CODEBOOK: Getting and Cleaning Data Course Project

This codebook describes the output data from *run_analysis.R* placed within *tidy.txt*.

Comprehensive Notes

A more comprehensive description of the data set can be found in the Human Activity Recognition Using Smartphones Dataset, inside *features_info.txt*. Materials used for this project was retrieved from this repository.

Algorithm Overview

1. Load the essential libraries dplyr and reshape2.
2. For each of the training and test data sets, load the corresponding isolated data into R for processing.
   - Subject Data *(subject_(test|train).txt)*
   - Activity Performed *(y_(test|train).txt)*
   - Features Measured *(X_(test|train).txt)*
3. For each group of isolated data, merge their respective training and data set. Print out the first data set as instructed into *merged.txt*.
4. Load the features measured into R *(features.txt)*.
5. Retrieve and rename(format) features of interest (mean and std).
6. Bind all three sub data sets and label the reconstructed table using the renamed column names.
7. Begin melting and casting the table based on two factors *subjectId* and *activity*.
8. Load the activity labels in *activity_labels.txt* and mutate the column in the reconstructed table for proper labeling.
9. Print the casted table to *tidy.txt* as instructed.

Discussed and Referenced Features

This section is based in the provided Human Activity Recognition Using Smartphones Dataset.

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

Data Set Columns

This section discusses the two categories of data produced by the analysis. The first two columns are the identifiers and the succeeding columns are based off the original data set and processed as per the instructions.

Key Columns and Identifiers

The following columns uniquely identify summarized information used to group the data. * subjectId - A unique ID of a volunteer in the Human Activity Recognition Using Smartphones Experiment. The values range from 1-30, representing the 30 volunteers who participated in the experiment. * activity - A specific activity executed by a volunteer and measured during the experiment. The column *activity* can accomodate any of the values below, based on the Human Activity Recognition Using Smartphones Dataset provided with the project instructions in *activity_labels.txt*. * WALKING * WALKING_UPSTAIRS * WALKING_DOWNSTAIRS * SITTING * STANDING * LAYING

Measurements on Features

The following variables have been summarized using *mean()* in R and have been melted using the

two variables above.

- tBodyAccMeanX
- tBodyAccMeanY
- tBodyAccMeanZ
- tBodyAccStdX
- tBodyAccStdY
- tBodyAccStdZ
- tGravityAccMeanX
- tGravityAccMeanY
- tGravityAccMeanZ
- tGravityAccStdX
- tGravityAccStdY
- tGravityAccStdZ
- tBodyAccJerkMeanX
- tBodyAccJerkMeanY
- tBodyAccJerkMeanZ
- tBodyAccJerkStdX
- tBodyAccJerkStdY
- tBodyAccJerkStdZ
- tBodyGyroMeanX
- tBodyGyroMeanY
- tBodyGyroMeanZ
- tBodyGyroStdX
- tBodyGyroStdY
- tBodyGyroStdZ
- tBodyGyroJerkMeanX
- tBodyGyroJerkMeanY
- tBodyGyroJerkMeanZ
- tBodyGyroJerkStdX
- tBodyGyroJerkStdY
- tBodyGyroJerkStdZ
- tBodyAccMagMean
- tBodyAccMagStd
- tGravityAccMagMean
- tGravityAccMagStd
- tBodyAccJerkMagMean
- tBodyAccJerkMagStd
- tBodyGyroMagMean
- tBodyGyroMagStd
- tBodyGyroJerkMagMean
- tBodyGyroJerkMagStd
- fBodyAccMeanX

- fBodyAccMeanY
- fBodyAccMeanZ
- fBodyAccStdX
- fBodyAccStdY
- fBodyAccStdZ
- fBodyAccMeanFreqX
- fBodyAccMeanFreqY
- fBodyAccMeanFreqZ
- fBodyAccJerkMeanX
- fBodyAccJerkMeanY
- fBodyAccJerkMeanZ
- fBodyAccJerkStdX
- fBodyAccJerkStdY
- fBodyAccJerkStdZ
- fBodyAccJerkMeanFreqX
- fBodyAccJerkMeanFreqY
- fBodyAccJerkMeanFreqZ
- fBodyGyroMeanX
- fBodyGyroMeanY
- fBodyGyroMeanZ
- fBodyGyroStdX
- fBodyGyroStdY
- fBodyGyroStdZ
- fBodyGyroMeanFreqX
- fBodyGyroMeanFreqY
- fBodyGyroMeanFreqZ
- fBodyAccMagMean
- fBodyAccMagStd
- fBodyAccMagMeanFreq
- fBodyBodyAccJerkMagMean
- fBodyBodyAccJerkMagStd
- fBodyBodyAccJerkMagMeanFreq
- fBodyBodyGyroMagMean
- fBodyBodyGyroMagStd
- fBodyBodyGyroMagMeanFreq
- fBodyBodyGyroJerkMagMean
- fBodyBodyGyroJerkMagStd
- fBodyBodyGyroJerkMagMeanFreq