Code Book

Bruce L

May 17, 2015

1. The purpose of this code book is to help understand the various codes used in the completion of the project for the Coursera Data Cleaning Course. The completed project results in a comma separated data file called: tidyDataSet.txt.

The R script "run_analysis.R"" reads in the original training and test data files, and then makes use of only some of the columns of data having to do with means and standard deviations. The complete list of variables of each feature vector is available in 'features.txt' which is included in the repo.

1.1 The specific column names for this data set are:

- 1 subject
- 2 activity
- 3 tBodyAcc-mean()-X
- 4 tBodyAcc-mean()-Y
- 5 tBodyAcc-mean()-Z
- 6 tGravityAcc-mean()-X
- 7 tGravityAcc-mean()-Y
- 8 tGravityAcc-mean()-Z
- 9 tBodyAccJerk-mean()-X
- 10 tBodyAccJerk-mean()-Y
- 11 tBodyAccJerk-mean()-Z
- 12 tBodyGyro-mean()-X
- 13 tBodyGyro-mean()-Y
- 14 tBodyGyro-mean()-Z
- 15 tBodyGyroJerk-mean()-X
- 16 tBodyGyroJerk-mean()-Y
- 17 tBodyGyroJerk-mean()-Z
- 18 tBodyAccMag-mean()
- 19 tGravityAccMag-mean()
- 20 tBodyAccJerkMag-mean()
- 21 tBodyGyroMag-mean()
- 22 tBodyGyroJerkMag-mean()
- 23 fBodyAcc-mean()-X
- 24 fBodyAcc-mean()-Y
- 25 fBodyAcc-mean()-Z
- 26 fBodyAcc-meanFreq()-X
- 27 fBodyAcc-meanFreq()-Y
- 28 fBodyAcc-meanFreq()-Z
- 29 fBodyAccJerk-mean()-X
- 30 fBodyAccJerk-mean()-Y
- 31 fBodyAccJerk-mean()-Z
- 32 fBodyAccJerk-meanFreq()-X
- 33 fBodyAccJerk-meanFreq()-Y
- 34 fBodyAccJerk-meanFreq()-Z
- 35 fBodyGyro-mean()-X

- 36 fBodyGyro-mean()-Y
- 37 fBodyGyro-mean()-Z
- 38 fBodyGyro-meanFreq()-X
- 39 fBodyGyro-meanFreq()-Y
- 40 fBodyGyro-meanFreq()-Z
- 41 fBodyAccMag-mean()
- 42 fBodyAccMag-meanFreq()
- 43 fBodyBodyAccJerkMag-mean()
- 44 fBodyBodyAccJerkMag-meanFreq()
- 45 fBodyBodyGyroMag-mean()
- 46 fBodyBodyGyroMag-meanFreq()
- 47 fBodyBodyGyroJerkMag-mean()
- 48 fBodyBodyGyroJerkMag-meanFreq()
- 49 angle(tBodyAccMean,gravity)
- 50 angle(tBodyAccJerkMean),gravityMean)
- 51 angle(tBodyGyroMean,gravityMean)
- 52 angle(tBodyGyroJerkMean,gravityMean)
- 53 angle(X,gravityMean)
- 54 angle(Y,gravityMean)
- 55 angle(Z,gravityMean)
- 56 tBodyAcc-std()-X
- 57 tBodyAcc-std()-Y
- 58 tBodyAcc-std()-Z
- 59 tGravityAcc-std()-X
- 60 tGravityAcc-std()-Y
- 61 tGravityAcc-std()-Z
- 62 tBodyAccJerk-std()-X
- 63 tBodyAccJerk-std()-Y
- 64 tBodyAccJerk-std()-Z
- 65 tBodyGyro-std()-X
- 66 tBodyGyro-std()-Y
- 67 tBodyGyro-std()-Z
- 68 tBodyGyroJerk-std()-X
- 69 tBodyGyroJerk-std()-Y
- 70 tBodyGyroJerk-std()-Z
- 71 tBodyAccMag-std()
- 72 tGravityAccMag-std()
- 73 tBodyAccJerkMag-std()
- 74 tBodyGyroMag-std()
- 75 tBodyGyroJerkMag-std()
- 76 fBodyAcc-std()-X
- 77 fBodyAcc-std()-Y
- 78 fBodyAcc-std()-Z
- 79 fBodyAccJerk-std()-X
- 80 fBodyAccJerk-std()-Y
- 81 fBodyAccJerk-std()-Z
- 82 fBodyGyro-std()-X
- 83 fBodyGyro-std()-Y
- 84 fBodyGyro-std()-Z
- 85 fBodyAccMag-std()

- 86 fBodyBodyAccJerkMag-std()
- 87 fBodyBodyGyroMag-std()
- 88 fBodyBodyGyroJerkMag-std()

2.0 Understanding the meaning of the column names.

The data linked to the course website represent data collected from the accelerometers from the Samsung Galaxy S smartphone. A full description is available at the site where the data was obtained:

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

Here are the data for the project:

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

Quoting from the features_info.txt available as part of the dataset:

Feature Selection

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.

- 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

The set of variables that were estimated from these signals are:

- 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.

Additional vectors obtained by averaging the signals in a signal window sample. These are used on the angle() variable:

- gravityMean
- tBodyAccMean
- tBodyAccJerkMean
- tBodyGyroMean
- tBodyGyroJerkMean

End of file