Getting and cleaning data course project

Author: Amlish

Date June 9, 2016

General Description of Data set

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).

Description of Abbreviations of Measurements

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

mean(): Mean value

std(): Standard deviation

mad(): Median absolute deviation

max(): Largest value in array

min(): Smallest value in array

sma(): Signal magnitude area

energy(): Energy measure. Sum of the squares divided by the number of values.

iqr(): Interquartile range

entropy(): Signal entropy

arCoeff(): Autorregresion coefficients with Burg order equal to 4

correlation(): correlation coefficient between two signals

maxInds(): index of the frequency component with largest magnitude

meanFreq(): Weighted average of the frequency components to obtain a mean frequency

skewness(): skewness of the frequency domain signal

kurtosis(): kurtosis of the frequency domain signal

bandsEnergy(): Energy of a frequency interval within the 64 bins of the FFT of each window.

angle(): Angle between to vectors.

Downloading the Data

```
if(!file.exists("./data")){dir.create("./data")}
fileUrl <- "https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip"
download.file(fileUrl,destfile="./data/Dataset.zip")
```

Unzip the file

```
unzip(zipfile="./data/Dataset.zip",exdir="./data")
```

```
path_rf <- file.path("./data" , "UCI HAR Dataset")
files<-list.files(path_rf, recursive=TRUE)
files
```

```
##  [1] "activity_labels.txt"
##  [2] "features.txt"
##  [3] "features_info.txt"
##  [4] "Hello-world/README.md"
##  [5] "README.txt"
##  [6] "test/Inertial Signals/body_acc_x_test.txt"
##  [7] "test/Inertial Signals/body_acc_y_test.txt"
##  [8] "test/Inertial Signals/body_acc_z_test.txt"
##  [9] "test/Inertial Signals/body_gyro_x_test.txt"
## [10] "test/Inertial Signals/body_gyro_y_test.txt"
## [11] "test/Inertial Signals/body_gyro_z_test.txt"
```

```
## [12] "test/Inertial Signals/total_acc_x_test.txt"
## [13] "test/Inertial Signals/total_acc_y_test.txt"
## [14] "test/Inertial Signals/total_acc_z_test.txt"
## [15] "test/subject_test.txt"
## [16] "test/X_test.txt"
## [17] "test/y_test.txt"
## [18] "train/Inertial Signals/body_acc_x_train.txt"
## [19] "train/Inertial Signals/body_acc_y_train.txt"
## [20] "train/Inertial Signals/body_acc_z_train.txt"
## [21] "train/Inertial Signals/body_gyro_x_train.txt"
## [22] "train/Inertial Signals/body_gyro_y_train.txt"
## [23] "train/Inertial Signals/body_gyro_z_train.txt"
## [24] "train/Inertial Signals/total_acc_x_train.txt"
## [25] "train/Inertial Signals/total_acc_y_train.txt"
## [26] "train/Inertial Signals/total_acc_z_train.txt"
## [27] "train/subject_train.txt"
## [28] "train/X_train.txt"
## [29] "train/y_train.txt"
```

The files that will be used to load data are listed as follows:

test/subject_test.txt

test/X_test.txt

test/y_test.txt

train/subject_train.txt

train/X_train.txt

train/y_train.txt

Reading data from the targeted files

```
dataActivityTest  <- read.table(file.path(path_rf, "test" , "Y_test.txt" ),header = FALSE)
dataActivityTrain <- read.table(file.path(path_rf, "train", "Y_train.txt"),header = FALSE)
str(dataActivityTest)
```

```
## 'data.frame':    2947 obs. of  1 variable:
##  $ V1: int  5 5 5 5 5 5 5 5 5 5 ...
```

```
str(dataActivityTrain)
```

```
## 'data.frame':    7352 obs. of  1 variable:
##  $ V1: int  5 5 5 5 5 5 5 5 5 5 ...
```

Reading the subject files

```
dataSubjectTrain <- read.table(file.path(path_rf, "train", "subject_train.txt"),header = FALSE)
dataSubjectTest  <- read.table(file.path(path_rf, "test" , "subject_test.txt"),header = FALSE)
str(dataSubjectTrain)
```

```
## 'data.frame':    7352 obs. of  1 variable:
##  $ V1: int  1 1 1 1 1 1 1 1 1 1 ...
```

```
str(dataSubjectTrain)
```

```
## 'data.frame':    7352 obs. of  1 variable:
##  $ V1: int  1 1 1 1 1 1 1 1 1 1 ...
```

Reading Featured files

```
dataFeaturesTest  <- read.table(file.path(path_rf, "test" , "X_test.txt" ),header = FALSE)
dataFeaturesTrain <- read.table(file.path(path_rf, "train", "X_train.txt"),header = FALSE)
```

Merging the training and the test sets to create one data set

```
dataSubject <- rbind(dataSubjectTrain, dataSubjectTest)
dataActivity<- rbind(dataActivityTrain, dataActivityTest)
dataFeatures<- rbind(dataFeaturesTrain, dataFeaturesTest)
```

Naming the Variables

```
names(dataSubject)<-c("subject")
names(dataActivity)<- c("activity")
dataFeaturesNames <- read.table(file.path(path_rf, "features.txt"),head=FALSE)
names(dataFeatures)<- dataFeaturesNames$V2
```

##.Merge columns to get the data frame Data for all data

```
dataCombine <- cbind(dataSubject, dataActivity)
Data <- cbind(dataFeatures, dataCombine)
```

The mean and standard deviation for each measurement

```
subdataFeaturesNames<-dataFeaturesNames$V2[grep("mean\\(\\)|std\\(\\)", dataFeaturesNames$V2)]
selectedNames<-c(as.character(subdataFeaturesNames), "subject", "activity" )
Data<-subset(Data,select=selectedNames)
```

Name the activities in the data set

```
activityLabels <- read.table(file.path(path_rf, "activity_labels.txt"),header = FALSE)
head(Data$activity,30)
```

```
##  [1] 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 4 4 4
```

labels the data set with descriptive variable names

In the former part, variables activity and subject and names of the activities have been labelled using descriptive names.In this part, Names of Feteatures will labelled using descriptive variable names.

prefix t is replaced by time

Acc is replaced by Accelerometer

Gyro is replaced by Gyroscope

prefix f is replaced by frequency

Mag is replaced by Magnitude

BodyBody is replaced by Body

```
names(Data)<-gsub("^t", "time", names(Data))
names(Data)<-gsub("^f", "frequency", names(Data))
names(Data)<-gsub("Acc", "Accelerometer", names(Data))
names(Data)<-gsub("Gyro", "Gyroscope", names(Data))
names(Data)<-gsub("Mag", "Magnitude", names(Data))
names(Data)<-gsub("BodyBody", "Body", names(Data))
```

```
(names(Data))
```

```
##  [1] "timeBodyAccelerometer-mean()-X"
##  [2] "timeBodyAccelerometer-mean()-Y"
##  [3] "timeBodyAccelerometer-mean()-Z"
##  [4] "timeBodyAccelerometer-std()-X"
##  [5] "timeBodyAccelerometer-std()-Y"
##  [6] "timeBodyAccelerometer-std()-Z"
##  [7] "timeGravityAccelerometer-mean()-X"
##  [8] "timeGravityAccelerometer-mean()-Y"
##  [9] "timeGravityAccelerometer-mean()-Z"
## [10] "timeGravityAccelerometer-std()-X"
## [11] "timeGravityAccelerometer-std()-Y"
## [12] "timeGravityAccelerometer-std()-Z"
## [13] "timeBodyAccelerometerJerk-mean()-X"
## [14] "timeBodyAccelerometerJerk-mean()-Y"
## [15] "timeBodyAccelerometerJerk-mean()-Z"
## [16] "timeBodyAccelerometerJerk-std()-X"
## [17] "timeBodyAccelerometerJerk-std()-Y"
## [18] "timeBodyAccelerometerJerk-std()-Z"
## [19] "timeBodyGyroscope-mean()-X"
## [20] "timeBodyGyroscope-mean()-Y"
## [21] "timeBodyGyroscope-mean()-Z"
## [22] "timeBodyGyroscope-std()-X"
## [23] "timeBodyGyroscope-std()-Y"
## [24] "timeBodyGyroscope-std()-Z"
## [25] "timeBodyGyroscopeJerk-mean()-X"
## [26] "timeBodyGyroscopeJerk-mean()-Y"
## [27] "timeBodyGyroscopeJerk-mean()-Z"
## [28] "timeBodyGyroscopeJerk-std()-X"
## [29] "timeBodyGyroscopeJerk-std()-Y"
## [30] "timeBodyGyroscopeJerk-std()-Z"
```

```
## [31] "timeBodyAccelerometerMagnitude-mean()"
## [32] "timeBodyAccelerometerMagnitude-std()"
## [33] "timeGravityAccelerometerMagnitude-mean()"
## [34] "timeGravityAccelerometerMagnitude-std()"
## [35] "timeBodyAccelerometerJerkMagnitude-mean()"
## [36] "timeBodyAccelerometerJerkMagnitude-std()"
## [37] "timeBodyGyroscopeMagnitude-mean()"
## [38] "timeBodyGyroscopeMagnitude-std()"
## [39] "timeBodyGyroscopeJerkMagnitude-mean()"
## [40] "timeBodyGyroscopeJerkMagnitude-std()"
## [41] "frequencyBodyAccelerometer-mean()-X"
## [42] "frequencyBodyAccelerometer-mean()-Y"
## [43] "frequencyBodyAccelerometer-mean()-Z"
## [44] "frequencyBodyAccelerometer-std()-X"
## [45] "frequencyBodyAccelerometer-std()-Y"
## [46] "frequencyBodyAccelerometer-std()-Z"
## [47] "frequencyBodyAccelerometerJerk-mean()-X"
## [48] "frequencyBodyAccelerometerJerk-mean()-Y"
## [49] "frequencyBodyAccelerometerJerk-mean()-Z"
## [50] "frequencyBodyAccelerometerJerk-std()-X"
## [51] "frequencyBodyAccelerometerJerk-std()-Y"
## [52] "frequencyBodyAccelerometerJerk-std()-Z"
## [53] "frequencyBodyGyroscope-mean()-X"
## [54] "frequencyBodyGyroscope-mean()-Y"
## [55] "frequencyBodyGyroscope-mean()-Z"
## [56] "frequencyBodyGyroscope-std()-X"
## [57] "frequencyBodyGyroscope-std()-Y"
## [58] "frequencyBodyGyroscope-std()-Z"
## [59] "frequencyBodyAccelerometerMagnitude-mean()"
## [60] "frequencyBodyAccelerometerMagnitude-std()"
## [61] "frequencyBodyAccelerometerJerkMagnitude-mean()"
## [62] "frequencyBodyAccelerometerJerkMagnitude-std()"
## [63] "frequencyBodyGyroscopeMagnitude-mean()"
## [64] "frequencyBodyGyroscopeMagnitude-std()"
## [65] "frequencyBodyGyroscopeJerkMagnitude-mean()"
## [66] "frequencyBodyGyroscopeJerkMagnitude-std()"
## [67] "subject"
## [68] "activity"
```

Independent tidy data set and ouput its out put

```
library(plyr);
```

```
## Warning: package 'plyr' was built under R version 3.2.5
```

```
Data2<-aggregate(. ~subject + activity, Data, mean)
Data2<-Data2[order(Data2$subject,Data2$activity),]
write.table(Data2, file = "tidydata.txt",row.name=FALSE)
```