The Human Activity Recognition Research

Manuel Cerda

12/24/2017

About the Analisis

The purpose of this analysis is to classify the manner how people perfom an excercise based on the *Human Ativity Recognition* research. We will use data from accelerometers on the belt, forearm, arm, and dumbell of 6 participants that were asked to perform barbell lifts correctly and incorrectly in 5 different ways.

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.

About the Data

The data was collected from devices such as *Jawbone Up*, *Nike FuelBand*, and *Fitbit*. 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.

The training data for this project are available here: https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv

The test data are available here: https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv

The data for this project come from this source: http://groupware.les.inf.puc-rio.br/har

Exploratory Analysis

```
require(ggplot2)
require(corrplot)
require(rpart)
require(rpart.plot)
require(caret)
require(randomForest)
require(rattle)

set.seed(123)

# Already downloaded to save some time
train <- read.csv("./data/pml-training.csv")
test  <- read.csv("./data/pml-testing.csv")
dim(train)
```

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

```
# Cleaning the data
trainClasse <- train$classe
train <- train[, colSums(is.na(train)) == 0]
train <- train[, !(grepl("^X|timestamp|window", names(train)))]
train <- train[, sapply(train, is.numeric)]
train$classe <- trainClasse
```

```
test  <- test[, colSums(is.na(test)) == 0]
test  <- test[, !(grepl("^X|timestamp|window", names(test)))]
test  <- test[, sapply(test, is.numeric)]
test[,"problem_id"] <- list(NULL)

dim(train)
```

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

```
# Desired clasifications
levels(train$classe)
```

```
## [1] "A" "B" "C" "D" "E"
```

Training Model

```
# Using a Random Forest algorithm
rf <- randomForest(classe ~ ., data=train, ntree=100)
rf
```

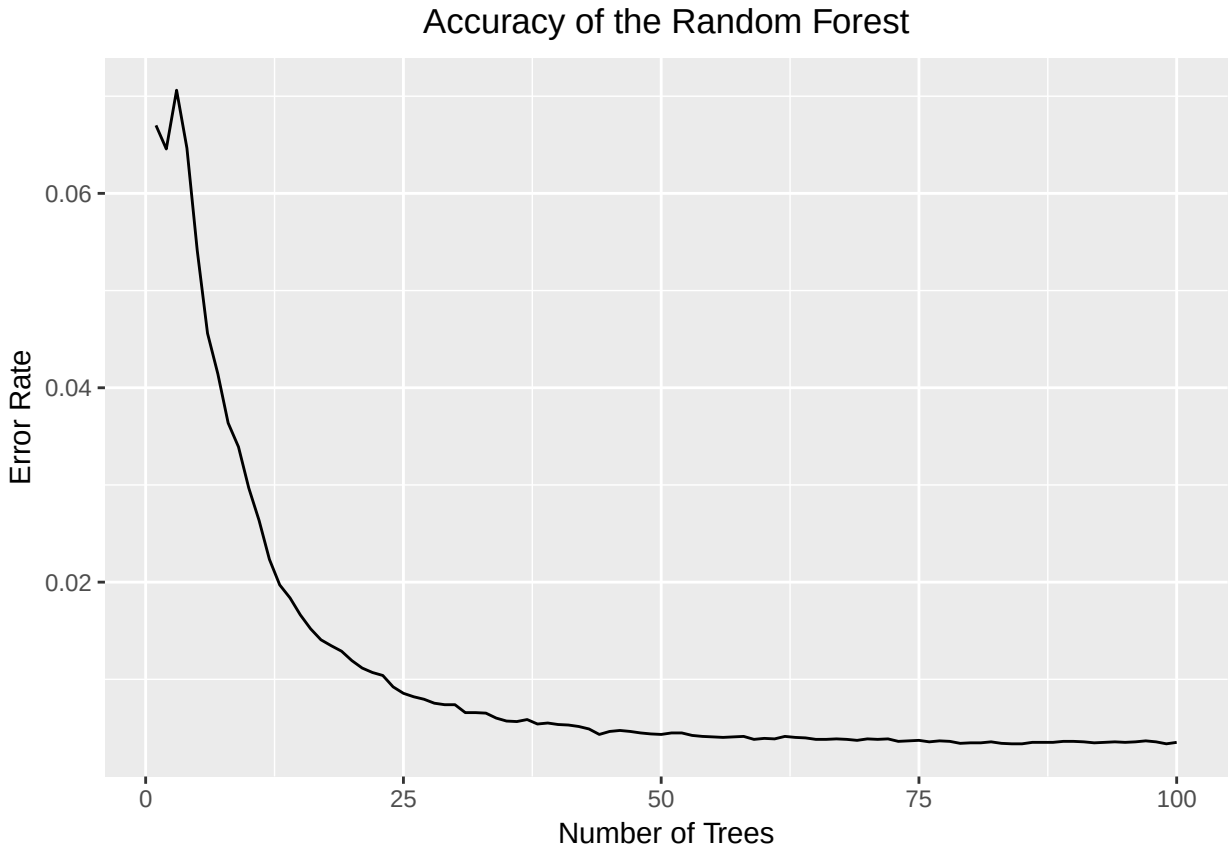
```
##
## Call:
##  randomForest(formula = classe ~ ., data = train, ntree = 100)
##                Type of random forest: classification
##                      Number of trees: 100
## No. of variables tried at each split: 7
##
##         OOB estimate of  error rate: 0.35%
## Confusion matrix:
##      A    B    C    D    E  class.error
## A 5577    3    0    0    0 0.0005376344
## B    9 3783    4    0    1 0.0036871214
## C    0   10 3412    0    0 0.0029222677
## D    0    0   31 3181    4 0.0108830846
## E    0    0    2    5 3600 0.0019406709
```

Test Prediction

```
pred <- predict(rf, test)
pred
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##  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
```

Appendix: Plots



Classification Tree (Training Data)