Getting and Cleaning Data

Code Book for Course Project

Study Design

This document is the code describes the variables generated by the script run\_analysis.R, as part of the course project for the Getting and Cleaning Data Coursera course.

The course project is based on the “Human Activity Recognition Using Smartphones” data set available in [this archive.](https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip) This data set formed the basis of the study done by Jorge L. Reyes-Ortiz, Davide Anguita, Alessandro Ghio, Luca Oneto at the Smartlab - Non Linear Complex Systems Laboratory in Genoa, Italy and represents a database built from the recordings of 30 subjects carrying a waist-mounted smartphone with embedded inertial sensors. Results were obtained on the dataset by exploiting a multiclass Support Vector Machine (SVM) (Anguita, et al., 2012).

A description of the full data set is included in the archive along with the raw and processed data, so it will not be reproduced here. This document concerns the data set that is the result of processing steps in the run\_analysis.R script. Specifically, the script performs the following steps:

1. Merges the training and the test sets in the UCI HAR database to create one data set.
2. Extracts only the measurements on the mean and standard deviation for each measurement.
3. Uses descriptive activity names to name the activities in the data set
4. Appropriately labels the data set with descriptive variable names.
5. Creates a second, independent tidy data set with the average of each variable for each activity and each subject.

It is in the last step that new variable are created that were not part of the Anguita et al. data set and which are described here, namely the means of the mean and standard deviation for each measurement, calculated for each of the 30 subjects and each of the six activities that the analysis detected, namely walking, walking upstairs, walking downstairs, sitting, standing, and laying down.

Code Book

The original dataset contained 561 “features” (variables). Of these, 66 variables were means or standards deviations (as determined from the occurrence of the character string “mean()” or “std()” in the feature name. These 66 variables were the ones extracted in Step 2 of the procedure and carried forward in the rest of the analysis.

To differentiate the mean variables that the run\_analysis.R script created from the original, yet still identify which variable was being measured, the script applied the following rule for naming the variables:

names(newdata) <- paste("Mean(",names(dataset),")",sep="")

Table 1 lists the variables and their meaning (all linear acceleration measurements are in standard freefall units of ‘g’ (32.174 ft/s/s) and rotational accelerations are in radians/s):

Table 1: Modified Variable Names Generated by run\_analysis.R

|  |  |
| --- | --- |
| **Variable Name** | **Mean of…** |
| Mean(tBodyAcc-mean()-X) | mean of time domain body acceleration-x |
| Mean(tBodyAcc-mean()-Y) | mean of time domain body acceleration-y |
| Mean(tBodyAcc-mean()-Z) | mean of time domain body acceleration-z |
| Mean(tGravityAcc-mean()-X) | mean of time domain gravity acceleration-x |
| Mean(tGravityAcc-mean()-Y) | mean of time domain gravity acceleration-y |
| Mean(tGravityAcc-mean()-Z) | mean of time domain gravity acceleration-z |
| Mean(tBodyAccJerk-mean()-X) | mean of time domain body jerk acceleration-x |
| Mean(tBodyAccJerk-mean()-Y) | mean of time domain body jerk acceleration-y |
| Mean(tBodyAccJerk-mean()-Z) | mean of time domain body jerk acceleration-z |
| Mean(tBodyGyro-mean()-X) | mean of time domain body rotation acc –x |
| Mean(tBodyGyro-mean()-Y) | mean of time domain body rotation acc –y |
| Mean(tBodyGyro-mean()-Z) | mean of time domain body rotation acc –z |
| Mean(tBodyGyroJerk-mean()-X) | mean of time domain body rotation jerk -x |
| Mean(tBodyGyroJerk-mean()-Y) | mean of time domain body rotation jerk –y |
| Mean(tBodyGyroJerk-mean()-Z) | mean of time domain body rotation jerk –z |
| Mean(tBodyAccMag-mean()) | mean of time domain body acc magnitude |
| Mean(tGravityAccMag-mean()) | mean of time domain gravity acc magnitude |
| Mean(tBodyAccJerkMag-mean()) | mean of time domain body acc jerk magnitude |
| Mean(tBodyGyroMag-mean()) | mean of time domain body rotation magnitude |
| Mean(tBodyGyroJerkMag-mean()) | mean of time domain body rotation jerk magnitude |
| Mean(fBodyAcc-mean()-X) | mean of freq domain body acceleration-x |
| Mean(fBodyAcc-mean()-Y) | mean of freq domain body acceleration-y |
| Mean(fBodyAcc-mean()-Z) | mean of freq domain body acceleration-z |
| Mean(fBodyAccJerk-mean()-X) | mean of freq domain body jerk acceleration-x |
| Mean(fBodyAccJerk-mean()-Y) | mean of freq domain body jerk acceleration-y |
| Mean(fBodyAccJerk-mean()-Z) | mean of freq domain body jerk acceleration-z |
| Mean(fBodyGyro-mean()-X) | mean of freq domain body rotation acc-x |
| Mean(fBodyGyro-mean()-Y) | mean of freq domain body rotation acc-y |
| Mean(fBodyGyro-mean()-Z) | mean of freq domain body rotation acc-z |
| Mean(fBodyAccMag-mean()) | mean of freq domain body accleration mag |
| Mean(fBodyBodyAccJerkMag-mean()) | mean of freq domain body jerk acc mag |
| Mean(fBodyBodyGyroMag-mean()) | mean of freq domain body rotation acc mag |
| Mean(fBodyBodyGyroJerkMag-mean()) | mean of freq domain body rotation jerk acc mag |
| Mean(tBodyAcc-std()-X) | std of time domain body acceleration-x |
| Mean(tBodyAcc-std()-Y) | std of time domain body acceleration-y |
| Mean(tBodyAcc-std()-Z) | std of time domain body acceleration-z |
| Mean(tGravityAcc-std()-X) | std of time domain gravity acceleration-x |
| Mean(tGravityAcc-std()-Y) | std of time domain gravity acceleration-y |
| Mean(tGravityAcc-std()-Z) | std of time domain gravity acceleration-z |
| Mean(tBodyAccJerk-std()-X) | std of time domain body jerk acceleration-x |
| Mean(tBodyAccJerk-std()-Y) | std of time domain body jerk acceleration-y |
| Mean(tBodyAccJerk-std()-Z) | std of time domain body jerk acceleration-z |
| Mean(tBodyGyro-std()-X) | std of time domain body rotation acc –x |
| Mean(tBodyGyro-std()-Y) | std of time domain body rotation acc –y |
| Mean(tBodyGyro-std()-Z) | std of time domain body rotation acc –z |
| Mean(tBodyGyroJerk-std()-X) | std of time domain body rotation jerk -x |
| Mean(tBodyGyroJerk-std()-Y) | std of time domain body rotation jerk –y |
| Mean(tBodyGyroJerk-std()-Z) | std of time domain body rotation jerk –z |
| Mean(tBodyAccMag-std()) | std of time domain body acc magnitude |
| Mean(tGravityAccMag-std()) | std of time domain gravity acc magnitude |
| Mean(tBodyAccJerkMag-std()) | std of time domain body acc jerk magnitude |
| Mean(tBodyGyroMag-std()) | std of time domain body rotation magnitude |
| Mean(tBodyGyroJerkMag-std()) | std of time domain body rotation jerk magnitude |
| Mean(fBodyAcc-std()-X) | std of freq domain body acceleration-x |
| Mean(fBodyAcc-std()-Y) | std of freq domain body acceleration-y |
| Mean(fBodyAcc-std()-Z) | std of freq domain body acceleration-z |
| Mean(fBodyAccJerk-std()-X) | std of freq domain body jerk acceleration-x |
| Mean(fBodyAccJerk-std()-Y) | std of freq domain body jerk acceleration-y |
| Mean(fBodyAccJerk-std()-Z) | std of freq domain body jerk acceleration-z |
| Mean(fBodyGyro-std()-X) | std of freq domain body rotation acc-x |
| Mean(fBodyGyro-std()-Y) | std of freq domain body rotation acc-y |
| Mean(fBodyGyro-std()-Z) | std of freq domain body rotation acc-z |
| Mean(fBodyAccMag-std()) | std of freq domain body accleration mag |
| Mean(fBodyBodyAccJerkMag-std()) | std of freq domain body jerk acc mag |
| Mean(fBodyBodyGyroMag-std()) | std of freq domain body rotation acc mag |
| Mean(fBodyBodyGyroJerkMag-std()) | std of freq domain body rotation jerk acc mag |

**References**

D. Anguita, A. Ghio, L. Oneto, X. Parra, and J.L. Reyes-Ortiz. Human activity recognition on smart-phones using a multiclass hardware-friendly support vector machine. InProceedings of the Interna-tional Workshop of Ambient Assited Living, 2012