Code Book

for the course Project of the Getting and Cleaning Data course

The script written should return a Tidy data set of the “Human Activity Recognition Using Smartphones Data Set” provided here:

https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip

The “Human Activity Recognition Using Smartphones Data Set” was built from the recordings of 30 subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors.

Data Set Information:

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.

For more information visit:
http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

The Dataset contained a directory called “UCI HAR 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.

My Script:

======

My script that is further described in the readme file of my GitHub working directory returns a Table that looks like this:

It lists 30 different tables for each subject 1-30

Each table provides the average of a bunch of observations for each activity performed.

The Activities listed in column 1 are:
WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING

The observations during the activities in the table are the following:

[2] "tBodyAcc-mean()-X"

[3] "tBodyAcc-mean()-Y"

[4] "tBodyAcc-mean()-Z"

[5] "tGravityAcc-mean()-X"

[6]"tGravityAcc-mean()-Y"

[7] "tGravityAcc-mean()-Z"

[8] "tBodyAccJerk-mean()-X"

[9] "tBodyAccJerk-mean()-Y"

[10] "tBodyAccJerk-mean()-Z"

[11] "tBodyGyro-mean()-X"

[12] "tBodyGyro-mean()-Y"

[13] "tBodyGyro-mean()-Z"

[14] "tBodyGyroJerk-mean()-X"

[15]"tBodyGyroJerk-mean()-Y"

[16] "tBodyGyroJerk-mean()-Z"

[17] "tBodyAccMag-mean()"

[18] "tGravityAccMag-mean()"

[19] "tBodyAccJerkMag-mean()"

[20] "tBodyGyroMag-mean()"

[21] "tBodyGyroJerkMag-mean()"

[22] "fBodyAcc-mean()-X"

[23] "fBodyAcc-mean()-Y"

[24] "fBodyAcc-mean()-Z"

[25] "fBodyAcc-meanFreq()-X"

[26] "fBodyAcc-meanFreq()-Y"

[27] "fBodyAcc-meanFreq()-Z"

[28] "fBodyAccJerk-mean()-X"

[29] "fBodyAccJerk-mean()-Y"

[30] "fBodyAccJerk-mean()-Z"

[31] "fBodyAccJerk-meanFreq()-X"

[32] "fBodyAccJerk-meanFreq()-Y"

[33]"fBodyAccJerk-meanFreq()-Z"

[34]"fBodyGyro-mean()-X"

[35] "fBodyGyro-mean()-Y"

[36] "fBodyGyro-mean()-Z"

[37] "fBodyGyro-meanFreq()-X"

[38] "fBodyGyro-meanFreq()-Y"

[39] “fBodyGyro-meanFreq()-Z"

[40] "fBodyAccMag-mean()"

[41] "fBodyAccMag-meanFreq()"

[42] "fBodyBodyAccJerkMag-mean()"

[43] "fBodyBodyAccJerkMag-meanFreq()"

[44] "fBodyBodyGyroMag-mean()"

[45]"fBodyBodyGyroMag-meanFreq()”

[46] "fBodyBodyGyroJerkMag-mean()"

[47] "fBodyBodyGyroJerkMag-meanFreq()"

[48] "tBodyAcc-std()-X"

[49] "tBodyAcc-std()-Y" [50] "tBodyAcc-std()-Z"

[51] "tGravityAcc-std()-X"

[52] "tGravityAcc-std()-Y"

[53] "tGravityAcc-std()-Z"

[54] "tBodyAccJerk-std()-X"

[55] "tBodyAccJerk-std()-Y"

[56] "tBodyAccJerk-std()-Z"

[57] "tBodyGyro-std()-X"

[58] "tBodyGyro-std()-Y"

[59] "tBodyGyro-std()-Z"

[60] "tBodyGyroJerk-std()-X"

[61] "tBodyGyroJerk-std()-Y"

[62] "tBodyGyroJerk-std()-Z"

[63] "tBodyAccMag-std()"

[64]"tGravityAccMag-std()"

[65] "tBodyAccJerkMag-std()"

[66]"tBodyGyroMag-std()"

[67]"tBodyGyroJerkMag-std()"

[68] "fBodyAcc-std()-X"

[69]"fBodyAcc-std()-Y"

[70]"fBodyAcc-std()-Z"

[71] "fBodyAccJerk-std()-X"

[72] "fBodyAccJerk-std()-Y"

[73] "fBodyAccJerk-std()-Z"

[74] "fBodyGyro-std()-X"

[75] "fBodyGyro-std()-Y"

[76] "fBodyGyro-std()-Z"

[77] "fBodyAccMag-std()"

[78] "fBodyBodyAccJerkMag-std()"

[79] "fBodyBodyGyroMag-std()"

[80] "fBodyBodyGyroJerkMag-std()"

The features selected for this database come from the accelerometer and gyroscope 3-axial raw signals tAcc-XYZ and tGyro-XYZ. These time domain signals (prefix 't' to denote time) were captured at a constant rate of 50 Hz. Then they were filtered using a median filter and a 3rd order low pass Butterworth filter with a corner frequency of 20 Hz to remove noise. Similarly, the acceleration signal was then separated into body and gravity acceleration signals (tBodyAcc-XYZ and tGravityAcc-XYZ) using another low pass Butterworth filter with a corner frequency of 0.3 Hz.

Subsequently, the body linear acceleration and angular velocity were derived in time to obtain Jerk signals (tBodyAccJerk-XYZ and tBodyGyroJerk-XYZ). Also the magnitude of these three-dimensional signals were calculated using the Euclidean norm (tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag).

Finally a Fast Fourier Transform (FFT) was applied to some of these signals producing fBodyAcc-XYZ, fBodyAccJerk-XYZ, fBodyGyro-XYZ, fBodyAccJerkMag, fBodyGyroMag, fBodyGyroJerkMag. (Note the 'f' to indicate frequency domain signals).

These signals were used to estimate variables of the feature vector for each pattern:

'-XYZ' is used to denote 3-axial signals in the X, Y and Z directions.

The set of variables that were estimated from these signals are:

mean(): Mean value

std(): Standard deviation