Improving people’s training

Tomas Fuenzalida

29 de mayo de 2020

## Objetive

The objective of this project is to put into practice the knowledge acquired in the Machine Learning course, part of the Data Science specialization in coursera taught by John Hopkins University. For this, the data delivered by different devices will be used when performing a physical exercise on different people, which will be rated according to the technique used. With these observations, a classification model will be calibrated that allows estimating 20 observations that have not been classified.

## 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://web.archive.org/web/20161224072740/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://web.archive.org/web/20161224072740/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.

## Download, Pre-process and analisis Data

library(dplyr)

Download the training dataset

data <- read.table("pml-training.csv", header = TRUE, sep = ",", na.strings = c("","NA"))

I remove the variables that do not generate a contribution to a future model such as time

data <- data[,-c(1,3,4,5,6,7)]

I remove all the columns that have NA

data <- data[ , colSums(is.na(data)) == 0]