Reproducing Qualitative Activity Recognition of Weight Lifting Exercises

Maria Velasco, Dec 21 2014

Notes

This document includes the Practical Machine Learning - Coursera Class Dec 2014 Project by mvelasco, published in https://github.com/mvelascoc/MachLearning

Exploratory data analysis done to complete this project included in the exploratory.md file.

Summary

Movement data is collected [1] performing a set of repetitions on 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).

The purpose of this project is to define model that can predict, based on the movement data, if the movement was done correctly. From a set of 160 variables, a subset of 12 was selected. Using caret package Gradient Boosted Machine algorithm a model was built with an acuracy of 0.45.

Data Analysis

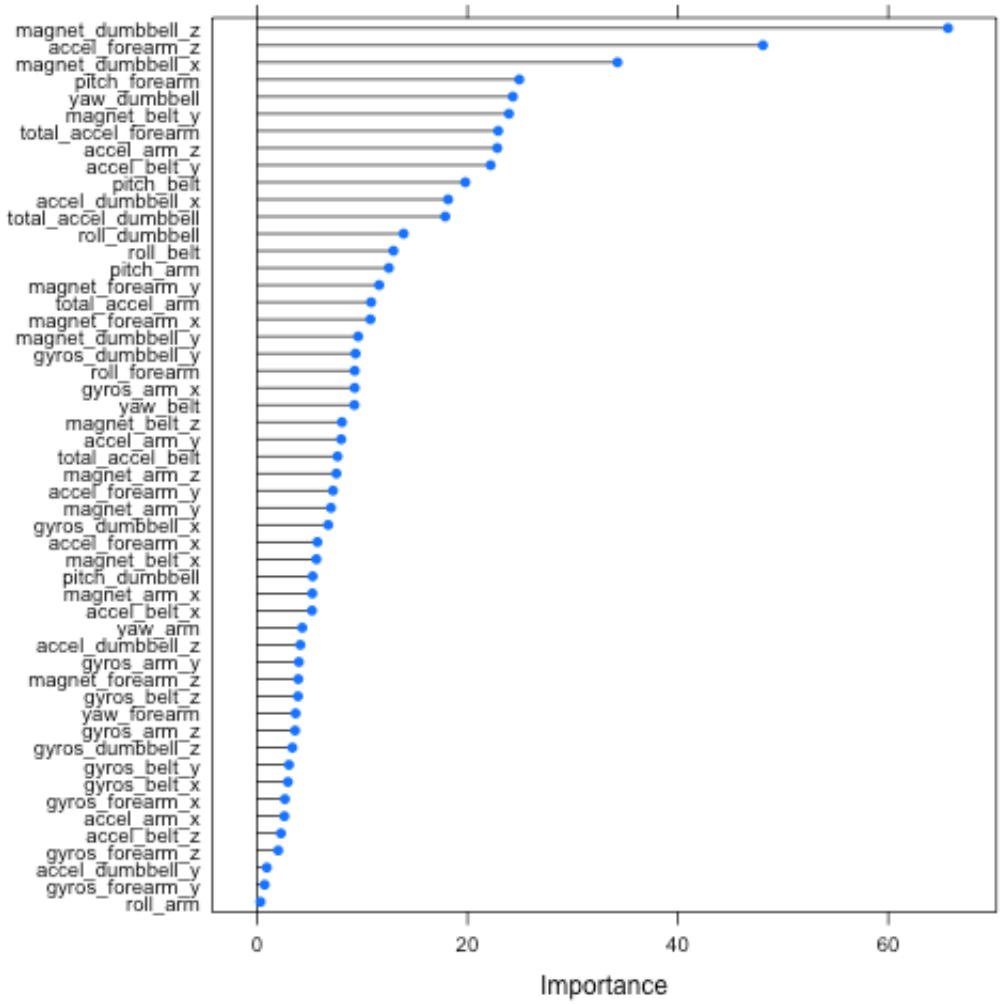
To define the features to be included in the algorith, several exploratory analysis were made. Deciding on using the most important features on a linear approach, the selected variables to be included were

Variable name	Overall importance
magnet_dumbbell_z	65.68

accel_forearm_z	48.09
magnet_dumbbell_x	34.25
pitch_forearm	24.92
yaw_dumbbell	24.31
magnet_belt_y	23.94
total_accel_forearm	22.91
accel_arm_z	22.83
accel_belt_y	22.20
pitch_belt	19.79
accel_dumbbell_x	18.15
total_accel_dumbbell	17.88

```
## Warning: package 'caret' was built under R version 3.1.1
```

```
## Loading required package: lattice
## Loading required package: ggplot2
## Find out what's changed in ggplot2 with
## news(Version == "1.0.0", package = "ggplot2")
```



Model Selection

Multiple models were proved using the caret package. Accuracy on the predictions using the training set were recorded, selecting the Gradient Boosted Machine gbm as the best one. Reported accuracies [1] using random forest model could not be reproduced.

Type of Model	Accuracy
lm	0.3014
rf	?
gbm	0.4547

brnn

Model Testing

Using the testing set of data, the confusion matrix is created. The model created does not have the accuracy of the one in the paper, but does have a significant prediction result, considered enough for this homwork project.

```
inTrain <- createDataPartition(y=target$class, p=0.75, list=FALSE)
training <- finaldata[inTrain,]
testing <- finaldata[-inTrain,]

modFitOp3 <- train(as.numeric(classe) ~ ., method="gbm", data=training)

predicted3 <- predict(modFitOp3, testing)
divide3 = cut(predicted3,5)
levels(divide3)[1] <- "A"
levels(divide3)[2] <- "B"
levels(divide3)[3] <- "C"
levels(divide3)[4] <- "D"
levels(divide3)[5] <- "E"

confusionMat3 <- confusionMatrix(divide3, testing$classe)
confusionMat3
```

References

[1] Data for this project comes from http://groupware.les.inf.puc-rio.br/har.

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.