Tidy_Data Code Book

Patrick Ball

Original Source & Script

This data set was created using data originally collected from http://archive.isc.uci.edu/ml./datasets/Human+Activity+Recognitition+Using+Smartphones. For this project, it was downloaded from https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip. The script run_analysis.R was run to tidy, subset, and aggregate the data.

Data Characteristics

The data have 180 rows and 82 columns and is 263 KB in size.

Data Variables

"subject" (numeric values 1-30) is the id of the person from whom the data row was collected.

"activity_code" (numeric values 1-6) is the activity id for the action performed while the data row was collected.

"activity_desc" (string values, 6 options) is a descriptive entry of the action performed while the data row was collected. Values can be: LAYING, SITTING, STANDING, WALKING, WALKING_DOWNSTAIRS, WALKING_UPSTAIRS

The remaining columns are considered features.

The features selected for this database come from the acceleromter and gyroscope 3-axial raw signals tAcc-XYZ and tGyro-XYZ. These time domain signals (prefix 't' to denot 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 aceleration 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).

Finally, a Fast Fourier Transform (FFT) was applied to some of these signals producing fBodyAcc-XYZ, fBodyAccJerk-XYZ, fBodyGyro-XYZ, fBodyAccJerkMag, fBodyGyroMag, gBodyGyroJerkMag. (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.

tBodyAcc-XYZ
tGravityAcc-XYZ
tBodyAccJerk-XYZ
tBodyGyro-XYZ
tBodyGyroJerk-XYZ
tBodyAccMag
tGravityAccMag
tBodyAccJerkMag
tBodyGyroMag

tBodyGyroJerkMag
fBodyAcc-XYZ
fBodyAccJerk-XYZ
fBodyGyro-XYZ
fBodyAccMag
fBodyAccJerkMag
fBodyGyroMag
fBodyGyroJerkMag

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

mean(): Mean value
std(): Standard deviation

Lastly, the mean of each feature was aggregated to each subject and each activity.