*Benjamin Menashe*
Getting and Cleaning Data course on Coursera – final assignment

# Programming Assignment 3 – Code Book

## Raw Data Information:

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 video-recorded 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.

> *Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. A Public Domain Dataset for Human Activity Recognition Using Smartphones. 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.*

## Transformations:

The raw data, which was separated to a training set and a test set, was merged into a single dataset. Only the variables which measured the mean or standard deviation of their corresponding measurement were extracted. Label names were fixed for readability.

## Final Dataset information:

Following the transformations, 2 datasets with identical variable names were created:

alldata.txt

This dataset contains the complete clean data for all the participants for all their runs.

meandata.txt

This dataset contains the mean of each variable, per participant per activity.

## Variable information:

subjectid – a unique number between 1-30 identifying a unique subject.

activity – a factor with 6 levels, each corresponding to an activity:

1. WALKING
2. WALKING_UPSTAIRS
3. WALKING_DOWNSTAIRS
4. SITTING
5. STANDING
6. LAYING

*Benjamin Menashe*
Getting and Cleaning Data course on Coursera – final assignment

Other variables:

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)

1. tBodyAcc-XYZ
2. tGravityAcc-XYZ
3. tBodyAccJerk-XYZ
4. tBodyGyro-XYZ
5. tBodyGyroJerk-XYZ
6. tBodyAccMag
7. tGravityAccMag
8. tBodyAccJerkMag
9. tBodyGyroMag
10. tBodyGyroJerkMag
11. fBodyAcc-XYZ
12. fBodyAccJerk-XYZ
13. fBodyGyro-XYZ
14. fBodyAccMag
15. fBodyAccJerkMag
16. fBodyGyroMag
17. fBodyGyroJerkMag

The set of variables that were estimated from these signals, per each direction are:

1. mean(): Mean value
2. std(): Standard deviation

Overall, there are 8 measurements that have X, Y, and Z directions and another 9 measurements, and the variables measures the mean and standard deviation for each measurement, for a total of (8*3+9)*2 = 66 variables.