Codebook for UCIHARtidy1.txt

# Study Design

(Taken from README file)

The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. 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.

The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low frequency components, therefore a filter with 0.3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain.

# Variable descriptors

For brevity, these are described in groups where similar. The descriptions are based on the document features\_info.txt.

activity  
 *Factor, 6 levels. Identifies activity being undertaken. Walk/Walkup/Walkdown/Stand/Sit/Lay*   
subject  
 *Integer 1 to 30. Identifies test subject.*

tbodyaccmeanx   
tbodyaccmeany  
tbodyaccmeanz  
tbodyaccstdx  
tbodyaccstdy  
tbodyaccstd   
 *mean and std deviations, body acceleration x, y, z axes. Units = g*

tgravityaccmeanx  
tgravityaccmeany  
tgravityaccmeanz  
tgravityaccstdx  
tgravityaccstdy  
tgravityaccstdz  
 *mean and std deviations, gravity acceleration x, y, z axes Units = g*

tbodyaccjerkmeanx  
tbodyaccjerkmeany  
tbodyaccjerkmeanz  
tbodyaccjerkstdx  
tbodyaccjerkstdy  
tbodyaccjerkstdz  
 *mean and std deviations, body linear jerk x, y, z axes Units = g*

tbodygyromeanx  
tbodygyromeany  
tbodygyromeanz  
tbodygyrostdx  
tbodygyrostdy  
tbodygyrostdz  
 *mean and std deviations, body gyroscope raw signals x, y, z axes Units = Radians/second*

tbodygyrojerkmeanx  
tbodygyrojerkmeany  
tbodygyrojerkmeanz  
tbodygyrojerkstdx  
tbodygyrojerkstdy  
tbodygyrojerkstdz  
 *mean and std deviations, body angular x, y, z axes Units = Radians/second*

tbodyaccmagmean  
tbodyaccmagstd  
tgravityaccmagmean  
tgravityaccmagstd  
tbodyaccjerkmagmean  
tbodyaccjerkmagstd  
tbodygyromagmean  
tbodygyromagstd  
tbodygyrojerkmagmean  
tbodygyrojerkmagstd  
 *mean and std deviation, magnitudes Units = g*

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  
 *mean and std deviations, fast fourier transforms Units = g (acceleration signals), radians/s (gyro signals)*

angletbodyaccmeangravity  
angletbodyaccjerkmeangravitymean  
angletbodygyromeangravitymean  
angletbodygyrojerkmeangravitymean  
anglexgravitymean  
angleygravitymean  
anglezgravitymean  
 *mean and std deviation angle between vectors. Units = g*

# Creation of the tidy dataset

The dataset was created from the documents in UCI HAR Dataset by applying the R script appended below. It works as follows:

The file run\_analysis.R requires that the uncompressed folder "UCI HAR Dataset" be in the working directory for R, wherever that may be. If in doubt run getwd() and place the "UCI HAR Dataset" folder in the directory identified. HOW DO YOU RUN IT??

The R script will then:

* read the file X\_train.txt - this file contains the actual data, 561 variables, 7352 observations
* read the file features.txt. The second column is used to name the variables in X-train.txt
* a search is run in the features labels for all variables containing 'mean' or 'std', and this is used to subset the X\_train dataset. This reduces it to 86 columns.
* read the file y\_train.txt - this file describes the activity being undertaken, 7352 rows. Name this 'activity' and add as a column to X\_train.
* read the file subject\_train.txt - this identifies the training subject (1 to 30). 7352 rows. Name this 'subject', and add as another column.
* Create a dataset with descriptive activity labels and merges this with the main dataset. This dataframe (trainmean)now has 88 columns (data plus activity plus subject).

The script will then perform a similar series of steps on the files in the 'test' folder, to create a dataframe called merged2.

The script will ignore the data in the Inertial Signals folder. Since these data do not contain any information on means or standard deviations, none of these figures are needed in the final tidy dataset. Not including them cuts out unnecessary steps and makes the scripts run faster.

merged1 and merged2 are combined with rbind() to create merged3, which has 88 columns and 7352 + 2947 = 10299 rows. The variable names are tidied up to remove caps, commas, and spurious brackets. This is saved as UCIHARtidy1.txt.

This is then grouped and summarized (means) and saved as UCIHARtidy2.txt

# R Script

run\_analysis <- function(){

#open the training data and label it with the features file

train <- read.table("UCI HAR Dataset/train/X\_train.txt")  
 features <- read.table("UCI HAR Dataset/features.txt")  
 colnames(train) <- features[,2]

#subset this to readings including mean or std

trainmean <-train[,grepl(  
 "\*[Mm]ean\*|\*[Ss]td\*",features[,2])]

#open and name the activity file

activity1 <- read.table("UCI HAR Dataset/train/y\_train.txt")  
 colnames(activity1) <- "activity1"

#open and name the subject file

subject1 <- read.table("UCI HAR  
 Dataset/train/subject\_train.txt")  
 colnames(subject1) <- "subject"

#join trainmean, activity1 and subject1

trainmean <- cbind(activity1,subject1,trainmean)

#open the test data and label it with the features file

test <- read.table("UCI HAR Dataset/test/X\_test.txt")  
 colnames(test) <- features[,2]

#subset this to readings including mean or std

testmean <- test[,grepl("\*[Mm]ean\*|\*[Ss]td\*",features[,2])]

#open and name the activity file

activity2 <- read.table("UCI HAR Dataset/test/y\_test.txt")  
 colnames(activity2) <- "activity1"

#open and name the subject file

subject2 <- read.table("UCI HAR  
 Dataset/test/subject\_test.txt")

colnames(subject2) <- "subject"

#join testmean, activity2 and subject2

testmean <- cbind(activity2,subject2,testmean)

#join the test and training files

merged1 <- rbind(trainmean,testmean)

#tidy up the columnnames by removing brackets dashes commas  
 and caps

colnames(merged1) <- tolower(gsub(  
 "\\(|\\)|-|,","",colnames(merged1)))

#create the activity descriptors

actlabels <- data.frame(1:6,c(  
 "walk","walkup","walkdown","sit","stand","lay"))  
 colnames(actlabels)<- c("activity1","activity")

#use merge to join the actlabels and merged1 dataframes

merged2 <- merge(actlabels,merged1)

#ditch activity1 as we no longer require it

merged2 <- merged2[,c(2:89)]

#write the first dataset

write.table(merged2, file=  
 "UCIHARtidy1.txt",row.names=FALSE,sep="\t")

#use dplyr to summarize each variable by group

library(dplyr)

merged3 <- tbl\_df(merged2) %>%

group\_by(subject, activity) %>%

summarise\_each(funs(mean),  
 tbodyaccmeanx:anglezgravitymean)

#write the second table

write.table(merged3, file=  
 "UCIHARtidy2.txt",row.names=FALSE,sep="\t")

}