Codebook GCD assignment

2023-09-16

Data overview

The dataset contains several measures of physical activity across six different types of activity and 30 persons. It is a tidy dataset that combines the different files of the *Human Activity Recognition Using Smartphones Dataset*. See below for the original description of the activity measures.

Description of variables in the data

Original Readme file

==========
Human Activity Recognition Using Smartphones Dataset
Version 1.0
==========
Jorge L. Reyes-Ortiz, Davide Anguita, Alessandro Ghio, Luca Oneto.
Smartlab - Non Linear Complex Systems Laboratory
DITEN - Università degli Studi di Genova.
Via Opera Pia 11A, I-16145, Genoa, Italy.
activityrecognition@smartlab.ws
www.smartlab.ws
==========

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. See 'features_info.txt' for more details.

For each record it is provided:

- Triaxial acceleration from the accelerometer (total acceleration) and the estimated body acceleration.
- Triaxial Angular velocity from the gyroscope.
- A 561-feature vector with time and frequency domain variables.
- Its activity label.
- An identifier of the subject who carried out the experiment.

Table 1: Aggregated

Name	Class	Values	Missing	Summary
person	integer	Num: 1 to 30	0	mean: 15.5, sd: 8.68, nuniq: 30
activity	factor	'WALKING' 'WALKING_UPSTAIRS' 'WALKING_DOWNSTAIRS' 'SITTING' 'STANDING' and 1 more	0	nuniq: 6
tBodyAcc-mean()-X	numeric	Num: 0.222 to 0.301	0	mean: 0.274, sd: 0.012, nuniq: 180
tBodyAcc-mean()-Y	numeric	Num: -0.041 to -0.001	0	mean: -0.018, sd: 0.006, nuniq: 180
tBodyAcc-mean()-Z	numeric	Num: -0.153 to -0.075	0	mean: -0.109, sd: 0.01, nuniq: 180
tBodyAcc-std()-X	numeric	Num: -0.996 to 0.627	0	mean: -0.558, sd: 0.452, nuniq: 180
tBodyAcc-std()-Y	numeric	Num: -0.99 to 0.617	0	mean: -0.46, sd: 0.497, nuniq: 180
tBodyAcc-std()-Z	numeric	Num: -0.988 to 0.609	0	mean: -0.576, sd: 0.396, nuniq: 180
tGravityAcc-mean()-X	numeric	Num: -0.68 to 0.975	0	mean: 0.697, sd: 0.487, nuniq: 180
tGravityAcc-mean()-Y	numeric	Num: -0.48 to 0.957	0	mean: -0.016, sd: 0.345, nuniq: 180
tGravityAcc-mean()-Z	numeric	Num: -0.495 to 0.958	0	mean: 0.074, sd: 0.289, nuniq: 180
tGravityAcc-std()-X	numeric	Num: -0.997 to -0.83	0	mean: -0.964, sd: 0.025, nuniq: 180
tGravityAcc-std()-Y	numeric	Num: -0.994 to -0.644	0	mean: -0.952, sd: 0.033, nuniq: 180
tGravityAcc-std()-Z	numeric	Num: -0.991 to -0.61	0	mean: -0.936, sd: 0.04, nuniq: 180
tBodyAccJerk-mean()-X	numeric	Num: 0.043 to 0.13	0	mean: 0.079, sd: 0.013, nuniq: 180
tBodyAccJerk-mean()-Y	numeric	Num: -0.039 to 0.057	0	mean: 0.008, sd: 0.014, nuniq: 180
tBodyAccJerk-mean()-Z	numeric	Num: -0.067 to 0.038	0	mean: -0.005, sd: 0.013, nuniq: 180
tBodyAccJerk-std()-X	numeric	Num: -0.995 to 0.544	0	mean: -0.595, sd: 0.418, nuniq: 180
tBodyAccJerk-std()-Y	numeric	Num: -0.99 to 0.355	0	mean: -0.565, sd: 0.433, nuniq: 180
tBodyAccJerk-std()-Z	numeric	Num: -0.993 to 0.031	0	mean: -0.736, sd: 0.277, nuniq: 180
tBodyGyro-mean()-X	numeric	Num: -0.206 to 0.193	0	mean: -0.032, sd: 0.054, nuniq: 180
tBodyGyro-mean()-Y	numeric	Num: -0.204 to 0.027	0	mean: -0.074, sd: 0.036, nuniq: 180
tBodyGyro-mean()-Z	numeric	Num: -0.072 to 0.179	0	mean: 0.087, sd: 0.036, nuniq: 180
tBodyGyro-std()-X	numeric	Num: -0.994 to 0.268	0	mean: -0.692, sd: 0.291, nuniq: 180
tBodyGyro-std()-Y	numeric	Num: -0.994 to 0.477	0	mean: -0.653, sd: 0.352, nuniq: 180
tBodyGyro-std()-Z	numeric	Num: -0.986 to 0.565	0	mean: -0.616, sd: 0.373, nuniq: 180
tBodyGyroJerk-mean()-X	numeric	Num: -0.157 to -0.022	0	mean: -0.096, sd: 0.023, nuniq: 180
tBodyGyroJerk-mean()-Y	numeric	Num: -0.077 to -0.013	0	mean: -0.043, sd: 0.01, nuniq: 180
tBodyGyroJerk-mean()-Z	numeric	Num: -0.092 to -0.007	0	mean: -0.055, sd: 0.012, nuniq: 180
tBodyGyroJerk-std()-X	numeric	Num: -0.997 to 0.179	0	mean: -0.704, sd: 0.301, nuniq: 180
tBodyGyroJerk-std()-Y	numeric	Num: -0.997 to 0.296	0	mean: -0.764, sd: 0.267, nuniq: 180
tBodyGyroJerk-std()-Z	numeric	Num: -0.995 to 0.193	0	mean: -0.71, sd: 0.305, nuniq: 180
tBodyAccMag-mean()	numeric	Num: -0.986 to 0.645	0	mean: -0.497, sd: 0.473, nuniq: 180
tBodyAccMag-std()	numeric	Num: -0.986 to 0.428	0	mean: -0.544, sd: 0.431, nuniq: 180
tGravityAccMag-mean()	numeric	Num: -0.986 to 0.645	0	mean: -0.497, sd: 0.473, nuniq: 180
tGravityAccMag-std()	numeric	Num: -0.986 to 0.428	0	mean: -0.544, sd: 0.431, nuniq: 180
tBodyAccJerkMag-mean()	numeric	Num: -0.993 to 0.434	0	mean: -0.608, sd: 0.397, nuniq: 180
tBodyAccJerkMag-std()	numeric	Num: -0.995 to 0.451	0	mean: -0.584, sd: 0.423, nuniq: 180
tBodyGyroMag-mean()	numeric	Num: -0.981 to 0.418	0	mean: -0.565, sd: 0.398, nuniq: 180
tBodyGyroMag-std()	numeric	Num: -0.981 to 0.3	0	mean: -0.63, sd: 0.337, nuniq: 180
tBodyGyroJerkMag-mean()	numeric	Num: -0.997 to 0.088	0	mean: -0.736, sd: 0.277, nuniq: 180
tBodyGyroJerkMag-std()	numeric	Num: -0.998 to 0.25	0	mean: -0.755, sd: 0.266, nuniq: 180
fBodyAcc-mean()-X	numeric	Num: -0.995 to 0.537	0	mean: -0.576, sd: 0.43, nuniq: 180
fBodyAcc-mean()-Y	numeric	Num: -0.989 to 0.524	0	mean: -0.489, sd: 0.481, nuniq: 180
fBodyAcc-mean()-Z	numeric	Num: -0.989 to 0.281	0	mean: -0.63, sd: 0.356, nuniq: 180
fBodyAcc-std()-X	numeric	Num: -0.997 to 0.659	0	mean: -0.552, sd: 0.46, nuniq: 180
fBodyAcc-std()-Y	numeric	Num: -0.991 to 0.56	0	mean: -0.481, sd: 0.474, nuniq: 180
fBodyAcc-std()-Z	numeric	Num: -0.987 to 0.687	0	mean: -0.582, sd: 0.388, nuniq: 180
fBodyAcc-meanFreq()-X	numeric	Num: -0.636 to 0.159	0	mean: -0.232, sd: 0.194, nuniq: 180
fBodyAcc-meanFreq()-Y	numeric	Num: -0.38 to 0.467	0	mean: 0.012, sd: 0.145, nuniq: 180
fBodyAcc-meanFreq()-Z	numeric	Num: -0.52 to 0.403	0	mean: 0.044, sd: 0.185, nuniq: 180
fBodyAccJerk-mean()-X	numeric	Num: -0.995 to 0.474	0	mean: -0.614, sd: 0.398, nuniq: 180
fBodyAccJerk-mean()-Y	numeric	Num: -0.989 to 0.277	0	mean: -0.588, sd: 0.408, nuniq: 180
fBodyAccJerk-mean()-Z	numeric	Num: -0.992 to 0.158	0	mean: -0.714, sd: 0.297, nuniq: 180
fBodyAccJerk-std()-X	numeric	Num: -0.995 to 0.477	0	mean: -0.612, sd: 0.4, nuniq: 180
fBodyAccJerk-std()-Y	numeric	Num: -0.99 to 0.35	0	mean: -0.571, sd: 0.432, nuniq: 180
fBodyAccJerk-std()-Z	numeric	Num: -0.993 to -0.006	0	mean: -0.756, sd: 0.257, nuniq: 180
fBodyAccJerk-meanFreq()-X	numeric	Num: -0.576 to 0.331	0	mean: -0.069, sd: 0.254, nuniq: 180
fBodyAccJerk-meanFreq()-Y	numeric	Num: -0.602 to 0.196	0	mean: -0.228, sd: 0.2, nuniq: 180
fBodyAccJerk-meanFreq()-Z	numeric	Num: -0.628 to 0.23	0	mean: -0.138, sd: 0.208, nuniq: 180
fBodyGyro-mean()-X	numeric	Num: -0.993 to 0.475	0	mean: -0.637, sd: 0.347, nuniq: 180
fBodyGyro-mean()-Y	numeric	Num: -0.994 to 0.329	0	mean: -0.677, sd: 0.332, nuniq: 180
fBodyGyro-mean()-Z	numeric	Num: -0.986 to 0.492	0	mean: -0.604, sd: 0.384, nuniq: 180
fBodyGyro-std()-X	numeric	Num: -0.995 to 0.197	0	mean: -0.711, sd: 0.273, nuniq: 180
fBodyGyro-std()-Y	numeric	Num: -0.994 to 0.646	0	mean: -0.645, sd: 0.363, nuniq: 180
fBodyGyro-std()-Z	numeric	Num: -0.987 to 0.522	0	mean: -0.658, sd: 0.336, nuniq: 180
fBodyGyro-meanFreq()-X	numeric	Num: -0.396 to 0.249	0	mean: -0.105, sd: 0.148, nuniq: 180
fBodyGyro-meanFreq()-Y	numeric	Num: -0.667 to 0.273	0	mean: -0.167, sd: 0.179, nuniq: 180
fBodyGyro-meanFreq()-Z	numeric	Num: -0.507 to 0.377	0	mean: -0.057, sd: 0.165, nuniq: 180
fBodyAccMag-mean()	numeric	Num: -0.987 to 0.587	0	mean: -0.537, sd: 0.452, nuniq: 180
fBodyAccMag-std()	numeric	Num: -0.988 to 0.179	0	mean: -0.621, sd: 0.353, nuniq: 180
fBodyAccMag-meanFreq()	numeric	Num: -0.312 to 0.436	0	mean: 0.076, sd: 0.14, nuniq: 180
fBodyBodyAccJerkMag-mean()	numeric	Num: -0.994 to 0.538	0	mean: -0.576, sd: 0.431, nuniq: 180
fBodyBodyAccJerkMag-std()	numeric	Num: -0.994 to 0.316	0	mean: -0.599, sd: 0.409, nuniq: 180
fBodyBodyAccJerkMag-meanFreq()	numeric	Num: -0.125 to 0.488	0	mean: 0.163, sd: 0.138, nuniq: 180
fBodyBodyGyroMag-mean()	numeric	Num: -0.987 to 0.204	0	mean: -0.667, sd: 0.318, nuniq: 180
fBodyBodyGyroMag-std()	numeric	Num: -0.981 to 0.237	0	mean: -0.672, sd: 0.293, nuniq: 180
fBodyBodyGyroMag-meanFreq()	numeric	Num: -0.457 to 0.41	0	mean: -0.036, sd: 0.181, nuniq: 180
fBodyBodyGyroJerkMag-mean()	numeric	Num: -0.998 to 0.147	0	mean: -0.756, sd: 0.263, nuniq: 180
fBodyBodyGyroJerkMag-std()	numeric	Num: -0.998 to 0.288	0	mean: -0.772, sd: 0.25, nuniq: 180
fBodyBodyGyroJerkMag-meanFreq()	numeric	Num: -0.183 to 0.426	0	mean: 0.126, sd: 0.108, nuniq: 180
angle(tBodyAccMean,gravity)	numeric	Num: -0.163 to 0.129	0	mean: 0.007, sd: 0.04, nuniq: 180
angle(tBodyAccJerkMean),gravityMean)	numeric	Num: -0.121 to 0.203	0	mean: 0.001, sd: 0.043, nuniq: 180
angle(tBodyGyroMean,gravityMean)	numeric	Num: -0.389 to 0.444	0	mean: 0.022, sd: 0.14, nuniq: 180
angle(tBodyGyroJerkMean,gravityMean)	numeric	Num: -0.224 to 0.182	0	mean: -0.011, sd: 0.073, nuniq: 180
angle(X,gravityMean)	numeric	Num: -0.947 to 0.738	0	mean: -0.524, sd: 0.481, nuniq: 180
angle(Y,gravityMean)	numeric	Num: -0.875 to 0.425	0	mean: 0.079, sd: 0.278, nuniq: 180
angle(Z,gravityMean)	numeric	Num: -0.874 to 0.39	0	mean: -0.04, sd: 0.231, nuniq: 180

The dataset includes the following files:

- 'README.txt'
- 'features_info.txt': Shows information about the variables used on the feature vector.
- 'features.txt': List of all features.
- 'activity_labels.txt': Links the class labels with their activity name.
- 'train/X_train.txt': Training set.
- 'train/y_train.txt': Training labels.
- 'test/X_test.txt': Test set.
- 'test/y_test.txt': Test labels.

The following files are available for the train and test data. Their descriptions are equivalent.

- 'train/subject_train.txt': Each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30.
- 'train/Inertial Signals/total_acc_x_train.txt': The acceleration signal from the smartphone accelerometer X axis in standard gravity units 'g'. Every row shows a 128 element vector. The same description applies for the 'total_acc_x_train.txt' and 'total_acc_z_train.txt' files for the Y and Z axis.
- 'train/Inertial Signals/body_acc_x_train.txt': The body acceleration signal obtained by subtracting the gravity from the total acceleration.
- 'train/Inertial Signals/body_gyro_x_train.txt': The angular velocity vector measured by the gyroscope for each window sample. The units are radians/second.

Notes:

- Features are normalized and bounded within [-1,1].
- Each feature vector is a row on the text file.

For more information about this dataset contact: activityrecognition@smartlab.ws

License:

Use of this dataset in publications must be acknowledged by referencing the following publication [1]

[1] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. International Workshop of Ambient Assisted Living (IWAAL 2012). Vitoria-Gasteiz, Spain. Dec 2012

Jorge L. Reyes-Ortiz, Alessandro Ghio, Luca Oneto, Davide Anguita. November 2012.