CLEANED SENSORS' DATA

The Variables

subject

    A group of 30 volunteers within an age bracket of 19-48 years

    1-30

groupset

    The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.

    1. test

    2. train

activities

    The activities subject performed when a smartphone (Samsung Galaxy S II) was wore on the waist

    1. WALKING

    2. WALKING_UPSTAIRS

    3. WALKING_DOWNSTAIRS

    4. SITTING

    5. STANDING

    6. LAYING

measurements

    This database come from the accelerometer(Acc) and gytoscope(Gyro) 3-axial raw signals. Prefix 't' represents time domain signals, prefix 'f' represents frequency domain signals. The acceleration signal can be separated into body and gravity acceleration signals, represented by 'Body' and 'Gravity' respectively. 'Jerk' indicates Jerk signals. 'Mag' indicates the magnitude of these three-dimensional signals were calculated using the Euclidean norm. Two variables were estimated from these signals are mean() and std(), represent mean value and standard deviation respectively.

The data used

Under UCI HAR Dataset folder:

features.txt: List of all features measured in the raw data sets.

activity_labels.txt: links the class lables with their activity names.

Under UCI HAR Dataset folder/text folder:

X_text.txt: text subjects' data

y_text.txt: text subjects' activity labels

subject_text.txt: text subjects.

Under UCI HAR Dataset folder/train folder:
X_train.txt: train subjects' data
y_train.txt: train subjects' activity labels
subject_train.txt: train subjects.

The procedures of code

1. Download the sensors' data collected from the Samsung Galaxy S smartphone and unzip it.

2. Read the interested data sets seperately and assigned corresponding names.

3. Merge the test and train groupset data together. First, add a new column named 'groupset' and assign 'text' and 'train' corresponding to different group of subject to identify them. Second, form a new data frame combining activity labels and data within each subject groups. Finally, merge these two new text and train's data together as a whole data frame called data_all.

4. Extract only the measurements on the mean(mean()) and standard deviation(std()) for each measurement, and form a subset data frame called sub_data.

5. Use descriptive activity names in the activity_labels.txt in the original folder to replace the representative number showed in the sub_data.

6. Appropriately labels in the sub_data using full descriptive variable names.
   'mean' replaced by 'Mean'
   'std' replaced by 'StandatdDeviation'
   't' replaced by 'TimeDomain'
   'f' replaced by 'FrequencyDomain'
   'Acc' replaced by 'Acceleration'
   'Gyro' replaced by 'Grroscope'
   'Mag' replaced by 'Magnitude'
   'Freq' replaced by 'Frequency'
   '..X' replaced by 'X'
   '..Y' replaced by 'Y'
   '..Z' replaced by 'Z'

7. Create a second, independent tidy data set with the average of each variable for each activity and each subject from sub_data, finally create and export a tidy_data into a tidy_data.txt file in local.