Prediction Assignment Writeup

Yaduttam Ganeriwala

July 23, 2021

Objective

The purpose of this project was to quantify how well the participants performed a barbell lifting exercise and to classify the measurement read from an accelerometer into 5 different classes (Class A:Class E).

Please reference the links below for the data sources:

http://groupware.les.inf.puc-rio.br/har

https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv

https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv

```
library(caret)
library(rpart)
library(randomForest)
```

Install/load the required packages needed for the creation of the model

```
train<-read.csv("C:/Users/aao1009/Desktop/pml-training.csv",na.strings=c("NA","#DIV/0!",""))
test<-read.csv("C:/Users/aao1009/Desktop/pml-testing.csv",na.strings=c("NA","#DIV/0!",""))
```

Load the training and testing datasets

```
test_clean <- names(test[,colSums(is.na(test)) == 0]) [8:59]
clean_train<-train[,c(test_clean,"classe")]
clean_test<-test[,c(test_clean,"problem_id")]
```

Remove null columns and the first 7 columns that will not be used

```
dim(clean_test)
dim(clean_train)
```

Check the dimensions of the clean test and train sets

```
set.seed(100)
inTrain<-createDataPartition(clean_train$classe, p=0.7, list=FALSE)
training<-clean_train[inTrain,]
testing<-clean_train[-inTrain,]
dim(training)
dim(testing)
```

Split the data into the training and testing datasets

Predicting the outcome using 3 different models

```
lda_model<-train(classe~ ., data=training, method="lda")
set.seed(200)
predict<-predict(lda_model,testing)
confusionMatrix(predict,testing$classe)
```

LDA Model   The LDA model gave a 70% accuracy on the testing set, with the expected out of sample error around 30%.

```
decision_tree_model<-rpart(classe~ ., data=training,method="class")
set.seed(300)
predict<-predict(decision_tree_model,testing,type="class")
confusionMatrix(predict,testing$classe)
```

Decision Tree Model   The Decision Tree Model gave a 74% accuracy on the testing set, with the expected out of sample error around 26%.

```
random_forest_mod<-randomForest(classe~ ., data=training, ntree=500)
set.seed(300)
predict<-predict(random_forest_mod, testing, type ="class")
confusionMatrix(predict,testing$classe)
```

Random Forest Model   The Random Forest Model gave a 99.6% accuracy on the testing set, with the expected out of sample error around 0.4%.

Conclusion

The greatest accuracy was achieved using the Random Forest Model, which gave an accuracy of 99.6%. Hence, this model was further used to make predictions on the exercise performance for 20 participants.

```
predict<-predict(random_forest_mod, clean_test, type ="class")
predict
```