CODEBOOK

Codebook

Dataset overview

The final data set includes appropriate variable names that have been processed from features.txt to have descriptive names that are appropriate for coding conventions. So, 'fBodyGyro-mean()-X' transforms into f_body_gyro_mean_x

Steps to Create Relevant Feature Names

Steps to clean and extract the 66 mean() / std() feature names

Load raw feature list Read features.txt, keeping only the second column (feature), which contains the 561 original variable names.

Standardise the names

Insert underscores between lower- and uppercase letters (tBody → t_Body).

Convert the entire string to lower-case.

Strip parentheses () and commas.

Replace dashes - with underscores _. Result: tidy, snake-case names like t_body_acc_mean_x.

Create an indexed table Build a data.table with two columns—idx (1-561) and the cleaned name—for easy sub-setting.

Identify target variables Use grep("mean\(\)|std\(\)", feature_names) on the original (uncleaned) names to capture the exact 66 indices corresponding to measurements derived from mean() and std() (excluding meanFreq and angle variables).

Keep only those rows Filter the data.table with the index vector, yielding features_keep: the definitive list of 66 cleaned variable names ready for sub-setting the merged dataset.

Steps to Create Final Dataset

The activity and subject column is combined with the train set and test set respectively, then the data sets are merged. Then the activity column is given meaningful labels so 1 is mapped as: 1 -> walking 2 -> walking_up 3 -> walking_down 4 -> sitting 5 -> standing 6 -> laying

The data set is then grouped by subject and activity and summarised by mean for all columns kept. Note that this reduces # of observations

Features

Features are in the range [-1, 1] for all suffixed with an x,y, or z Features which have mag in it are in range [0, 1], this is because the euclidean distance is non-negative and original features were normalized after passing respective filters.

Feature Names

t_body_acc_mean_x
t_body_acc_mean_y
t_body_acc_mean_z
t_body_acc_std_x
t_body_acc_std_y
t_body_acc_std_z
t_gravity_acc_mean_x
t_gravity_acc_mean_y
t_gravity_acc_mean_z
t_gravity_acc_std_x
t_gravity_acc_std_y
t_gravity_acc_std_z
t_body_acc_jerk_mean_x
t_body_acc_jerk_mean_y
t_body_acc_jerk_mean_z
t_body_acc_jerk_std_x
t_body_acc_jerk_std_y
t_body_acc_jerk_std_z
t_body_gyro_mean_x
t_body_gyro_mean_y
t_body_gyro_mean_z
t_body_gyro_std_x
t_body_gyro_std_y
t_body_gyro_std_z
t_body_gyro_jerk_mean_x
t_body_gyro_jerk_mean_y
t_body_gyro_jerk_mean_z
t_body_gyro_jerk_std_x
t_body_gyro_jerk_std_y
t_body_gyro_jerk_std_z
t_body_acc_mag_mean
t_body_acc_mag_std
t_gravity_acc_mag_mean
t_gravity_acc_mag_std
t_body_acc_jerk_mag_mean
t_body_acc_jerk_mag_std
t_body_gyro_mag_mean
t_body_gyro_mag_std
t_body_gyro_jerk_mag_mean
t_body_gyro_jerk_mag_std
f_body_acc_mean_x
f_body_acc_mean_y
f_body_acc_mean_z
f_body_acc_std_x
f_body_acc_std_y
f_body_acc_std_z
f_body_acc_jerk_mean_x
f_body_acc_jerk_mean_y
f_body_acc_jerk_mean_z
f_body_acc_jerk_std_x
f_body_acc_jerk_std_y
f_body_acc_jerk_std_z

f_body_gyro_mean_x
f_body_gyro_mean_y
f_body_gyro_mean_z
f_body_gyro_std_x
f_body_gyro_std_y
f_body_gyro_std_z
f_body_acc_mag_mean
f_body_acc_mag_std
f_body_body_acc_jerk_mag_mean
f_body_body_acc_jerk_mag_std
f_body_body_gyro_mag_mean
f_body_body_gyro_mag_std
f_body_body_gyro_jerk_mag_mean
f_body_body_gyro_jerk_mag_std