Codebook for the wide tidy data

November 20, 2015

1 Important notes

- The codebook for the final tidy averaged data would be the same. Indeed both tables contain the same variables, the averaged data being just an average over some observation of the wide data.
- all the features have been normalized in order to be in $[-1, 1]$.
- This notebook takes full support of the original one, that can be found at `http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones`
- The tidy data set was obtained having in mind the two following referecences
  - `www.jstatsoft.org/article/view/v059i10/v59i10.pdf`
  - `https://thoughtfulbloke.wordpress.com/2015/09/09/getting-and-cleaning-the-assignment/`

1.1 subject (column 1)

This feature indicates which subject is concerned. Labeled from 1 to 30.

1.2 activity (column 2)

This feature indicates which activity is concerned. 6 possible activities

- laying: the person is laying
- sitting: the person is sitting
- standing: the person is standing
- walking: the person is walking
- walkingdownstairs: the person is walking upstairs
- walkingupstairs: the person is walking downstairs

1.3 Other features (from column 3 to column 81

The other features selected for this database come from the accelerometer and gyroscope 3-axial raw signals timeacc-xyz and timegyro-xyz. These time domain signals were captured at a constant rate of 50 Hz.

They were then 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 (timebodyacc-xyz and timegravityacc-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 (timebodyaccmag, timegravityaccmag, timebodyaccjerkmag, timebodygyromag, timebodygyrojerkmag).

Finally a Fast Fourier Transform (FFT) was applied to some of these signals producing fbodyacc-xyz, frequencybodyaccjerk-xyz, frequencybodygyro-xyz, frequencybodyaccjerkmag, frequencybodygyromag, frequencybodygyrojerkmag.

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):

- timebodyacc-xyz
- timegravityacc-xyz
- timebodyaccJerk-xyz
- timebodygyro-xyz
- timebodygyroJerk-xyz
- timebodyaccmag
- timegravityaccmag
- timebodyaccJerkmag
- timebodygyromag
- timebodygyroJerkmag
- frequencybodyacc-xyz
- frequencybodyaccJerk-xyz
- frequencybodygyro-xyz
- frequencybodyaccmag
- frequencybodyaccJerkmag
- frequencybodygyromag
- frequencybodygyroJerkmag

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

- mean: mean value
- std: Standard deviation