**Study Design**

The study is about human activity recognition using smartphones.

The experiments have been carried out with a group of 30 volunteers. Each person performed six activities wearing a Samsung Galaxy S II smartphone on the waist: WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING and LAYING.

Using the smartphone’s embedded accelerometer and gyroscope the 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz were captured. 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.

A full description of the experiments is available at the site where the data was obtained:

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

**Dataset Information**

The dataset was downloaded from UCI repository and includes the following files:

* README.txt
* features\_info.txt: 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.
* train/subject\_train.txt: Subjects who performed the activity for each training window sample, ranging from 1 to 30.
* train/subject\_test.txt: Subjects who performed the activity for each testing window sample, ranging 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.

Notes:

* Features are normalized and bounded within [-1,1].
* Each feature vector is a row on the text file.
* The units used for the accelerations (total and body) are 'g's (gravity of earth -> 9.80665 m/seg2).
* The gyroscope units are rad/seg.

**Code book**

The generated tidy dataset avg.txt is described in the table below.

| **Variable Information** | | | |
| --- | --- | --- | --- |
| Variable | Position | Measurement Level | Measurement Unit |
| Subject | 1 | Categorical |  |
| Activity | 2 | Categorical |  |
| avg\_tBodyAcc-mean()-X | 3 | Continuous | Gravity unit 'g' |
| avg\_tBodyAcc-mean()-Y | 4 | Continuous | Gravity unit 'g' |
| avg\_tBodyAcc-mean()-Z | 5 | Continuous | Gravity unit 'g' |
| avg\_tBodyAcc-std()-X | 6 | Continuous | Gravity unit 'g' |
| avg\_tBodyAcc-std()-Y | 7 | Continuous | Gravity unit 'g' |
| avg\_tBodyAcc-std()-Z | 8 | Continuous | Gravity unit 'g' |
| avg\_tGravityAcc-mean()-X | 9 | Continuous | Gravity unit 'g' |
| avg\_tGravityAcc-mean()-Y | 10 | Continuous | Gravity unit 'g' |
| avg\_tGravityAcc-mean()-Z | 11 | Continuous | Gravity unit 'g' |
| avg\_tGravityAcc-std()-X | 12 | Continuous | Gravity unit 'g' |
| avg\_tGravityAcc-std()-Y | 13 | Continuous | Gravity unit 'g' |
| avg\_tGravityAcc-std()-Z | 14 | Continuous | Gravity unit 'g' |
| avg\_tBodyAccJerk-mean()-X | 15 | Continuous | Gravity unit 'g' |
| avg\_tBodyAccJerk-mean()-Y | 16 | Continuous | Gravity unit 'g' |
| avg\_tBodyAccJerk-mean()-Z | 17 | Continuous | Gravity unit 'g' |
| avg\_tBodyAccJerk-std()-X | 18 | Continuous | Gravity unit 'g' |
| avg\_tBodyAccJerk-std()-Y | 19 | Continuous | Gravity unit 'g' |
| avg\_tBodyAccJerk-std()-Z | 20 | Continuous | Gravity unit 'g' |
| avg\_tBodyGyro-mean()-X | 21 | Continuous | rad/seg |
| avg\_tBodyGyro-mean()-Y | 22 | Continuous | rad/seg |
| avg\_tBodyGyro-mean()-Z | 23 | Continuous | rad/seg |
| avg\_tBodyGyro-std()-X | 24 | Continuous | rad/seg |
| avg\_tBodyGyro-std()-Y | 25 | Continuous | rad/seg |
| avg\_tBodyGyro-std()-Z | 26 | Continuous | rad/seg |
| avg\_tBodyGyroJerk-mean()-X | 27 | Continuous | rad/seg |
| avg\_tBodyGyroJerk-mean()-Y | 28 | Continuous | rad/seg |
| avg\_tBodyGyroJerk-mean()-Z | 29 | Continuous | rad/seg |
| avg\_tBodyGyroJerk-std()-X | 30 | Continuous | rad/seg |
| avg\_tBodyGyroJerk-std()-Y | 31 | Continuous | rad/seg |
| avg\_tBodyGyroJerk-std()-Z | 32 | Continuous | rad/seg |
| avg\_tBodyAccMag-mean() | 33 | Continuous | Gravity unit 'g' |
| avg\_tBodyAccMag-std() | 34 | Continuous | Gravity unit 'g' |
| avg\_tGravityAccMag-mean() | 35 | Continuous | Gravity unit 'g' |
| avg\_tGravityAccMag-std() | 36 | Continuous | Gravity unit 'g' |
| avg\_tBodyAccJerkMag-mean() | 37 | Continuous | Gravity unit 'g' |
| avg\_tBodyAccJerkMag-std() | 38 | Continuous | Gravity unit 'g' |
| avg\_tBodyGyroMag-mean() | 39 | Continuous | rad/seg |
| avg\_tBodyGyroMag-std() | 40 | Continuous | rad/seg |
| avg\_tBodyGyroJerkMag-mean() | 41 | Continuous | rad/seg |
| avg\_tBodyGyroJerkMag-std() | 42 | Continuous | rad/seg |
| avg\_fBodyAcc-mean()-X | 43 | Continuous | Gravity unit 'g' |
| avg\_fBodyAcc-mean()-Y | 44 | Continuous | Gravity unit 'g' |
| avg\_fBodyAcc-mean()-Z | 45 | Continuous | Gravity unit 'g' |
| avg\_fBodyAcc-std()-X | 46 | Continuous | Gravity unit 'g' |
| avg\_fBodyAcc-std()-Y | 47 | Continuous | Gravity unit 'g' |
| avg\_fBodyAcc-std()-Z | 48 | Continuous | Gravity unit 'g' |
| avg\_fBodyAcc-meanFreq()-X | 49 | Continuous | Gravity unit 'g' |
| avg\_fBodyAcc-meanFreq()-Y | 50 | Continuous | Gravity unit 'g' |
| avg\_fBodyAcc-meanFreq()-Z | 51 | Continuous | Gravity unit 'g' |
| avg\_fBodyAccJerk-mean()-X | 52 | Continuous | Gravity unit 'g' |
| avg\_fBodyAccJerk-mean()-Y | 53 | Continuous | Gravity unit 'g' |
| avg\_fBodyAccJerk-mean()-Z | 54 | Continuous | Gravity unit 'g' |
| avg\_fBodyAccJerk-std()-X | 55 | Continuous | Gravity unit 'g' |
| avg\_fBodyAccJerk-std()-Y | 56 | Continuous | Gravity unit 'g' |
| avg\_fBodyAccJerk-std()-Z | 57 | Continuous | Gravity unit 'g' |
| avg\_fBodyAccJerk-meanFreq()-X | 58 | Continuous | Gravity unit 'g' |
| avg\_fBodyAccJerk-meanFreq()-Y | 59 | Continuous | Gravity unit 'g' |
| avg\_fBodyAccJerk-meanFreq()-Z | 60 | Continuous | Gravity unit 'g' |
| avg\_fBodyGyro-mean()-X | 61 | Continuous | rad/seg |
| avg\_fBodyGyro-mean()-Y | 62 | Continuous | rad/seg |
| avg\_fBodyGyro-mean()-Z | 63 | Continuous | rad/seg |
| avg\_fBodyGyro-std()-X | 64 | Continuous | rad/seg |
| avg\_fBodyGyro-std()-Y | 65 | Continuous | rad/seg |
| avg\_fBodyGyro-std()-Z | 66 | Continuous | rad/seg |
| avg\_fBodyGyro-meanFreq()-X | 67 | Continuous | rad/seg |
| avg\_fBodyGyro-meanFreq()-Y | 68 | Continuous | rad/seg |
| avg\_fBodyGyro-meanFreq()-Z | 69 | Continuous | rad/seg |
| avg\_fBodyAccMag-mean() | 70 | Continuous | Gravity unit 'g' |
| avg\_fBodyAccMag-std() | 71 | Continuous | Gravity unit 'g' |
| avg\_fBodyAccMag-meanFreq() | 72 | Continuous | Gravity unit 'g' |
| avg\_fBodyBodyAccJerkMag-mean() | 73 | Continuous | Gravity unit 'g' |
| avg\_fBodyBodyAccJerkMag-std() | 74 | Continuous | Gravity unit 'g' |
| avg\_fBodyBodyAccJerkMag-meanFreq() | 75 | Continuous | Gravity unit 'g' |
| avg\_fBodyBodyGyroMag-mean() | 76 | Continuous | rad/seg |
| avg\_fBodyBodyGyroMag-std() | 77 | Continuous | rad/seg |
| avg\_fBodyBodyGyroMag-meanFreq() | 78 | Continuous | rad/seg |
| avg\_fBodyBodyGyroJerkMag-mean() | 79 | Continuous | rad/seg |
| avg\_fBodyBodyGyroJerkMag-std() | 80 | Continuous | rad/seg |
| avg\_fBodyBodyGyroJerkMag-meanFreq() | 81 | Continuous | rad/seg |

| **Variable Values** | |
| --- | --- |
| Value | Label |
| Activity 1 | WALKING |
| 2 | WALKING\_UPSTAIRS |
| 3 | WALKING DOWNSTAIRS |
| 4 | SITTING |
| 5 | STANDING |
| 6 | LAYING |

**Instruction List / Script**

**run\_analysis.R**

1. Step 1 - Read the training and testing datasets and merge them to form the complete set, named completeSet, with 561 variables and 10299 rows;
2. Step 2 - Read the features names, select the names with measurements on the mean and standard deviation, extract the columns of interest from completeSet and name the variable columns. The new set, called selectSet, contains 81 variables.
3. Step 3 - Read the activities perfomed by the subjects from the respective training and testing datasets and merge these activity data into a single set named activitySet.
4. Step 4 - Read activity labels and replace the activity column values in activitySet with the descriptive activity names.
5. Step 5 - Read the subjects who performed the activities from the respective training and testing datasets and merge these subject data into a single set named subjectSet;
6. Step 6 - Create a tidy dataset with the columns Subject, Activity and Average of each feature for each activity and each subject.
7. Step 7 - Write the tidy dataset into the avgSet.txt file.

The software was used in Mac system more then once to confirm it gave the same results.