DATA CALCULATIONS ON THE HUMAN ACTIVITY REPORT DATASET OBTAINED FROM UCI

OVERVIEW

This codebook provides information on how the UCI HAR Dataset has been used to compute certain parameters as part of a Data Cleaning Project.

For the purpose of this Data Cleaning (learning) project, the dataset was obtained here:
https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip

The original dataset and a full description can be obtained here:
http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

Information and interpretation of the dataset as *originally obtained* from the dataset is ***reproduced, without modification, in the next section***.

DATASET INFORMATION RECEIVED WITH THE DATASET (Reproduced)

```
==================================================================
Human Activity Recognition Using Smartphones Dataset
Version 1.0
==================================================================
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
==================================================================
```

The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.

The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low frequency components, therefore a filter with 0.3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain. See 'features_info.txt' for more details.

For each record it is provided:
======================================

- Triaxial acceleration from the accelerometer (total acceleration) and the estimated body acceleration.
- Triaxial Angular velocity from the gyroscope.
- A 561-feature vector with time and frequency domain variables.
- Its activity label.
- An identifier of the subject who carried out the experiment.

The dataset includes the following files:
===

- '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.

For more information about this dataset contact: activityrecognition@smartlab.ws

COMPUTATION PERFORMED ON THS DATASET

1. Measurements on the mean and standard deviation for each measurement was extracted in separate datasets from the 'test' and the 'train'ing data
2. The activities noted as numbers (from 1 to 6) in the data set, have been re-labeled with one of the following six descriptive activity names
    a. WALKING
    b. WALKING_UPSTAIRS
    c. WALKING_DOWNSTAIRS
    d. SITTING
    e. STANDING
    f. LAYING
3. The training and the test sets were then merged to create a single data set
4. The resulting super data set was then labeled with descriptive variable names taking reference from `features.txt` file provided with the original dataset

5. From the data set in Step 4, an independent tidy data set with the average of each variable for each activity (6 different activities) and each subject (30 different subjects - average taken after a group_by on subject and activity)
6. The resulting dataset contains 6 different activity rows for each of the 30 subjects, totaling a 180 rows

   NOTE: Most of this is the same as the instructions provided with the exercise, but please be cognizant that the order of the various steps are different (this was done to avoid creating huge datasets having over 500 variables)

The resulting data was written in a text file (UCI_HAR.txt) and can be obtained from these 2 locations:

- UCI_HAR
- ari-git/datacleaningprojectUCIHAR

Information provided in the next section will be beneficial in interpreting the information contained in the UCI_HAR.txt file.

INFORMATION ON THE DATA PRODUCED POST EXECUTING THE TRANSFORMATION STEPS OUTLINED ABOVE

> **NOTE 1:** The UCI_HAR.txt is a plain text file, created without using a delimiter. In R programming platform, the file can be loaded into a data frame by executing *read.table ("./UCI_HAR.txt", header = TRUE)* from an R console (assuming the file is in the R working directory.)

There are 68 variables (columns) spread across 180 readings (rows). Here is the interpretation of the variables:

- **subject_id**: Representative number for the human subject (range is 1-30)
- **activity_label**: Represents one of the activities (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING)
- **Mean_Time_Body_Acc_X_axis, Mean_Time_Body_Acc_Y_axis and Mean_Time_Body_Acc_Z_axis**: Average of the mean of the time domain component of the estimated body acceleration (on the X, Y and Z axes respectively). Gravity units 'g'
- **StDev_Time_Body_Acc_X_axis, StDev_Time_Body_Acc_Y_axis and StDev_Time_Body_Acc_Z_axis**: Average of the standard deviation of the time domain component of the estimated body acceleration (on the X, Y and Z axes respectively). Gravity units 'g'
- **Mean_Time_Gravity_Acc_X_axis, Mean_Time_Gravity_Acc_Y_axis and Mean_Time_Gravity_Acc_Z_axis**: Average of the mean of the time domain component of the estimated gravity acceleration from the accelerometer (on the X, Y and Z axes rspectively). Gravity units 'g'
- **StDev_Time_Gravity_Acc_X_axis, StDev_Time_Gravity_Acc_Y_axis and StDev_Time_Gravity_Acc_Z_axis**: Average of the standard deviation of the time domain component of the estimated gravity acceleration from the accelerometer (on the X, Y and Z axes respectively). Gravity units 'g'
- **Mean_Time_Body_Acc_Jerk_X_axis, Mean_Time_Body_Acc_Jerk_Y_axis and Mean_Time_Body_Acc_Jerk_Z_axis**: Average of the mean of the time domain component of the estimated body acceleration jerk (on the X, Y and Z axes respectively). Gravity units 'g'
- **StDev_Time_Body_Acc_Jerk_X_axis, StDev_Time_Body_Acc_Jerk_Y_axis and StDev_Time_Body_Acc_Jerk_Z_axis**: Average of the standard deviation of the time domain component of the estimated body acceleration jerk (on the X, Y and Z axes respectively). Gravity units 'g'
- **Mean_Time_Body_AngVelocity_X_axis, Mean_Time_Body_AngVelocity_Y_axis and Mean_Time_Body_AngVelocity_Z_axis**: Average of the mean of the time domain component of the angular velocity from the gyroscope (on the X, Y and Z axes respectively). Units are radians/second
- **StDev_Time_Body_AngVelocity_X_axis, StDev_Time_Body_AngVelocity_Y_axis and**
- **StDev_Time_Body_AngVelocity_Z_axis**: Average of the standard deviation of the time domain component of the angular velocity from the gyroscope (on the X, Y and Z axes respectively). Units are radians/second
- **Mean_Time_Body_AngVelocity_Jerk_X_axis, Mean_Time_Body_AngVelocity_Jerk_Y_axis and Mean_Time_Body_AngVelocity_Jerk_Z_axis**: Average of the mean of the time domain component of the angular velocity jerk (on the X, Y and Z axes respectively). Units are radians/second
- **StDev_Time_Body_AngVelocity_Jerk_X_axis, StDev_Time_Body_AngVelocity_Jerk_Y_axis and**

- **StDev_Time_Body_AngVelocity_Jerk_Z_axis**: Average of the standard deviation of the time domain component of the angular velocity jerk (on the X, Y and Z axes respectively). Units are radians/second
- **Mean_Time_Body_Acc_Magnitude**: Average of the mean of the time domain component of the total body acceleration. Gravity units 'g'
- **StDev_Time_Body_Acc_Magnitude**: Average of the standard deviation of the time domain component of the total body acceleration. Gravity units 'g'
- **Mean_Time_Gravity_Acc_Magnitude**: Average of the mean of the time domain component of the total gravity acceleration from the accelerometer. Gravity units 'g'
- **StDev_Time_Gravity_Acc_Magnitude**: Average of the standard deviation of the time domain component of the total gravity acceleration from the accelerometer. Gravity units 'g'
- **Mean_Time_Body_Acc_Jerk_Magnitude**: Average of the mean of the time domain component of the total body acceleration jerk. Gravity units 'g'
- **StDev_Time_Body_Acc_Jerk_Magnitude**: Average of the standard deviation of the time domain component of the total body acceleration jerk. Gravity units 'g'
- **Mean_Time_Body_AngVelocity_Magnitude**: Average of the mean of the time domain component of the total angular velocity. Units are radians/second
- **StDev_Time_Body_AngVelocity_Magnitude**: Average of the standard deviation of the time domain component of the total angular velocity. Units are radians/second
- **Mean_Time_Body_AngVelocity_Jerk_Magnitude**: Average of the mean of the time domain component of the total angular velocity jerk. Units are radians/second
- **StDev_Time_Body_AngVelocity_Jerk_Magnitude**: Average of the standard deviation of the time domain component of the total angular velocity jerk. Units are radians/second

The rest of the variables (43 to 68) have *similar interpretation* except that they are the ***frequency domain*** counterparts (*as opposed to the time domain variables defined above*)

[43] Mean_Freq_Body_Acc_X_axis
[44] Mean_Freq_Body_Acc_Y_axis
[45] Mean_Freq_Body_Acc_Z_axis
[46] StDev_Freq_Body_Acc_X_axis
[47] StDev_Freq_Body_Acc_Y_axis
[48] StDev_Freq_Body_Acc_Z_axis
[49] Mean_Freq_Body_Acc_Jerk_X_axis
[50] Mean_Freq_Body_Acc_Jerk_Y_axis
[51] Mean_Freq_Body_Acc_Jerk_Z_axis
[52] StDev_Freq_Body_Acc_Jerk_X_axis
[53] StDev_Freq_Body_Acc_Jerk_Y_axis
[54] StDev_Freq_Body_Acc_Jerk_Z_axis
[55] Mean_Freq_Body_AngVelocity_X_axis
[56] Mean_Freq_Body_AngVelocity_Y_axis
[57] Mean_Freq_Body_AngVelocity_Z_axis
[58] StDev_Freq_Body_AngVelocity_X_axis
[59] StDev_Freq_Body_AngVelocity_Y_axis
[60] StDev_Freq_Body_AngVelocity_Z_axis
[61] Mean_Freq_Body_Acc_Magnitude
[62] StDev_Freq_Body_Acc_Magnitude
[63] Mean_Freq_Body_Acc_Jerk_Magnitude
[64] StDev_Freq_Body_Acc_Jerk_Magnitude
[65] Mean_Freq_Body_AngVelocity_Magnitude
[66] StDev_Freq_Body_AngVelocity_Magnitude
[67] Mean_Freq_Body_Body_AngVelocity_Magnitude
[68] StDev_Freq_Body_Body_AngVelocity_Magnitude

REFERENCES

[1] *Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz*. Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. International Workshop of Ambient Assisted Living (IWAAL 2012). Vitoria-Gasteiz, Spain. Dec 2012