**CODE BOOK FOR ASSIGNEMENT OF WEEK 4**

1. **Description of the original (raw) data.**

**The raw data used is taken from:**

Human Activity Recognition Using Smartphones Dataset

Version 1.0

Authors:

Jorge L. Reyes-Ortiz, Davide Anguita, Alessandro Ghio, Luca Oneto.

Smartlab - Non Linear Complex Systems Laboratory

DITEN - Università degli Studi di Genova.

Via Opera Pia 11A, I-16145, Genoa, Italy.

activityrecognition@smartlab.ws

www.smartlab.ws

Data can be downloaded from: <https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>

1. **Description of Datasets containing raw data used in this assignement.**

- 'README.txt'

- 'features\_info.txt': Shows information about the variables used on the feature vector.

- 'features.txt': List of all features.

- 'activity\_labels.txt': Links the class labels with their activity name.

- 'train/X\_train.txt': Training set.

- 'train/y\_train.txt': Training labels.

- 'test/X\_test.txt': Test set.

- 'test/y\_test.txt': Test labels.

The following files are available for the train and test data. Their descriptions are equivalent.

- 'train/subject\_train.txt': Each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30.

- 'train/Inertial Signals/total\_acc\_x\_train.txt': The acceleration signal from the smartphone accelerometer X axis in standard gravity units 'g'. Every row shows a 128 element vector. The same description applies for the 'total\_acc\_x\_train.txt' and 'total\_acc\_z\_train.txt' files for the Y and Z axis.

- 'train/Inertial Signals/body\_acc\_x\_train.txt': The body acceleration signal obtained by subtracting the gravity from the total acceleration.

- 'train/Inertial Signals/body\_gyro\_x\_train.txt': The angular velocity vector measured by the gyroscope for each window sample. The units are radians/second.

Notes:

======

- Features are normalized and bounded within [-1,1].

- Each feature vector is a row on the text file.

1. **Transformations applied to datasets with raw data**

The objective of this assignement is to create one R script called run\_analysis.R that does the following.:

* Merges the training and the test sets to create one data set.
* Extracts only the measurements on the mean and standard deviation for each measurement.
* Uses descriptive activity names to name the activities in the data set
* Appropriately labels the data set with descriptive variable names.
* From the data set in step 4, creates a second, independent tidy data set with the average of each variable for each activity and each subject.

The following is a description of the transformation applied, using the code contained in **“run\_analysis.R”**

**General description**

# Read into R and display "subject\_train.txt".

# Has subject ID's and is of size 7352 x 1

# Read into R and display "y\_train.txt"

# Has Code of activities as 1 - 6. Of Size 7352 x 1

# Read into R and display "X\_train.txt"

# Is a dataframe with measurements. Size 7352 x 561

# Features.txt has the names of the columns of "X\_train.txt"

# Use it to substitute the column names ("V1-..."V561") for the descriptive names in "Features.txt"

# CODE : To facilitate comprehension, the code has been setup in "processes"

**# Process 1**

# 1. Create a dataframe with "train" data (7352 x 1)

# 1.1 Capture 561 colnames from "features.txt"

# 1.2 Change colnames of "X\_train.txt" using "features.txt" -> df\_train1 (7352 x 561)

# 1.3 Join data from "subject\_train.txt" with "y\_train.txt" with "df\_train1" -> df\_train2

## RESULTS IN A DATAFRAME "df\_train2" which has all data from "train" observations (7352 x 563)

**# Process 2**

# 2. Create a dataframe with "test" data (2947 x 561)

# 2.1 Capture 561 colnames from "features.txt"

# 2.2 Change colnames of "X\_test.txt" using "features.txt" -> df\_test1

# 2.3 Join data from "subject\_test.txt" + "y\_test.txt" with "df\_test1" -> df\_test2

## RESULTS IN A DATAFRAME "df\_test2" which has all data from "test" observations (2947 x 563)

**# Process 3**

# Join "df\_train2" and "df\_test2" dataframes

# 3.1 Correct first two column labels (they are both "V1") renaming to "id" and "code"

# for "df\_train2" and “df\_test2”

# 3.2 merge the dataframes to obtain "total" dataframe dimension 10,299 x 563

## RESULTS IN A DATAFRAME "total" which has all data from "test" and “train” observations (10,299 x 563)

**# Process 4**

# filter columns of “total” dataframe, keeping only the ones referring to "means and std deviations" .

# 4.1 filter dataframes

# 4.2 Select cols that have "mean" or "std" only. Keep "id" and "code"

my\_colnames\_final <- my\_colnames1[c(my\_colnames2)]

str(my\_colnames\_final)

## RESULTS IN A VECTOR "my\_colnames\_final" which has the names of the 88 selected columns

**# Process 5**

# Create a smaller data frame as result of selection in Process 4. THIS IS THE FINAL OUTPUT

# RESULTS IN A DATAFRAME "df\_final" which has all data from "test" observations (10,299 x 88)