Activity Data CodeBook

*Rajat Kumar*

*vimalrajat@gmail.com*

Introduction

This is the CodeBook for the Getting and Cleaning Data Assignment from Week 4.

Raw Data

The raw data for this project came from the UCI Machine Learning Repository. It was

collected from the accelerometers from the Samsung Galaxy S smartphone.

The data can be downloaded here

(https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset

The full description of the experiments can be read here

(http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones).

Description

The experiments have been 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, we captured 3-axial linear acceleration and 3-axial

angular velocity at a constant rate of 50Hz. The experiments have been videorecorded

to label the data manually. The obtained dataset has been randomly

partitioned into two sets, where 70% of the volunteers was selected for generating

the training data and 30% the test data.

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 lowpass

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.

Data Files

The raw dataset included the following files:

‘README.txt’

‘features\_info.txt’: Shows information about the variables used on the feature

vector.

‘features.txt’: List of all features.

‘activity\_labels.txt’: Links the class labels with their activity name.

‘train/X\_train.txt’: Training set.

‘train/y\_train.txt’: Training labels.

‘test/X\_test.txt’: Test set.

‘test/y\_test.txt’: Test labels.

The following files are available for the train and test data. Their descriptions are

equivalent.

‘train/subject\_train.txt’: Each row identifies the subject who performed the

activity for each window sample. Its range is from 1 to 30.

‘train/Inertial Signals/total\_acc\_x\_train.txt’: The acceleration signal from the

smartphone accelerometer X axis in standard gravity units ‘g’. Every row shows

a 128 element vector. The same description applies for the

‘total\_acc\_x\_train.txt’ and ‘total\_acc\_z\_train.txt’ files for the Y and Z axis.

‘train/Inertial Signals/body\_acc\_x\_train.txt’: The body acceleration signal

obtained by subtracting the gravity from the total acceleration.

‘train/Inertial Signals/body\_gyro\_x\_train.txt’: The angular velocity vector

measured by the gyroscope for each window sample. The units are

radians/second.

Tidy Data

The Final dataset for this project consists of 1 File - “tidydata.txt”.

Description

Each row represents 1 Subject and 1 Activity. The raw data included multiple

measurement observations for each Subject and Activity, so these were averaged

together for the final tidy dataset.

Identifiers

2 columns in the dataset serve as Identifiers:

**SubjectID** - The ID for the subject participating in the experiment

Ranges from 1 to 30

**Activity** - The type of activity performed by the subject for which the

measurements were taken

WALKING

WALKING\_UPSTAIRS

WALKING\_DOWNSTAIRS

SITTING

STANDING

LAYING

Measurements

The remaining columns are measurements taken by the smartphone for each Subject

and Activity

TimeBodyAccelerometerMeanX

TimeBodyAccelerometerMeanY

TimeBodyAccelerometerMeanZ

TimeGravityAccelerometerMeanX

TimeGravityAccelerometerMeanY

TimeGravityAccelerometerMeanZ

TimeBodyAccelerometerJerkMeanX

TimeBodyAccelerometerJerkMeanY

TimeBodyAccelerometerJerkMeanZ

TimeBodyGyroscopeMeanX

TimeBodyGyroscopeMeanY

TimeBodyGyroscopeMeanZ

TimeBodyGyroscopeJerkMeanX

TimeBodyGyroscopeJerkMeanY

TimeBodyGyroscopeJerkMeanZ

TimeBodyAccelerometerMagnitudeMean

TimeGravityAccelerometerMagnitudeMean

TimeBodyAccelerometerJerkMagnitudeMean

TimeBodyGyroscopeMagnitudeMean

TimeBodyGyroscopeJerkMagnitudeMean

FrequencyBodyAccelerometerMeanX

FrequencyBodyAccelerometerMeanY

FrequencyBodyAccelerometerMeanZ

FrequencyBodyAccelerometerJerkMeanX

FrequencyBodyAccelerometerJerkMeanY

FrequencyBodyAccelerometerJerkMeanZ

FrequencyBodyGyroscopeMeanX

FrequencyBodyGyroscopeMeanY

FrequencyBodyGyroscopeMeanZ

FrequencyBodyAccelerometerMagnitudeMean

FrequencyBodyAccelerometerJerkMagnitudeMean

FrequencyBodyGyroscopeMagnitudeMean

FrequencyBodyGyroscopeJerkMagnitudeMean

TimeBodyAccelerometerStdX

TimeBodyAccelerometerStdY

TimeBodyAccelerometerStdZ

TimeGravityAccelerometerStdX

TimeGravityAccelerometerStdY

TimeGravityAccelerometerStdZ

TimeBodyAccelerometerJerkStdX

TimeBodyAccelerometerJerkStdY

TimeBodyAccelerometerJerkStdZ

TimeBodyGyroscopeStdX

TimeBodyGyroscopeStdY

TimeBodyGyroscopeStdZ

TimeBodyGyroscopeJerkStdX

TimeBodyGyroscopeJerkStdY

TimeBodyGyroscopeJerkStdZ

TimeBodyAccelerometerMagnitudeStd

TimeGravityAccelerometerMagnitudeStd

TimeBodyAccelerometerJerkMagnitudeStd

TimeBodyGyroscopeMagnitudeStd

TimeBodyGyroscopeJerkMagnitudeStd

FrequencyBodyAccelerometerStdX

FrequencyBodyAccelerometerStdY

FrequencyBodyAccelerometerStdZ

FrequencyBodyAccelerometerJerkStdX

FrequencyBodyAccelerometerJerkStdY

FrequencyBodyAccelerometerJerkStdZ

FrequencyBodyGyroscopeStdX

FrequencyBodyGyroscopeStdY

FrequencyBodyGyroscopeStdZ

FrequencyBodyAccelerometerMagnitudeStd

FrequencyBodyAccelerometerJerkMagnitudeStd

FrequencyBodyGyroscopeMagnitudeStd

FrequencyBodyGyroscopeJerkMagnitudeStd

Transformation

To get from the Raw Data to the Tidy Data the following transformations were made.

These transformations can be executed with the script ‘run\_analysis.R’.

1. Merged the training and test sets to create one data set.

2. Extracted only the measurements on the mean and standard deviation for each

measurement.

3. Used descriptive activity names to name the activities in the data set.

4. Appropriately labeled the data set with descriptive variable names.

5. From the data set in step 4, created a second, independent tidy data set with

the average of each variable for each activity and each subject