Human Activity Tracking Codebook

Dominik Sudwischer

22 October 2017

Extracting the Required Base Data

The original data set comprises 10299 obvervations collected from 30 volunteers that participated in a study capturing different movement measurements via the sensors of a Samsung Galacy S2 smartphone. According to the description of the original data set, the following facts hold true: The six activities (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) had their corresponding motions tracked by an embedded accelerometer and a gyroscope. To be more precise, 3-axial linear acceleration and angular velocity have been captured at a constant rate of 50 measurements per seconds. The labeling of the data has been performend manually using recorded video files. In the original data, the 10299 observations have been split into two datasets with 70% ('X_train.txt') and 30% ('X_test.txt') of the records respectively. Originally, there are 561 features. The corresponding files 'y_train' and 'y_test' contain the id of the subject. The text file 'activity labels.txt' explain each activity id, a number between 1 and 6, in plain text.

These 5 files were concatenated in the following way: for $s \in \{\text{"train"}, \text{"test"}\}$, X_s and y_s have been combined using cbind. After that, both data frames have been concatenated using rbind. The activity ids have been replaced the the actual name strings for each activity that can be seen in 'activity labels.txt'.

Changes to Variable Names

The names have been changed to be more readable by humans. In particular, leading 't' and 'f' have been renamed to 'time' and 'frequency', respectively. Also, 'Acc', 'Gyro' and 'Mag' have been renamed to 'acceleration', 'gyroscope' and 'magnitude'. All names have been transformed to lower case.

Included Variables

After the initial transformations performed by the run_analysis script the data consists of a single large data frame including a total of 10299 records and 68 variables, 66 of which are extracted from the original dataset. The last two ones are the subject ID, an integer between 1 and 30, and the activity as a captial letter string. The 66 extracted variables are those that are mean and standard deviations of measurements. Please note that frequency means ('MeanFreq()') are not included.

The resulting data set comprises means of each of those 66 variables, grouping the records by subject id and activity. The following variables are contained:

```
##  [1] "activity"
##  [2] "subject"
##  [3] "mean.of.time.body.acceleration.mean().x"
##  [4] "mean.of.time.body.acceleration.mean().y"
##  [5] "mean.of.time.body.acceleration.mean().z"
##  [6] "mean.of.time.body.acceleration.std().x"
##  [7] "mean.of.time.body.acceleration.std().y"
##  [8] "mean.of.time.body.acceleration.std().z"
##  [9] "mean.of.time.gravity.acceleration.mean().x"
## [10] "mean.of.time.gravity.acceleration.mean().y"
## [11] "mean.of.time.gravity.acceleration.mean().z"
```

```
## [12] "mean.of.time.gravity.acceleration.std().x"
## [13] "mean.of.time.gravity.acceleration.std().y"
## [14] "mean.of.time.gravity.acceleration.std().z"
## [15] "mean.of.time.body.acceleration.jerk.mean().x"
## [16] "mean.of.time.body.acceleration.jerk.mean().y"
## [17] "mean.of.time.body.acceleration.jerk.mean().z"
## [18] "mean.of.time.body.acceleration.jerk.std().x"
## [19] "mean.of.time.body.acceleration.jerk.std().y"
## [20] "mean.of.time.body.acceleration.jerk.std().z"
## [21] "mean.of.time.body.gyroscope.mean().x"
## [22] "mean.of.time.body.gyroscope.mean().y"
## [23] "mean.of.time.body.gyroscope.mean().z"
## [24] "mean.of.time.body.gyroscope.std().x"
## [25] "mean.of.time.body.gyroscope.std().y"
## [26] "mean.of.time.body.gyroscope.std().z"
## [27] "mean.of.time.body.gyroscope.jerk.mean().x"
## [28] "mean.of.time.body.gyroscope.jerk.mean().y"
## [29] "mean.of.time.body.gyroscope.jerk.mean().z"
## [30] "mean.of.time.body.gyroscope.jerk.std().x"
## [31] "mean.of.time.body.gyroscope.jerk.std().y"
## [32] "mean.of.time.body.gyroscope.jerk.std().z"
## [33] "mean.of.time.body.acceleration.magnitude.mean()"
## [34] "mean.of.time.body.acceleration.magnitude.std()"
## [35] "mean.of.time.gravity.acceleration.magnitude.mean()"
## [36] "mean.of.time.gravity.acceleration.magnitude.std()"
## [37] "mean.of.time.body.acceleration.jerkmagnitude.mean()"
## [38] "mean.of.time.body.acceleration.jerkmagnitude.std()"
## [39] "mean.of.time.body.gyroscope.magnitude.mean()"
## [40] "mean.of.time.body.gyroscope.magnitude.std()"
## [41] "mean.of.time.body.gyroscope.jerkmagnitude.mean()"
## [42] "mean.of.time.body.gyroscope.jerkmagnitude.std()"
## [43] "mean.of.frequency.body.acceleration.mean().x"
## [44] "mean.of.frequency.body.acceleration.mean().y"
## [45] "mean.of.frequency.body.acceleration.mean().z"
## [46] "mean.of.frequency.body.acceleration.std().x"
## [47] "mean.of.frequency.body.acceleration.std().y"
## [48] "mean.of.frequency.body.acceleration.std().z"
## [49] "mean.of.frequency.body.acceleration.jerk.mean().x"
## [50] "mean.of.frequency.body.acceleration.jerk.mean().y"
## [51] "mean.of.frequency.body.acceleration.jerk.mean().z"
## [52] "mean.of.frequency.body.acceleration.jerk.std().x"
## [53] "mean.of.frequency.body.acceleration.jerk.std().y"
## [54] "mean.of.frequency.body.acceleration.jerk.std().z"
## [55] "mean.of.frequency.body.gyroscope.mean().x"
## [56] "mean.of.frequency.body.gyroscope.mean().y"
## [57] "mean.of.frequency.body.gyroscope.mean().z"
## [58] "mean.of.frequency.body.gyroscope.std().x"
## [59] "mean.of.frequency.body.gyroscope.std().y"
## [60] "mean.of.frequency.body.gyroscope.std().z"
## [61] "mean.of.frequency.body.acceleration.magnitude.mean()"
## [62] "mean.of.frequency.body.acceleration.magnitude.std()"
## [63] "mean.of.frequency.bodybody.acceleration.jerkmagnitude.mean()"
## [64] "mean.of.frequency.bodybody.acceleration.jerkmagnitude.std()"
## [65] "mean.of.frequency.bodybody.gyroscope.magnitude.mean()"
```

```
## [66] "mean.of.frequency.bodybody.gyroscope.magnitude.std()"
## [67] "mean.of.frequency.bodybody.gyroscope.jerkmagnitude.mean()"
## [68] "mean.of.frequency.bodybody.gyroscope.jerkmagnitude.std()"
```

Each of these values have been generated by grouping the original data set by subject and activity and then computing the mean for each of the resulting 180 groups. That is, it contains the mean value of every single variable for each of the 30 subjects doing each of the six activities.

Units

The units are the same as in the original data. Please note that each feature has been normalized in the original data to be in [-1, 1]. The exact methods of generating the data, including filtering noise and applying Fourier transformations, can be looked up in the documentation of the original dataset. More detailed information can be found at http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones.