Kristen Bell

Code Book

Data Science / Coursera

Getting Data – Class Project

Introduction

This summary dataset is derived from a more detailed located here: <http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>

1. VARIABLES

The variables included in this summary data set are labeled and defined as follows:

|  |  |
| --- | --- |
| Variable Name | Description |
| SubjectID | Unique subject ID (1 to 30) |
| Activity | One of 6 possible activities the subject was doing during the experiment: WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING |
| tBodyAcc-mean()-X | Mean body acceleration, along the X, Y, and Z axes, calculated from the time domain, averaged across each subject on each activity |
| tBodyAcc-mean()-Y |
| tBodyAcc-mean()-Z |
| tGravityAcc-mean()-X | Mean gravitational acceleration, along the X, Y, and Z axes, calculated from the time domain, averaged across each subject on each activity |
| tGravityAcc-mean()-Y |
| tGravityAcc-mean()-Z |
| tBodyAccJerk-mean()-X | Mean body linear acceleration - derived in time to obtain Jerk signals along the X, Y, and Z axes, averaged across each subject on each activity |
| tBodyAccJerk-mean()-Y |
| tBodyAccJerk-mean()-Z |
| tBodyGyro-mean()-X | Mean angular velocity, along the X, Y, and Z axes, calculated from the time domain, averaged across each subject on each activity |
| tBodyGyro-mean()-Y |
| tBodyGyro-mean()-Z |
| tBodyGyroJerk-mean()-X | Mean body angular velocity - derived in time to obtain Jerk signals along the X, Y, and Z axes, averaged across each subject on each activity |
| tBodyGyroJerk-mean()-Y |
| tBodyGyroJerk-mean()-Z |
| fBodyAcc-mean()-X | Mean body acceleration, along the X, Y, and Z axes, calculated from the frequency domain, averaged across each subject on each activity |
| fBodyAcc-mean()-Y |
| fBodyAcc-mean()-Z |
| fBodyAccJerk-mean()-X | Mean body linear acceleration - derived from the frequency domain to obtain Jerk signals along the X, Y, and Z axes, averaged across each subject on each activity |
| fBodyAccJerk-mean()-Y |
| fBodyAccJerk-mean()-Z |
| fBodyGyro-mean()-X | Mean angular velocity, along the X, Y, and Z axes, calculated from the frequency domain, averaged across each subject on each activity |
| fBodyGyro-mean()-Y |
| fBodyGyro-mean()-Z |

Continued next page

|  |  |
| --- | --- |
| Variable Name | Description |
| tBodyAcc-std()-X | Average standard deviation for body acceleration, along the X, Y, and Z axes, calculated from the time domain (Std deviations were averaged across each subject on each activity) |
| tBodyAcc-std()-Y |
| tBodyAcc-std()-Z |
| tGravityAcc-std()-X | Average standard deviation for gravitational acceleration, along the X, Y, and Z axes, calculated from the time domain (Std deviations were averaged across each subject on each activity) |
| tGravityAcc-std()-Y |
| tGravityAcc-std()-Z |
| tBodyAccJerk-std()-X | Average standard deviation for body linear acceleration was derived in time to obtain Jerk signals along the X, Y, and Z axes (Std deviations were averaged across each subject on each activity) |
| tBodyAccJerk-std()-Y |
| tBodyAccJerk-std()-Z |
| tBodyGyro-std()-X | Average standard deviation for angular velocity, along the X, Y, and Z axes, calculated from the time domain (Std deviations were averaged across each subject on each activity) |
| tBodyGyro-std()-Y |
| tBodyGyro-std()-Z |
| tBodyGyroJerk-std()-X | Average standard deviation for body angular velocity, derived in time to obtain Jerk signals along the X, Y, and Z axes (Std deviations were averaged across each subject on each activity) |
| tBodyGyroJerk-std()-Y |
| tBodyGyroJerk-std()-Z |
| fBodyAcc-std()-X | Average standard deviation for body acceleration, along the X, Y, and Z axes, calculated from the frequency domain (Std deviations were averaged across each subject on each activity) |
| fBodyAcc-std()-Y |
| fBodyAcc-std()-Z |
| fBodyAccJerk-std()-X | Average standard deviation for body linear acceleration, derived from the frequency domain to obtain Jerk signals along the X, Y, and Z axes (Std deviations were averaged across each subject on each activity) |
| fBodyAccJerk-std()-Y |
| fBodyAccJerk-std()-Z |
| fBodyGyro-std()-X | Average standard deviation for angular velocity, along the X, Y, and Z axes, calculated from the frequency domain (Std deviations were averaged across each subject on each activity) |
| fBodyGyro-std()-Y |
| fBodyGyro-std()-Z |
| tBodyAccMag-mean() | Apparently, the magnitude of these three-dimensional signals was calculated using the Euclidean norm (tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag).  The mean values were averaged across subjects for each activity. |
| tGravityAccMag-mean() |
| tBodyAccJerkMag-mean() |
| tBodyGyroMag-mean() |
| tBodyGyroJerkMag-mean() |
| fBodyAccMag-mean() |
| fBodyBodyAccJerkMag-mean() |
| fBodyBodyGyroMag-mean() |
| fBodyBodyGyroJerkMag-mean() |
| tBodyAccMag-std() | Apparently, the magnitude of these three-dimensional signals was calculated using the Euclidean norm (tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag).  The standard deviations were averaged across subjects for each activity. Who knows why one would want to do this, but the assignment asked us to and so we did ☺ |
| tGravityAccMag-std() |
| tBodyAccJerkMag-std() |
| tBodyGyroMag-std() |
| tBodyGyroJerkMag-std() |
| fBodyAccMag-std() |
| fBodyBodyAccJerkMag-std() |
| fBodyBodyGyroMag-std() |
| fBodyBodyGyroJerkMag-std() |

1. SUMMARY CHOICES

The script uses grep to extract the above variable names. The variables were extracted based on the occurrence of the words “mean”, “std”, and “Mean” in the variable name.

Numeric variable names were replaced with descriptive variable names using a series of cbind steps.

The training and the test sets were combined to create one data set using an rbind.

The average of each variable for each activity and each subject was calculated using the aggregate function in R.

1. STUDY DESIGN

According to the researchers at UCI, the experiments were 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, they captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments were video-recorded to label the data manually. The obtained dataset was randomly partitioned by the research team into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data. [Note – per the instructions of this assignment, the data have been recombined into a single dataset].  
  
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.

For each record in the original UCI dataset, there is:

- Triaxial acceleration from the accelerometer (total acceleration) and the estimated body acceleration.   
- Triaxial Angular velocity from the gyroscope.   
- A 561-feature vector with time and frequency domain variables.   
- Its activity label.   
- An identifier of the subject who carried out the experiment.

 Additional information about the study and the original dataset can be found here: <http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>