Codebook

Ioannis

30 May 2016

Variable Information

Detailed information about the raw variables can be found in the corresponding site where the data was obtained: http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

Variable units: variables that start with t are measures in seconds, while variables that start wth f are measured in Hz (frequency).

Summary choices: Variables were summarized as explained in the assignment. Specifically:

- The training and testing datasets were merged into one dataset.
- We only extracted the measurements on the mean and standard deviation for each measurement. This was interpreted as all columns that included the part 'mean()' or 'meanFreq()' or 'std()'
- Initial activity labels (numbers 1 to 6) were replaced by actual activity labels (strings), as described in the project website.
- Variable labels were replaces with actual variable names. The variable names were not substituted, although one could argue that they could be made more readable. For example, variable `fBodyAccMag-std()`, that denotes the standard deviation of the magnitude of the frequency obtained by the accelerometer could had been renamed as `frequenctBodyAccelerometerMagnitude-StDeviation`. It was a concious choice to leave the variable names to their original values.
- The final dataset was grouped by subject and activity, and the mean of each variable (for each subject-activity combination) was reported.

Experimental Study Design: This is described on the website where the dataset was obtained (see link above).

Instruction list

The script file that obtained and cleans the data is `run_analysis.R`. It downloads the data into a newly created folder, called `data` and unzips them. Then, it pulls data from the corresponding files, it combines them and cleanses them, as described in the assignment instructions. The file outputs two files, `tidy_data.csv`, and `merged.csv`, which are the two files that correspond to the final and the merged datasets respectively.

The file was run on a Mac under RStudio and it is not tested under Windows or Linux.