Coursera – Getting and Cleaning Data – Course Project

Code Book of the output file created by the project

Output File Structure

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).

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.

'-mean()' and '-std()' is used to denote mean and std calculated values.

|  |  |  |
| --- | --- | --- |
| Name | Description | Range |
| Subject | Identifier of a volunteer whose measures appear in the row | 1…30 |
| Activity | Description of the activity whose measures were taken for the Subject | WALKING  WALKING\_UPSTAIRS  WALKING\_DOWNSTAIRS  SITTING  STANDING  LAYING |
| tBodyAcc-mean()-XYZ | a column for each of X,Y and Z |  |
| tBodyAcc-std()-XYZ | a column for each of X,Y and Z |  |
| tGravityAcc-mean()-XYZ | a column for each of X,Y and Z |  |
| tGravityAcc-std()-XYZ | a column for each of X,Y and Z |  |
| tBodyAccJerk-mean()-XYZ | a column for each of X,Y and Z |  |
| tBodyAccJerk-std()-XYZ | a column for each of X,Y and Z |  |
| tBodyGyro-mean()-XYZ | a column for each of X,Y and Z |  |
| tBodyGyro-std()-XYZ | a column for each of X,Y and Z |  |
| tBodyGyroJerk-mean()-XYZ | a column for each of X,Y and Z |  |
| tBodyGyroJerk-std()-XYZ | a column for each of X,Y and Z |  |
| tBodyAccMag-mean() |  |  |
| tBodyAccMag-std() |  |  |
| tGravityAccMag-mean() |  |  |
| tGravityAccMag-std() |  |  |
| tBodyAccJerkMag-mean() |  |  |
| tBodyAccJerkMag-std() |  |  |
| tBodyGyroMag-mean() |  |  |
| tBodyGyroMag-std() |  |  |
| tBodyGyroJerkMag-mean() |  |  |
| tBodyGyroJerkMag-std() |  |  |
| fBodyAcc-mean()-XYZ | a column for each of X,Y and Z |  |
| fBodyAcc-std()-XYZ | a column for each of X,Y and Z |  |
| fGravityAcc-mean()-XYZ | a column for each of X,Y and Z |  |
| fGravityAcc-std()-XYZ | a column for each of X,Y and Z |  |
| fBodyAccJerk-mean()-XYZ | a column for each of X,Y and Z |  |
| fBodyAccJerk-std()-XYZ | a column for each of X,Y and Z |  |
| fBodyGyro-mean()-XYZ | a column for each of X,Y and Z |  |
| fBodyGyro-std()-XYZ | a column for each of X,Y and Z |  |
| fBodyGyroJerk-mean()-XYZ | a column for each of X,Y and Z |  |
| fBodyGyroJerk-std()-XYZ | a column for each of X,Y and Z |  |
| fBodyAccMag-mean() |  |  |
| fBodyAccMag-std() |  |  |
| fGravityAccMag-mean() |  |  |
| fGravityAccMag-std() |  |  |
| fBodyAccJerkMag-mean() |  |  |
| fBodyAccJerkMag-std() |  |  |
| fBodyGyroMag-mean() |  |  |
| fBodyGyroMag-std() |  |  |
| fBodyGyroJerkMag-mean() |  |  |
| fBodyGyroJerkMag-std() |  |  |

Processing

The data represent data collected from the accelerometers from the Samsung Galaxy S smartphone. A full description is available at the site where the data was obtained:   
  
<http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>

The raw data was downloaded from <https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>

The dataset was created by merging the measures for both training and test volunteer groups to a single dataset. Only the mean and std features were included. The Activity identifier from the raw data was replaced by a descriptive string.

The average of each of the features was calculated per grouping of Subject and Activity.