Codebook for Getting and Cleaning Data

Dataset

The data set referred to in this code-book is in the tidy_dataset.txt file of this repository.

Data

The tidy_dataset.txt file contains space separated values. The first row contains the variable names given below.

Variables

Each row contains 79 average signal measurements for a particular subject id and activity label

Identifiers

- subjectid identifies subject and ranges in [1, 30]
- label activity label, factor with 6 possible values:
  - WALKING: subject was walking
  - WALKING_UPSTAIRS: subject was walking upstairs
  - WALKING_DOWNSTAIRS: subject was walking downstairs
  - SITTING: subject was sitting
  - STANDING: subject was standing
  - LAYING: subject was laying

Measurement Averages

- All measurements are numeric values, normalised and bounded in [-1, 1]
- Variables are below, t stands for time domain, f for frequency domain
- X, Y, and Z refer to the 3 axes
- tBodyAcc.mean...X
- tBodyAcc.mean...Y
- tBodyAcc.mean...Z
- tBodyAcc.std...X
- tBodyAcc.std...Y
- tBodyAcc.std...Z
- tGravityAcc.mean...X
- tGravityAcc.mean...Y
- tGravityAcc.mean...Z
- tGravityAcc.std...X
- tGravityAcc.std...Y
- tGravityAcc.std...Z

- tBodyAccJerk.mean...X
- tBodyAccJerk.mean...Y
- tBodyAccJerk.mean...Z
- tBodyAccJerk.std...X
- tBodyAccJerk.std...Y
- tBodyAccJerk.std...Z
- tBodyGyro.mean...X
- tBodyGyro.mean...Y
- tBodyGyro.mean...Z
- tBodyGyro.std...X
- tBodyGyro.std...Y
- tBodyGyro.std...Z
- tBodyGyroJerk.mean...X
- tBodyGyroJerk.mean...Y
- tBodyGyroJerk.mean...Z
- tBodyGyroJerk.std...X
- tBodyGyroJerk.std...Y
- tBodyGyroJerk.std...Z
- tBodyAccMag.mean..
- tBodyAccMag.std..
- tGravityAccMag.mean..
- tGravityAccMag.std..
- tBodyAccJerkMag.mean..
- tBodyAccJerkMag.std..
- tBodyGyroMag.mean..
- tBodyGyroMag.std..
- tBodyGyroJerkMag.mean..
- tBodyGyroJerkMag.std..
- fBodyAcc.mean...X
- fBodyAcc.mean...Y
- fBodyAcc.mean...Z
- fBodyAcc.std...X
- fBodyAcc.std...Y
- fBodyAcc.std...Z
- fBodyAcc.meanFreq...X
- fBodyAcc.meanFreq...Y
- fBodyAcc.meanFreq...Z
- fBodyAccJerk.mean...X
- fBodyAccJerk.mean...Y
- fBodyAccJerk.mean...Z
- fBodyAccJerk.std...X
- fBodyAccJerk.std...Y
- fBodyAccJerk.std...Z
- fBodyAccJerk.meanFreq...X
- fBodyAccJerk.meanFreq...Y
- fBodyAccJerk.meanFreq...Z
- fBodyGyro.mean...X
- fBodyGyro.mean...Y
- fBodyGyro.mean...Z
- fBodyGyro.std...X
- fBodyGyro.std...Y
- fBodyGyro.std...Z
- fBodyGyro.meanFreq...X
- fBodyGyro.meanFreq...Y

- fBodyGyro.meanFreq...Z
- fBodyAccMag.mean..
- fBodyAccMag.std..
- fBodyAccMag.meanFreq..
- fBodyBodyAccJerkMag.mean..
- fBodyBodyAccJerkMag.std..
- fBodyBodyAccJerkMag.meanFreq..
- fBodyBodyGyroMag.mean..
- fBodyBodyGyroMag.std..
- fBodyBodyGyroMag.meanFreq..
- fBodyBodyGyroJerkMag.mean..
- fBodyBodyGyroJerkMag.std..
- fBodyBodyGyroJerkMag.meanFreq..

Transformations

The following transformations were applied to the source data which is from https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip :

- The training and test sets were loaded and given column names using the features text file
- The training and test sets were merged to create a finalset
- The measurements on mean and standard deviation were extracted and other columns discared
- The activity identifiers were replaced with activity labels (names)
- The final data set was created with the average of each variable for each subjectid and each label
- The finaltidy set was written into tidy_dataset.txt