Practice Machine Learning

Andres Felipe Lagos Ruiz

12/12/2020

Introdution

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 dumbell of 6 participants. They were asked to perform barbell lifts correctly and incorrectly in 5 different ways

```
suppressWarnings(suppressMessages(library(dplyr)))
suppressWarnings(suppressMessages(library("PerformanceAnalytics")))
suppressWarnings(suppressMessages(library(corrplot)))
suppressWarnings(suppressMessages(library(caret)))
suppressWarnings(suppressMessages(library(rattle)))
suppressWarnings(suppressMessages(library(rattle)))
```

Preprocesing of the data

the fist step loading the data

```
download.file("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv","training.csv")
download.file("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv","testing.csv")
training<-read.csv("training.csv")
testing<-read.csv("testing.csv")
```

taking only some variables

```
training$user_name=as.factor(training$user_name)
testing$user_name=as.factor(testing$user_name)
training$classe=as.factor(training$classe)
#using only the variables that start with gyros,accel and magnet
training<-select(training,user_name,grep('^accel|^gyros|^magnet',names(training)),classe)
testing<-select(testing,user_name,grep('^accel|^gyros|^magnet',names(testing)))
#searching for nulls
which(is.na(training))
```

```
## integer(0)
```

```r
correlation<-cor(training[, -c(1,38)])
co<-abs(correlation)>0.8
which(co==T, arr.ind=TRUE)
```

```
##                   row col
## gyros_belt_x        1   1
## gyros_belt_y        2   2
## gyros_belt_z        3   3
## accel_belt_x        4   4
## magnet_belt_x       7   4
## accel_belt_y        5   5
## accel_belt_z        6   5
## accel_belt_y        5   6
## accel_belt_z        6   6
## accel_belt_x        4   7
## magnet_belt_x       7   7
## magnet_belt_y       8   8
## magnet_belt_z       9   9
## gyros_arm_x        10  10
## gyros_arm_y        11  10
## gyros_arm_x        10  11
## gyros_arm_y        11  11
## gyros_arm_z        12  12
## accel_arm_x        13  13
## magnet_arm_x       16  13
## accel_arm_y        14  14
## accel_arm_z        15  15
## accel_arm_x        13  16
## magnet_arm_x       16  16
## magnet_arm_y       17  17
## magnet_arm_z       18  17
## magnet_arm_y       17  18
## magnet_arm_z       18  18
## gyros_dumbbell_x   19  19
## gyros_dumbbell_z   21  19
## gyros_forearm_z    30  19
## gyros_dumbbell_y   20  20
## gyros_dumbbell_x   19  21
## gyros_dumbbell_z   21  21
## gyros_forearm_z    30  21
## accel_dumbbell_x   22  22
## accel_dumbbell_y   23  23
## accel_dumbbell_z   24  24
## magnet_dumbbell_x  25  25
## magnet_dumbbell_y  26  26
## magnet_dumbbell_z  27  27
## gyros_forearm_x    28  28
## gyros_forearm_y    29  29
## gyros_forearm_z    30  29
## gyros_dumbbell_x   19  30
## gyros_dumbbell_z   21  30
## gyros_forearm_y    29  30
## gyros_forearm_z    30  30
```

```
## accel_forearm_x    31  31
## accel_forearm_y    32  32
## accel_forearm_z    33  33
## magnet_forearm_x   34  34
## magnet_forearm_y   35  35
## magnet_forearm_z   36  36
```

```
#it is deleted the variables
training<-training[,-c(11,16,18,21,30)]
testing<-testing[,-c(11,16,18,21,30)]
summary(training)
```

```
##     user_name     gyros_belt_x        gyros_belt_y       gyros_belt_z
##  adelmo  :3892   Min.   :-1.040000   Min.   :-0.64000   Min.   :-1.4600
##  carlitos:3112   1st Qu.:-0.030000   1st Qu.: 0.00000   1st Qu.:-0.2000
##  charles :3536   Median : 0.030000   Median : 0.02000   Median :-0.1000
##  eurico  :3070   Mean   :-0.005592   Mean   : 0.03959   Mean   :-0.1305
##  jeremy  :3402   3rd Qu.: 0.110000   3rd Qu.: 0.11000   3rd Qu.:-0.0200
##  pedro   :2610   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     gyros_arm_y       gyros_arm_z
##  Min.   :354.0   Min.   :-623.0   Min.   :-3.4400   Min.   :-2.3300
##  1st Qu.:581.0   1st Qu.:-375.0   1st Qu.:-0.8000   1st Qu.:-0.0700
##  Median :601.0   Median :-320.0   Median :-0.2400   Median : 0.2300
##  Mean   :593.7   Mean   :-345.5   Mean   :-0.2571   Mean   : 0.2695
##  3rd Qu.:610.0   3rd Qu.:-306.0   3rd Qu.: 0.1400   3rd Qu.: 0.7200
##  Max.   :673.0   Max.   : 293.0   Max.   : 2.8400   Max.   : 3.0200
##   accel_arm_x       accel_arm_y      magnet_arm_x     magnet_arm_z
##  Min.   :-404.00   Min.   :-318.0   Min.   :-584.0   Min.   :-597.0
##  1st Qu.:-242.00   1st Qu.: -54.0   1st Qu.:-300.0   1st Qu.: 131.2
##  Median : -44.00   Median :  14.0   Median : 289.0   Median : 444.0
##  Mean   : -60.24   Mean   :  32.6   Mean   : 191.7   Mean   : 306.5
##  3rd Qu.:  84.00   3rd Qu.: 139.0   3rd Qu.: 637.0   3rd Qu.: 545.0
##  Max.   : 437.00   Max.   : 308.0   Max.   : 782.0   Max.   : 694.0
##  gyros_dumbbell_x    gyros_dumbbell_z  accel_dumbbell_x  accel_dumbbell_y
##  Min.   :-204.0000   Min.   : -2.380   Min.   :-419.00   Min.   :-189.00
##  1st Qu.:  -0.0300   1st Qu.: -0.310   1st Qu.: -50.00   1st Qu.:  -8.00
##  Median :   0.1300   Median : -0.130   Median :  -8.00   Median :  41.50
##  Mean   :   0.1611   Mean   : -0.129   Mean   : -28.62   Mean   :  52.63
##  3rd Qu.:   0.3500   3rd Qu.:  0.030   3rd Qu.:  11.00   3rd Qu.: 111.00
##  Max.   :   2.2200   Max.   :317.000   Max.   : 235.00   Max.   : 315.00
##  accel_dumbbell_z  magnet_dumbbell_x magnet_dumbbell_y magnet_dumbbell_z
##  Min.   :-334.00   Min.   :-643.0    Min.   :-3600     Min.   :-262.00
##  1st Qu.:-142.00   1st Qu.:-535.0    1st Qu.:  231     1st Qu.: -45.00
##  Median :  -1.00   Median :-479.0    Median :  311     Median :  13.00
##  Mean   : -38.32   Mean   :-328.5    Mean   :  221     Mean   :  46.05
##  3rd Qu.:  38.00   3rd Qu.:-304.0    3rd Qu.:  390     3rd Qu.:  95.00
##  Max.   : 318.00   Max.   : 592.0    Max.   :  633     Max.   : 452.00
```

```
##  gyros_forearm_x   gyros_forearm_z    accel_forearm_x   accel_forearm_y
##  Min.   :-22.000   Min.   : -8.0900   Min.   :-498.00   Min.   :-632.0
##  1st Qu.: -0.220   1st Qu.: -0.1800   1st Qu.:-178.00   1st Qu.:  57.0
##  Median :  0.050   Median :  0.0800   Median : -57.00   Median : 201.0
##  Mean   :  0.158   Mean   :  0.1512   Mean   : -61.65   Mean   : 163.7
##  3rd Qu.:  0.560   3rd Qu.:  0.4900   3rd Qu.:  76.00   3rd Qu.: 312.0
##  Max.   :  3.970   Max.   :231.0000   Max.   : 477.00   Max.   : 923.0
##  accel_forearm_z   magnet_forearm_x  magnet_forearm_y magnet_forearm_z classe
##  Min.   :-446.00   Min.   :-1280.0   Min.   :-896.0   Min.   :-973.0   A:5580
##  1st Qu.:-182.00   1st Qu.: -616.0   1st Qu.:   2.0   1st Qu.: 191.0   B:3797
##  Median : -39.00   Median : -378.0   Median : 591.0   Median : 511.0   C:3422
##  Mean   : -55.29   Mean   : -312.6   Mean   : 380.1   Mean   : 393.6   D:3216
##  3rd Qu.:  26.00   3rd Qu.:  -73.0   3rd Qu.: 737.0   3rd Qu.: 653.0   E:3607
##  Max.   : 291.00   Max.   :  672.0   Max.   :1480.0   Max.   :1090.0
```

Training the network

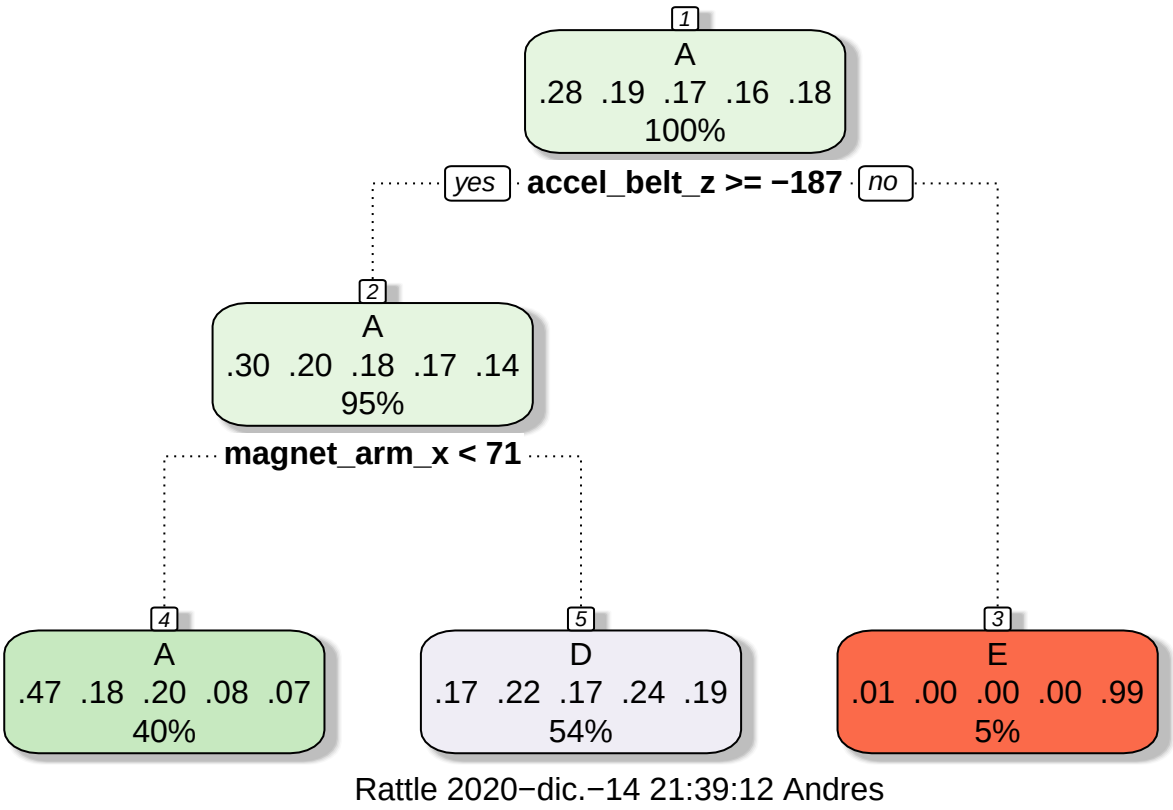
```
modFit<-train(classe~.,data=training,method="rpart")
print(modFit$finalModel)
```

```
## n= 19622
##
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
## 1) root 19622 14042 A (0.28 0.19 0.17 0.16 0.18)
##   2) accel_belt_z>=-187.5 18587 13013 A (0.3 0.2 0.18 0.17 0.14)
##     4) magnet_arm_x< 70.5 7907  4196 A (0.47 0.18 0.2 0.081 0.071) *
##     5) magnet_arm_x>=70.5 10680  8104 D (0.17 0.22 0.17 0.24 0.19) *
##   3) accel_belt_z< -187.5 1035     7 E (0.0058 0.00097 0 0 0.99) *
```

```
predict(modFit,newdata = testing)
```

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

```
fancyRpartPlot(modFit$finalModel)
```



##conclucions without the classe variable in the testing data, it not possible to say if the model classified in a good way, but seen that only has 3 of the 5 classes it is possible that it fails. it tried to used other models like a random forest but it takes more than 30 min.