

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, the goal is 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.

Variable “Class” description

Six young health participants were asked to perform one set of 10 repetitions of the Unilateral Dumbbell Biceps Curl in five different fashions: exactly according to the specification (Class A), throwing the elbows to the front (Class B), lifting the dumbbell only halfway (Class C), lowering the dumbbell only halfway (Class D) and throwing the hips to the front (Class E).

Project requirements

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.

1.The submission of project should consist of a link to a Github repo with R markdown and compiled HTML file describing your analysis. It You should also apply machine learning algorithm to the 20 test cases available in the test data above and submit predictions in appropriate format to the programming assignment for automated grading. See the programming assignment for additional details.

Load libraries

```
library(AppliedPredictiveModeling)
library(caret)
library(lattice)
library(ggplot2)
library(rpart)
library(rpart.plot)
#install.packages("randomForest")
library(randomForest)
```

Read Files ,cleaning & review data

```
setwd("F:/Research_Studies/Coursera_EDX/Data specialization Hoopkins/8-Practical Machine Learning/project")

#training dataset
training0=read.csv("pml - training.csv", na.strings=c("NA", "#DIV/0!", ""))
#names(training)
#str(training)
#head(training)
head(training0$classe)

## [1] A A A A A A
## Level: A B C D E

#dim(training) #[1] 19622 60

#testing dataset
testing0=read.csv("pml - testing.csv", na.strings=c("NA", "#DIV/0!", ""))
#names(testing)
#str(testing)
#dim(testing) #[1] 20 60
```

Cleaning data sets & decrease unnecessary variables due to RF model limitations (less than 53 variables required)

```
training<- training0[, colSums(is.na(training0)) == 0]
#dim(training)
testing <- testing0[, colSums(is.na(testing0)) == 0]
#dim(testing)

training <- training[, -c(1:7)] # remove 7 first variables
dim(training)
```

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

```
testing <- testing[, -c(1:7)] # remove 7 first variables
dim(testing)
```

```
## [1] 20 53
```

Random Number Generation

```
set.seed(20000)
```

Partitioning data for verification & validation

```
SSamples <- createDataPartition(y=training$classe, p=0.8, list=FALSE) # 0.8/0.2
STraining <- training[SSamples, ]
dim(STraining)
```

```
## [1] 15699 53
```

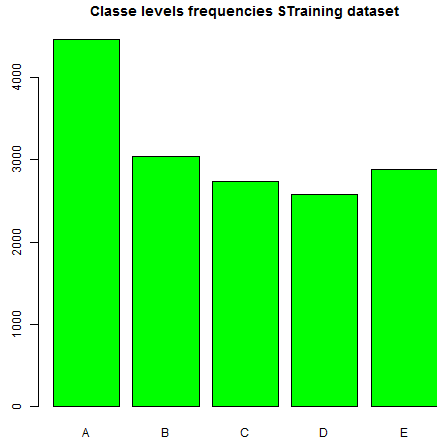
```
#head(STraining)
STesting <- training[-SSamples, ]
dim(STesting)
```

```
## [1] 3923 53
```

```
#head(STesting)
```

Training dataset Variable “class” barplot & frequencies

```
plot(STraining$classe, col="green", main="Classe levels frequencies STraining dataset")
```



```
a=nrow(STraining[STraining$classe=="A",]); b=nrow(STraining[STraining$classe=="B",]); c=nrow(STraining[STraining$classe=="C",]); d=nrow(STraining[STraining$classe=="D",]); e=nrow(STraining[STraining$classe=="E",])
classe=c(a, b, c, d, e)
classe
```

```
## [1] 4464 3038 2738 2573 2886
```

prediction model used: Random Forest (as per Lecture 3-3)

```
#names(STraining)
#dim(STraining)
```

```
#dim(STesting)
model_RF <- randomForest(classe ~., data=STraining, method="class") # randomForest model

prediction_RF <- predict(model_RF, STesting, type = "class") #predict Sample testing

# Test results of prediction Sample testing:
confusionMatrix(prediction_RF, STesting$classe)
```

## Confusion Matrix and Statistics						
##						
## Reference						
## Prediction	A	B	C	D	E	
## A	1116	0	0	0	0	
## B	0	757	3	0	0	
## C	0	2	681	8	0	
## D	0	0	0	635	2	
## E	0	0	0	0	719	
##						
## Overall Statistics						
##						
## Accuracy : 0.996						
## 95% CI : (0.994, 0.998)						
## No Information Rate : 0.284						
## P-Value [Acc > NIR] : <2e-16						
##						
## Kappa : 0.995						
## McNemar's Test P-Value : NA						
##						
## Statistics by Class:						
##						
## Class: A Class: B Class: C Class: D Class: E						
## Sensitivity	1.000	0.997	0.996	0.988	0.997	
## Specificity	1.000	0.999	0.997	0.999	1.000	
## Pos Pred Value	1.000	0.996	0.986	0.997	1.000	
## Neg Pred Value	1.000	0.999	0.999	0.998	0.999	
## Prevalence	0.284	0.193	0.174	0.164	0.184	
## Detection Rate	0.284	0.193	0.174	0.162	0.183	
## Detection Prevalence	0.284	0.194	0.176	0.162	0.183	
## Balanced Accuracy	1.000	0.998	0.996	0.993	0.999	

Reported Accuracy of R.F model is 0.996

Apply machine learning algorithm to the 20 test cases available in the test data above

```
final_Predict <- predict(model_RF, testing, type="class")
final_Predict

##  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
```

Submission prediction files

```
pml_write_files = function(x) {
  n = length(x)
  for(i in 1:n){
    filename = paste0("problem_id_", i, ".txt")
    write.table(x[i], file=filename, quote=FALSE, row.names=FALSE, col.names=FALSE)
  }
}
pml_write_files(final_Predict)
```

References

1)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.

2)<http://groupware.les.inf.puc-rio.br/har#ixzz3MWSmduBn> (21.12.2014-11:23 AM)