Getting and Cleaning Data Project

One of the most exciting areas in all of data science right now is wearable computing. Companies like Fitbit, Nike, and Jawbone Up are racing to develop the most advanced algorithms to attract new users. sdf

In this project, data collected from the accelerometer and gyroscope of the Samsung Galaxy S smartphone was retrieved, worked with, and cleaned, to prepare a tidy data that can be used for later analysis.

This repository contains the following files: - README.pdf, this file, that gives an overview of the project - tidy_dataset.txt , containing the data set - Codebook.pdf, the codebook that describes the contents of the data set - run_analysis.R, the R Script that was used to create the data set

The Project

The data was obtained from The Human Activity Recognition Using Smartphones Data Set, which describes how data was collected:

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.

Training and test datasets were merged, and then the mean and standard deviation measurements were extracted, after which they were averaged for each activity and subject to create the final dataset.

Dataset creation

run_analysis.R was the R script used to create the data set and it came be found in the same repository as this README.pdf. - The R script reads the data, whilst renaming columns based on features

- Training and test were merged
- Mean and standard deviation containing columns were alone extracted
- Feature names were used to describe variables
- A new tidy dataset was created with average of each variable grouped by activity label and subject
- Data was written to tidy_dataset.txt

R version 3.6.1 (2019-07-05) – "Action of the Toes" was used to create this on MacOS Catalina. This script requires the dplyr package to run.