

CodeBook Description

This document is a codebook that provides descriptions of the variables, the data, and all transformations and work that I performed to clean up the data.

##The Data Source

Source

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

Description of the dataset from the source

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

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.

##The data

The dataset includes the following files (some of which are currently too large to upload to this Github repository):

'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.

##Transformation Specifics- 5 functions/requirements and How Source Code run_analysis.R Meets Those Requirements

1. Merges the training and the test sets to create one data set. Source code "run_analysis.R" loads both test and train data, processes them, and merges the results into one dataset.
2. Extracts only the measurements on the mean and standard deviation for each measurement. Source code "run_analysis.R" extracts the mean and standard deviation data into one dataset with appropriate column names.
3. Uses descriptive activity names to name the activities in the data set. Source code "run_analysis.R" loads the descriptive feature and activity labels.
4. Appropriately labels the data set with descriptive variable names Source code "run_analysis.R" adds appropriately descriptive variable names to the large dataset columns (variables).
5. 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 Source code

"run_analysis.R" calculates the average for all measurement columns grouped by variables Activity and Subject and then writes the output to a local text file named "tidydata.txt"