.   :-54.97   Min.   :-103.3200   Min.   : 1.0
##  1st Qu.:-51.89   1st Qu.: -75.2809   1st Qu.: 7.0
##  Median :-40.81   Median :  -8.2863   Median :15.5
##  Mean   :-19.47   Mean   :  -0.9385   Mean   :17.2
##  3rd Qu.: 16.12   3rd Qu.:  55.8335   3rd Qu.:29.0
##  Max.   : 96.87   Max.   : 132.2337   Max.   :31.0
##  gyros_dumbbell_x  gyros_dumbbell_y  gyros_dumbbell_z accel_dumbbell_x
##  Min.   :-1.0300   Min.   :-1.1100   Min.   :-1.180   Min.   :-159.00
##  1st Qu.: 0.1600   1st Qu.:-0.2100   1st Qu.:-0.485   1st Qu.:-140.25
##  Median : 0.3600   Median : 0.0150   Median :-0.280   Median : -19.00
##  Mean   : 0.2690   Mean   : 0.0605   Mean   :-0.266   Mean   : -47.60
##  3rd Qu.: 0.4625   3rd Qu.: 0.1450   3rd Qu.:-0.165   3rd Qu.:  15.75
##  Max.   : 1.0600   Max.   : 1.9100   Max.   : 1.100   Max.   : 185.00
##  accel_dumbbell_y accel_dumbbell_z magnet_dumbbell_x magnet_dumbbell_y
##  Min.   :-30.00   Min.   :-221.0   Min.   :-576.0    Min.   :-558.0
##  1st Qu.:  5.75   1st Qu.:-192.2   1st Qu.:-528.0    1st Qu.: 259.5
##  Median : 71.50   Median :  -3.0   Median :-508.5    Median : 316.0
##  Mean   : 70.55   Mean   : -60.0   Mean   :-304.2    Mean   : 189.3
##  3rd Qu.:151.25   3rd Qu.:  76.5   3rd Qu.:-317.0    3rd Qu.: 348.2
##  Max.   :166.00   Max.   : 100.0   Max.   : 523.0    Max.   : 403.0
```

```
##  magnet_dumbbell_z  roll_forearm     pitch_forearm      yaw_forearm
##  Min.   :-164.00   Min.   :-176.00   Min.   :-63.500   Min.   :-168.000
##  1st Qu.: -33.00   1st Qu.: -40.25   1st Qu.:-11.457   1st Qu.: -93.375
##  Median :  49.50   Median :  94.20   Median :  8.830   Median : -19.250
##  Mean   :  71.40   Mean   :  38.66   Mean   :  7.099   Mean   :   2.195
##  3rd Qu.:  96.25   3rd Qu.: 143.25   3rd Qu.: 28.500   3rd Qu.: 104.500
##  Max.   : 368.00   Max.   : 176.00   Max.   : 59.300   Max.   : 159.000
##  total_accel_forearm gyros_forearm_x   gyros_forearm_y   gyros_forearm_z
##  Min.   :21.00       Min.   :-1.0600   Min.   :-5.9700   Min.   :-1.2600
##  1st Qu.:24.00       1st Qu.:-0.5850   1st Qu.:-1.2875   1st Qu.:-0.0975
##  Median :32.50       Median : 0.0200   Median : 0.0350   Median : 0.2300
##  Mean   :32.05       Mean   :-0.0200   Mean   :-0.0415   Mean   : 0.2610
##  3rd Qu.:36.75       3rd Qu.: 0.2925   3rd Qu.: 2.0475   3rd Qu.: 0.7625
##  Max.   :47.00       Max.   : 1.3800   Max.   : 4.2600   Max.   : 1.8000
##  accel_forearm_x  accel_forearm_y  accel_forearm_z  magnet_forearm_x
##  Min.   :-212.0   Min.   :-331.0   Min.   :-282.0   Min.   :-714.0
##  1st Qu.:-114.8   1st Qu.:   8.5   1st Qu.:-199.0   1st Qu.:-427.2
##  Median :  86.0   Median : 138.0   Median :-148.5   Median :-189.5
##  Mean   :  38.8   Mean   : 125.3   Mean   : -93.7   Mean   :-159.2
##  3rd Qu.: 166.2   3rd Qu.: 268.0   3rd Qu.: -31.0   3rd Qu.:  41.5
##  Max.   : 232.0   Max.   : 406.0   Max.   : 179.0   Max.   : 532.0
##  magnet_forearm_y magnet_forearm_z
##  Min.   :-787.0   Min.   :-32.0
##  1st Qu.:-328.8   1st Qu.:275.2
##  Median : 487.0   Median :491.5
##  Mean   : 191.8   Mean   :460.2
##  3rd Qu.: 720.8   3rd Qu.:661.5
##  Max.   : 800.0   Max.   :884.0
```

Model Fit

creation of training and a testing set

```
set.seed(1342)
inTrain <- createDataPartition(y=train2$classe,
                               p=0.75, list=FALSE)
training <- train2[inTrain,]
testing <- train2[-inTrain,]
```

set training parameters: we will run 5-fold cross validation

```
fitControl <- trainControl(method = "cv", number=5)
```

knn

We run a k-nearest neighbours models with 1, 2 and 5 neighbours

```
tuneGrid <-  expand.grid(k = c(1,2,5))
set.seed(32343)
modelFit <- train(classe ~. ,data=training, method="knn",
```

```
                  trControl = fitControl, tuneGrid=tuneGrid, preProcess=c("center", "scale"))
modelFit
```

```
## k-Nearest Neighbors
##
## 14718 samples
##    53 predictor
##     5 classes: 'A', 'B', 'C', 'D', 'E'
##
## Pre-processing: centered (57), scaled (57)
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 11775, 11775, 11775, 11773, 11774
## Resampling results across tuning parameters:
##
##   k  Accuracy   Kappa
##   1  0.9866830  0.9831547
##   2  0.9730263  0.9658804
##   5  0.9574676  0.9461943
##
## Accuracy was used to select the optimal model using  the largest value.
## The final value used for the model was k = 1.
```

The best result is achieved for k=1, with an accuracy of 98.66%

naive Bayes

```
library("naivebayes")
library("fastICA")
tuneGrid <-  expand.grid(fL = c(0,0.5,1),
                         usekernel = c(TRUE,FALSE),
                         adjust=c(TRUE,FALSE))

set.seed(32343)
modelFit <- train(classe ~. ,data=training, method="naive_bayes",
                  trControl = fitControl,  tuneGrid=tuneGrid)
modelFit
```

```
## Naive Bayes
##
## 14718 samples
##    53 predictor
##     5 classes: 'A', 'B', 'C', 'D', 'E'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 11775, 11775, 11775, 11773, 11774
## Resampling results across tuning parameters:
##
##   fL   usekernel  adjust  Accuracy   Kappa
##   0.0  FALSE      FALSE   0.4888584  0.3684171
##   0.0  FALSE       TRUE   0.4888584  0.3684171
##   0.0   TRUE      FALSE   0.7363102  0.6627087
##   0.0   TRUE       TRUE   0.7363102  0.6627087
##   0.5  FALSE      FALSE   0.4888584  0.3684171
```

```
##   0.5  FALSE       TRUE   0.4888584  0.3684171
##   0.5   TRUE      FALSE   0.7363102  0.6627087
##   0.5   TRUE       TRUE   0.7363102  0.6627087
##   1.0  FALSE      FALSE   0.4888584  0.3684171
##   1.0  FALSE       TRUE   0.4888584  0.3684171
##   1.0   TRUE      FALSE   0.7363102  0.6627087
##   1.0   TRUE       TRUE   0.7363102  0.6627087
##
## Accuracy was used to select the optimal model using  the largest value.
## The final values used for the model were fL = 0, usekernel = TRUE
##  and adjust = FALSE.
```

The best result is an accuracy of 73.6%

random Forest

```
library("randomForest")
rfGrid <-  expand.grid(mtry = 7)
set.seed(32343)
model <- train(classe~., data=training, trControl=fitControl, method="rf",
               tuneGrid=rfGrid, ntree=100)
model
```

```
## Random Forest
##
## 14718 samples
##    53 predictor
##     5 classes: 'A', 'B', 'C', 'D', 'E'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 11775, 11775, 11775, 11773, 11774
## Resampling results:
##
##   Accuracy  Kappa
##   0.99307   0.9912331
##
## Tuning parameter 'mtry' was held constant at a value of 7
```

```
model$finalModel
```

```
##
## Call:
##  randomForest(x = x, y = y, ntree = 100, mtry = param$mtry)
##                Type of random forest: classification
##                      Number of trees: 100
## No. of variables tried at each split: 7
##
##         OOB estimate of  error rate: 0.62%
## Confusion matrix:
##      A    B    C    D    E  class.error
## A 4182    1    1    1    0 0.0007168459
## B   14 2827    7    0    0 0.0073735955
## C    1   23 2540    3    0 0.0105181145
```

```
## D    0    0   27 2382    3 0.0124378109
## E    0    0    2    8 2696 0.0036954915
```

Accuracy is predicted at 99.31%, for an out of bag error estimate of 0.62%. This is our best model, let's double-check its accuracy on our test set

```
predictions <- predict.train(model,newdata=testing[,1:53])
confusionMatrix(predictions,testing$classe)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1392    3    0    0    0
##          B    2  943    9    0    0
##          C    0    3  846   11    2
##          D    0    0    0  793    3
##          E    1    0    0    0  896
##
## Overall Statistics
##
##                Accuracy : 0.9931
##                  95% CI : (0.9903, 0.9952)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9912
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9978   0.9937   0.9895   0.9863   0.9945
## Specificity            0.9991   0.9972   0.9960   0.9993   0.9998
## Pos Pred Value         0.9978   0.9885   0.9814   0.9962   0.9989
## Neg Pred Value         0.9991   0.9985   0.9978   0.9973   0.9988
## Prevalence             0.2845   0.1935   0.1743   0.1639   0.1837
## Detection Rate         0.2838   0.1923   0.1725   0.1617   0.1827
## Detection Prevalence   0.2845   0.1945   0.1758   0.1623   0.1829
## Balanced Accuracy      0.9985   0.9954   0.9928   0.9928   0.9971
```

The results for the test set are consistent with the cross-validation, we now apply our model to the second set and create our submission.

```
submission <- predict.train(model,newdata=test2)
submission
```

```
##  [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
```