Practical Machine Learning - Prediction Assignment Writeup

Arturo Penas Rial

November 22, 2020

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 dumbell 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://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

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. 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.

Executive Summary

This is a report of the Peer Assessment project from the Practical Machine Learning course. The goal of this analysis is to predict the manner in which the six participants performed their exercises. The machine learning algorithm, uses the "classe" variable in the training set, is applied to the 20 test cases available in the test data.

What you should submit

The goal of your project is to predict the manner in which they did the exercise. This is the "classe" variable in the training set. You may use any of the other variables to predict with. You should create a report describing how you built your model, how you used cross validation, what you think the expected out of sample error is, and why you made the choices you did. You will also use your prediction model to predict 20 different test cases.

Your submission should consist of a link to a Github repo with your R markdown and compiled HTML file describing your analysis. Please constrain the text of the writeup to < 2000 words and the number of figures to be less than 5. It will make it easier for the graders if you submit a repo with a gh-pages branch so the HTML page can be viewed online (and you always want to make it easy on graders :-). You should also apply your machine learning algorithm to the 20 test cases available in the test data above. Please submit your predictions in appropriate format to the programming assignment for automated grading. See the programming assignment for additional details. Reproducibility

Due to security concerns with the exchange of R code, your code will not be run during the evaluation by your classmates. Please be sure that if they download the repo, they will be able to view the compiled HTML version of your analysis.

Data Selection

We first read training data and testing data from "*pml-training.csv*" and "*pml-testing.csv*", respectively. The variables we choosen for training are **user_name**, **classe** and information from accelerometers. Note that there is no **classe** in "*pml-testing.csv*".

```
rawdata <- read.csv("./pml-training.csv")
rawdata_test <- read.csv("./pml-testing.csv")

col_select_train <- c("user_name","gyros_belt_x","gyros_belt_y","gyros_belt_z","accel_belt_x","accel_bel
data_selected <- rawdata[rawdata$new_window != "yes",col_select_train]

col_select_test <- col_select_train[1:length(col_select_train)-1]
testdata_selected <- rawdata_test[rawdata_test$new_window != "yes",col_select_test]
```

Partition of Training Data

In training data, we use 75% as true training data and 25% as quizing data.

```
set.seed(10)
inTrain <- createDataPartition(data_selected$classe, p = 3/4, list=F)
training <- data_selected[inTrain,]
quizing <- data_selected[-inTrain,]
```

Train the Model

We use **random forest** as classification method. Although it takes about 1 hour for fitting model, it provide 100% accuracy for quizing data.

```
ptm <- proc.time()
modFit <- train(classe ~ .,data=data_selected,method="rf")
time <- proc.time() - ptm
pred_quiz <- predict(modFit,quizing)
str(quizing$classe)
```

```
##  chr [1:4802] "A" "A" "A" "A" "A" "A" "A" "A" "A" "A" "A" "A" "A" "A" "A" ...
```

```
str(pred_quiz)
```

```
##  Factor w/ 5 levels "A","B","C","D",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
cmfit <- confusionMatrix(table(quizing$classe,pred_quiz))
# cmfit <- confusionMatrix(quizing$classe,pred_quiz)
```

```r
time
```

```
##    user  system elapsed 
## 3876.13   42.47 3964.36
```

```r
cmfit$table
```

```
##    pred_quiz
##        A    B    C    D    E
##   A 1367    0    0    0    0
##   B    0  929    0    0    0
##   C    0    0  838    0    0
##   D    0    0    0  786    0
##   E    0    0    0    0  882
```

Prediction for testing data

Finally we predict testing data using fitted model.

```r
pred_test <- predict(modFit,testdata_selected)
pred_test
```

```
##  [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
```

Result of Course Project Prediction Quiz

The fitted model gives 100% accuracy for testing data.