Getting and cleaning data course project codebook

Contents

Introduction:

In Dec 2012 an experiment was carried out on 30 volunteers aged between 19-48 years on Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine[1] Each person performed six activities (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II). Using the embedded accelerometer and gyroscope3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz was captured for a number of different variables. The data generated from this experiment was randomly partitioned to create two datasets; with the data from 70% of the volunteers used to create a training dataset and that from the remainder used to create a test dataset.

The dataset created as part of this project creates a wide format tidy dataset of a subset of the combined experimental data variables – namely the mean and standard deviation measurements of each variable. For each of these variables the dataset provides the average recorded value for each volunteer by activity.

Variables:

From the original experiment: The variables selected for this database come from the accelerometer and gyroscope 3-axial raw signals tAcc-XYZ and tGyro-XYZ. These time domain signals (prefix 't' to denote time) were captured at a constant rate of 50 Hz.. Similarly, the acceleration signal was then separated into body and gravity acceleration signals (tBodyAcc-XYZ and tGravityAcc-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 (tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag). Finally a Fast Fourier Transform (FFT) was applied to some of these signals producing fBodyAcc-XYZ, fBodyAccJerk-XYZ, fBodyGyro-XYZ, fBodyAccJerkMag, fBodyGyroMag, fBodyGyroJerkMag. (Note the 'f' to indicate frequency domain signals).

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.

Tidied dataset: The set of variables used these signals are mean and standard deviation (std).

tBodyAcc-XYZ

tGravityAcc-XYZ

tBodyAccJerk-XYZ

tBodyGyroXYZ

tBodyGyroJerkXYZ

tBodyAccMag

tGravityAccMag

tBodyAccJerkMag

tBodyGyroMag

tBodyGyroJerkMag

fBodyAccXYZ

fBodyAccJerkXYZ

fBodyGyroXYZ

fBodyAccMag

fBodyAccJerkMag

fBodyGyroMag

fBodyGyroJerkMag

Tidied Dataset:

WALKING_1: mean values calculated for volunteer # 1 while walking. Data type: numeric

WALKING_2: mean values calculated for volunteer # 2 while walking. Data type: numeric

WALKING_3: mean values calculated for volunteer # 3 while walking. Data type: numeric

WALKING_4: mean values calculated for volunteer # 4 while walking. Data type: numeric

WALKING_5: mean values calculated for volunteer # 5 while walking. Data type: numeric

WALKING_6: mean values calculated for volunteer # 6 while walking. Data type: numeric

WALKING_7: mean values calculated for volunteer # 7 while walking. Data type: numeric

WALKING_8: mean values calculated for volunteer # 8 while walking. Data type: numeric

WALKING_9: mean values calculated for volunteer # 9 while walking. Data type: numeric

WALKING_10: mean values calculated for volunteer # 10 while walking. Data type: numeric

WALKING_11: mean values calculated for volunteer # 11 while walking. Data type: numeric

WALKING_12: mean values calculated for volunteer # 12 while walking. Data type: numeric

WALKING_13: mean values calculated for volunteer # 13 while walking. Data type: numeric

WALKING_14: mean values calculated for volunteer # 14 while walking. Data type: numeric

WALKING_15: mean values calculated for volunteer # 15 while walking. Data type: numeric

WALKING_16: mean values calculated for volunteer # 16 while walking. Data type: numeric

WALKING_17: mean values calculated for volunteer # 17 while walking. Data type: numeric

WALKING_18: mean values calculated for volunteer # 18 while walking. Data type: numeric

WALKING_19: mean values calculated for volunteer # 19 while walking. Data type: numeric

WALKING_20: mean values calculated for volunteer # 20 while walking. Data type: numeric

WALKING_21: mean values calculated for volunteer # 21 while walking. Data type: numeric

WALKING_22: mean values calculated for volunteer # 22 while walking. Data type: numeric

WALKING_23: mean values calculated for volunteer # 23 while walking. Data type: numeric

WALKING_24: mean values calculated for volunteer # 24 while walking. Data type: numeric

WALKING_25: mean values calculated for volunteer # 25 while walking. Data type: numeric

WALKING_26: mean values calculated for volunteer # 26 while walking. Data type: numeric

WALKING_27: mean values calculated for volunteer # 27 while walking. Data type: numeric

WALKING_28: mean values calculated for volunteer # 28 while walking. Data type: numeric

WALKING_29: mean values calculated for volunteer # 29 while walking. Data type: numeric

WALKING_30: mean values calculated for volunteer # 30 while walking. Data type: numeric

WALKING_UPSTAIRS_1: mean values calculated for volunteer # 1 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_2: mean values calculated for volunteer # 2 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_3: mean values calculated for volunteer # 3 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_4: mean values calculated for volunteer # 4 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_5: mean values calculated for volunteer # 5 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_6: mean values calculated for volunteer # 6 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_7: mean values calculated for volunteer # 7 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_8: mean values calculated for volunteer # 8 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_9: mean values calculated for volunteer # 9 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_10: mean values calculated for volunteer # 10 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_11: mean values calculated for volunteer # 11 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_12: mean values calculated for volunteer # 12 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_13: mean values calculated for volunteer # 13 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_14: mean values calculated for volunteer # 14 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_15: mean values calculated for volunteer # 15 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_16: mean values calculated for volunteer # 16 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_17: mean values calculated for volunteer # 17 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_18: mean values calculated for volunteer # 18 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_19: mean values calculated for volunteer # 19 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_20: mean values calculated for volunteer # 20 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_21: mean values calculated for volunteer # 21 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_22: mean values calculated for volunteer # 22 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_23: mean values calculated for volunteer # 23 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_24: mean values calculated for volunteer # 24 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_25: mean values calculated for volunteer # 25 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_26: mean values calculated for volunteer # 26 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_27: mean values calculated for volunteer # 27 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_28: mean values calculated for volunteer # 28 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_29: mean values calculated for volunteer # 29 while walking upstairs. Data type: numeric

WALKING_UPSTAIRS_30: mean values calculated for volunteer # 30 while walking upstairs. Data type: numeric

WALKING_DOWNSTAIRS_1: mean values calculated for volunteer # 1 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_2: mean values calculated for volunteer # 2 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_3: mean values calculated for volunteer # 3 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_4: mean values calculated for volunteer # 4 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_5: mean values calculated for volunteer # 5 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_6: mean values calculated for volunteer # 6 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_7: mean values calculated for volunteer # 7 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_8: mean values calculated for volunteer # 8 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_9: mean values calculated for volunteer # 9 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_10: mean values calculated for volunteer # 10 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_11: mean values calculated for volunteer # 11 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_12: mean values calculated for volunteer # 12 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_13: mean values calculated for volunteer # 13 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_14: mean values calculated for volunteer # 14 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_15: mean values calculated for volunteer # 15 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_16: mean values calculated for volunteer # 16 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_17: mean values calculated for volunteer # 17 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_18: mean values calculated for volunteer # 18 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_19: mean values calculated for volunteer # 19 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_20: mean values calculated for volunteer # 20 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_21: mean values calculated for volunteer # 21 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_22: mean values calculated for volunteer # 22 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_23: mean values calculated for volunteer # 23 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_24: mean values calculated for volunteer # 24 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_25: mean values calculated for volunteer # 25 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_26: mean values calculated for volunteer # 26 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_27: mean values calculated for volunteer # 27 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_28: mean values calculated for volunteer # 28 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_29: mean values calculated for volunteer # 29 while walking downstairs. Data type: numeric

WALKING_DOWNSTAIRS_30: mean values calculated for volunteer # 30 while walking downstairs. Data type: numeric

SITTING_1: mean values calculated for volunteer # 1 sitting. Data type: numeric

SITTING_2: mean values calculated for volunteer # 2 sitting. Data type: numeric

SITTING_3: mean values calculated for volunteer # 3 sitting. Data type: numeric

SITTING_4: mean values calculated for volunteer # 4 sitting. Data type: numeric

SITTING_5: mean values calculated for volunteer # 5 sitting. Data type: numeric

SITTING_6: mean values calculated for volunteer # 6 sitting. Data type: numeric

SITTING_7: mean values calculated for volunteer # 7 sitting. Data type: numeric

SITTING_8: mean values calculated for volunteer # 8 sitting. Data type: numeric

SITTING_9: mean values calculated for volunteer # 9 sitting. Data type: numeric

SITTING_10: mean values calculated for volunteer # 10 sitting. Data type: numeric

SITTING_11: mean values calculated for volunteer # 11 sitting. Data type: numeric

SITTING_12: mean values calculated for volunteer # 12 sitting. Data type: numeric

SITTING_13: mean values calculated for volunteer # 13 sitting. Data type: numeric

SITTING_14: mean values calculated for volunteer # 14 sitting. Data type: numeric

SITTING_15: mean values calculated for volunteer # 15 sitting. Data type: numeric

SITTING_16: mean values calculated for volunteer # 16 sitting. Data type: numeric

SITTING_17: mean values calculated for volunteer # 17 sitting. Data type: numeric

SITTING_18: mean values calculated for volunteer # 18 sitting. Data type: numeric

SITTING_19: mean values calculated for volunteer # 19 sitting. Data type: numeric

SITTING_20: mean values calculated for volunteer # 20 sitting. Data type: numeric

SITTING_21: mean values calculated for volunteer # 21 sitting. Data type: numeric

SITTING_22: mean values calculated for volunteer # 22 sitting. Data type: numeric

SITTING_23: mean values calculated for volunteer # 23 sitting. Data type: numeric

SITTING_24: mean values calculated for volunteer # 24 sitting. Data type: numeric

SITTING_25: mean values calculated for volunteer # 25 sitting. Data type: numeric

SITTING_26: mean values calculated for volunteer # 26 sitting. Data type: numeric

SITTING_27: mean values calculated for volunteer # 27 sitting. Data type: numeric

SITTING_28: mean values calculated for volunteer # 28 sitting. Data type: numeric

SITTING_29: mean values calculated for volunteer # 29 sitting. Data type: numeric

SITTING_30: mean values calculated for volunteer # 30 sitting. Data type: numeric

STANDING_1: mean values calculated for volunteer # 1 while standing. Data type: numeric

STANDING_2: mean values calculated for volunteer # 2 while standing. Data type: numeric

STANDING_3: mean values calculated for volunteer # 3 while standing. Data type: numeric

STANDING_4: mean values calculated for volunteer # 4 while standing. Data type: numeric

STANDING_5: mean values calculated for volunteer # 5 while standing. Data type: numeric

STANDING_6: mean values calculated for volunteer # 6 while standing. Data type: numeric

STANDING_7: mean values calculated for volunteer # 7 while standing. Data type: numeric

STANDING_8: mean values calculated for volunteer # 8 while standing. Data type: numeric

STANDING_9: mean values calculated for volunteer # 9 while standing. Data type: numeric

STANDING_10: mean values calculated for volunteer # 10 while standing. Data type: numeric

STANDING_11: mean values calculated for volunteer # 11 while standing. Data type: numeric

STANDING_12: mean values calculated for volunteer # 12 while standing. Data type: numeric

STANDING_13: mean values calculated for volunteer # 13 while standing. Data type: numeric

STANDING_14: mean values calculated for volunteer # 14 while standing. Data type: numeric

STANDING_15: mean values calculated for volunteer # 15 while standing. Data type: numeric

STANDING_16: mean values calculated for volunteer # 16 while standing. Data type: numeric

STANDING_17: mean values calculated for volunteer # 17 while standing. Data type: numeric

STANDING_18: mean values calculated for volunteer # 18 while standing. Data type: numeric

STANDING_19: mean values calculated for volunteer # 19 while standing. Data type: numeric

STANDING_20: mean values calculated for volunteer # 20 while standing. Data type: numeric

STANDING_21: mean values calculated for volunteer # 21 while standing. Data type: numeric

STANDING_22: mean values calculated for volunteer # 22 while standing. Data type: numeric

STANDING_23: mean values calculated for volunteer # 23 while standing. Data type: numeric

STANDING_24: mean values calculated for volunteer # 24 while standing. Data type: numeric

STANDING_25: mean values calculated for volunteer # 25 while standing. Data type: numeric

STANDING_26: mean values calculated for volunteer # 26 while standing. Data type: numeric

STANDING_27: mean values calculated for volunteer # 27 while standing. Data type: numeric

STANDING_28: mean values calculated for volunteer # 28 while standing. Data type: numeric

STANDING_29: mean values calculated for volunteer # 29 while standing. Data type: numeric

STANDING_30: mean values calculated for volunteer # 30 while standing. Data type: numeric

LAYING_1: mean values calculated for volunteer # 1 while laying down. Data type: numeric

LAYING_2: mean values calculated for volunteer # 2 while laying down. Data type: numeric

LAYING_3: mean values calculated for volunteer # 3 while laying down. Data type: numeric

LAYING_4: mean values calculated for volunteer # 4 while laying down. Data type: numeric

LAYING_5: mean values calculated for volunteer # 5 while laying down. Data type: numeric

LAYING_6: mean values calculated for volunteer # 6 while laying down. Data type: numeric

LAYING_7: mean values calculated for volunteer # 7 while laying down. Data type: numeric

LAYING_8: mean values calculated for volunteer # 8 while laying down. Data type: numeric

LAYING_9: mean values calculated for volunteer # 9 while laying down. Data type: numeric

LAYING_10: mean values calculated for volunteer # 10 while laying down. Data type: numeric

LAYING_11: mean values calculated for volunteer # 11 while laying down. Data type: numeric

LAYING_12: mean values calculated for volunteer # 12 while laying down. Data type: numeric

LAYING_13: mean values calculated for volunteer # 13 while laying down. Data type: numeric

LAYING_14: mean values calculated for volunteer # 14 while laying down. Data type: numeric

LAYING_15: mean values calculated for volunteer # 15 while laying down. Data type: numeric

LAYING_16: mean values calculated for volunteer # 16 while laying down. Data type: numeric

LAYING_17: mean values calculated for volunteer # 17 while laying down. Data type: numeric

LAYING_18: mean values calculated for volunteer # 18 while laying down. Data type: numeric

LAYING_19: mean values calculated for volunteer # 19 while laying down. Data type: numeric

LAYING_20: mean values calculated for volunteer # 20 while laying down. Data type: numeric

LAYING_21: mean values calculated for volunteer # 21 while laying down. Data type: numeric

LAYING_22: mean values calculated for volunteer # 22 while laying down. Data type: numeric

LAYING_23: mean values calculated for volunteer # 23 while laying down. Data type: numeric

LAYING_24: mean values calculated for volunteer # 24 while laying down. Data type: numeric

LAYING_25: mean values calculated for volunteer # 25 while laying down. Data type: numeric

LAYING_26: mean values calculated for volunteer # 26 while laying down. Data type: numeric

LAYING_27: mean values calculated for volunteer # 27 while laying down. Data type: numeric

LAYING_28: mean values calculated for volunteer # 28 while laying down. Data type: numeric

LAYING_29: mean values calculated for volunteer # 29 while laying down. Data type: numeric

LAYING_30: mean values calculated for volunteer # 30 while laying down. Data type: numeric

References:

[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