Practical Machine Learning peer

Sahaj G

11/11/20

##Introduction

In todays world , it is very easy to collect many types of personal data and use it for analysis thru variety of fitness bands which takes help of various monitoring various conditions and they can come in very handy and cheap. The aim of this project is to use all this collected statistical data from waise, arms, wrists etc while the people were asked to exercis in a certain way. We want to predict how they exercised exactly using `classe` variable in the training set.

Data description

The outcome variable of the data set is `classe`, a factor variable that have five different levels. For the given data set, all the six participants were asked to perform one set of 10 repetitions of the Unilateral Dumbbell Biceps Curl in five different methods which defines 5 different classes: 10reps of biseps dumbell singly curl in 5 diff methods

- Cl P : Same
- Cl Q : front elbow
- Cl R : Half dubell
- Cl S : lower half dumbell
- Cl T : hip front

Initial configuration

Loading and installing some imp R Packages.

```
#Data variables
training.file   <- './data/pml-training.csv'
test.cases.file <- './data/pml-testing.csv'
training.url    <- 'http://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv'
test.cases.url  <- 'http://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv'
#Directories
if (!file.exists("data")){
  dir.create("data")
}
if (!file.exists("data/submission")){
  dir.create("data/submission")
}
```

```r
#R-Packages
IscaretInstalled <- require("caret")
```

```
## Loading required package: caret
```

```
## Warning: package 'caret' was built under R version 3.6.3
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```r
if(!IscaretInstalled){
    install.packages("caret")
    library("caret")
    }
IsrandomForestInstalled <- require("randomForest")
```

```
## Loading required package: randomForest
```

```
## Warning: package 'randomForest' was built under R version 3.6.3
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```r
if(!IsrandomForestInstalled){
    install.packages("randomForest")
    library("randomForest")
    }
IsRpartInstalled <- require("rpart")
```

```
## Loading required package: rpart
```

```r
if(!IsRpartInstalled){
    install.packages("rpart")
    library("rpart")
    }
IsRpartPlotInstalled <- require("rpart.plot")
```

```
## Loading required package: rpart.plot
```

```
## Warning: package 'rpart.plot' was built under R version 3.6.3
```

```r
if(!IsRpartPlotInstalled){
    install.packages("rpart.plot")
    library("rpart.plot")
    }
```

```r
# Set seed for reproducability
set.seed(9999)
```

Data processing

Downloading of data and then proocessing it. Will do transformations and data cleaning to remove the unnecessary nullified values from raw data. Also some non important columns are removed.

```r
# Download data
download.file(training.url, training.file)
download.file(test.cases.url,test.cases.file )
# Clean data
training   <-read.csv(training.file, na.strings=c("NA","#DIV/0!", ""))
testing <-read.csv(test.cases.file , na.strings=c("NA", "#DIV/0!", ""))
training<-training[,colSums(is.na(training)) == 0]
testing <-testing[,colSums(is.na(testing)) == 0]
# Subset data
training   <-training[,-c(1:7)]
testing <-testing[,-c(1:7)]
```