**Getting and Cleaning Data Course Project**

Human Activity Recognition Using Smartphones Dataset [1]

**Source Experiment & Data Set**

The source data set represents data collected from the accelerometers from the Samsung Galaxy S smartphone. The data set archive can be downloaded from:

Data Set URL: <https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>

One of the most exciting areas in all of data science right now is wearable computing - see for example this article . Companies like Fitbit, Nike, and Jawbone Up are racing to develop the most advanced algorithms to attract new users.

The data is made available by Reyes-Ortiz, Anguita, Ghio, and Oneto from the Smartlab Non Linear Complex Systems Laboratory in Genoa, Italy.

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 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.

Use of this dataset in publications must be acknowledged by referencing the following publication [1].

[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

For more information please visit [2]

**Source Files used**

The source files we use to create the tidy data set from the extracted data archive are:

* 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.

**Variables**

Variables within the tidy data set are as examined as follows (taken from [1]).

The activity labels consist of WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING

The subjects who carried out the experiment are represented as a number from 1 to 30.

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).

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.

For each measurement the mean and standard deviation were calculated.

**Transformations**

Training and test data set rows were appended and then a unified data set created from the source files. Measurements were extracted for mean, standard deviation for each measurement.

Variable names for the tidy dataset are as follows:-

Output data set has 40 obs. of 69 variables:

$ SubjectId:

$ ActivityId :

$ ActivityDesc:

|  |
| --- |
| $ TimeBodyAccelerometerMeanX : |
| $ TimeBodyAccelerometerMeanY : |
| $ TimeBodyAccelerometerMeanZ : |
| $ TimeGravityAccelerometerMeanX : |
| $ TimeGravityAccelerometerMeanY : |
| $ TimeGravityAccelerometerMeanZ : |
| $ TimeBodyAccelerometerJerkMeanX : |
| $ TimeBodyAccelerometerJerkMeanY : |
| $ TimeBodyAccelerometerJerkMeanZ : |
| $ TimeBodyGyroscopeMeanX : |
| $ TimeBodyGyroscopeMeanY : |
| $ TimeBodyGyroscopeMeanZ : |
| $ TimeBodyGyroscopeJerkMeanX : |
| $ TimeBodyGyroscopeJerkMeanY : |
| $ TimeBodyGyroscopeJerkMeanZ : |
| $ TimeBodyAccelerometerMagMean : |
| $ TimeGravityAccelerometerMagMean : |
| $ TimeBodyAccelerometerJerkMagMean : |
| $ TimeBodyGyroscopeMagMean : |
| $ TimeBodyGyroscopeJerkMagMean : |
| $ FrequencyBodyAccelerometerMeanX : |
| $ FrequencyBodyAccelerometerMeanY : |
| $ FrequencyBodyAccelerometerMeanZ : |
| $ FrequencyBodyAccelerometerJerkMeanX : |
| $ FrequencyBodyAccelerometerJerkMeanY : |
| $ FrequencyBodyAccelerometerJerkMeanZ : |
| $ FrequencyBodyGyroscopeMeanX : |
| $ FrequencyBodyGyroscopeMeanY : |
| $ FrequencyBodyGyroscopeMeanZ : |
| $ FrequencyBodyAccelerometerMagMean : |
| $ FrequencyBodyAccelerometerJerkMagMean : |
| $ FrequencyBodyGyroscopeMagMean : |
| $ FrequencyBodyGyroscopeJerkMagMean : |
| $ TimeBodyAccelerometerStdevX : |
| $ TimeBodyAccelerometerStdevY : |
| $ TimeBodyAccelerometerStdevZ : |
| $ TimeGravityAccelerometerStdevX : |
| $ TimeGravityAccelerometerStdevY : |
| $ TimeGravityAccelerometerStdevZ : |
| $ TimeBodyAccelerometerJerkStdevX : |
| $ TimeBodyAccelerometerJerkStdevY : |
| $ TimeBodyAccelerometerJerkStdevZ : |
| $ TimeBodyGyroscopeStdevX : |
| $ TimeBodyGyroscopeStdevY : |
| $ TimeBodyGyroscopeStdevZ : |
| $ TimeBodyGyroscopeJerkStdevX : |
| $ TimeBodyGyroscopeJerkStdevY : |
| $ TimeBodyGyroscopeJerkStdevZ : |
| $ TimeBodyAccelerometerMagStdev : |
| $ TimeGravityAccelerometerMagStdev : |
| $ TimeBodyAccelerometerJerkMagStdev : |
| $ TimeBodyGyroscopeMagStdev : |
| $ TimeBodyGyroscopeJerkMagStdev : |
| $ FrequencyBodyAccelerometerStdevX : |
| $ FrequencyBodyAccelerometerStdevY : |
| $ FrequencyBodyAccelerometerStdevZ : |
| $ FrequencyBodyAccelerometerJerkStdevX : |
| $ FrequencyBodyAccelerometerJerkStdevY : |
| $ FrequencyBodyAccelerometerJerkStdevZ : |
| $ FrequencyBodyGyroscopeStdevX : |
| $ FrequencyBodyGyroscopeStdevY : |
| $ FrequencyBodyGyroscopeStdevZ : |
| $ FrequencyBodyAccelerometerMagStdev : |
| $ FrequencyBodyAccelerometerJerkMagStdev: |
| $ FrequencyBodyGyroscopeMagStdev : |
| $ FrequencyBodyGyroscopeJerkMagStdev : |

Labels were cleaned to make the variable names readable.

> gsub("-", "", colnames(Mergeddata2))

> gsub("\\(\\)", "", colnames(Mergeddata2))

> gsub("\\.", "", colnames(Mergeddata2))

> gsub("BodyBody", "Body", colnames(Mergeddata2))

> gsub("^t", "Time", colnames(Mergeddata2))

> gsub("^f", "Frequency", colnames(Mergeddata2))

> gsub("Acc", "Accelerometer", colnames(Mergeddata2))

> gsub("Gyro", "Gyroscope", colnames(Mergeddata2))

> gsub("mean", "Mean", colnames(Mergeddata2))

> gsub("std", "Stdev", colnames(Mergeddata2))

Results were output as an indepenent tidy data set at HumanActivityLog\_SummaryData.txt

**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<http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>

[2] <http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>