Tidy Dataset Code Book

Information Source

This source dataset is obtained from the "Human Activity Recognition Using Smartphones Data Set" from the web location at
http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

Data Processing

The downloaded data is then processed to obtain a tidy data set. The tidy data set is named based on different attributes in data. The data in general is about the data captured by accelerometer and gyrometer sensors from the Samsung mobile phones tied to 30 subjects. 70% of subject's data was used for training and remaining 30% is used for testing. Both these data are available as seperate folders.

Feature Selection

The original codebook for the data is clear to great extent and the information from which is given below. Information is added wherever necessary or not clear. The following information can be obtained from readme.txt and features_info.txt files

The features 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. Then they were filtered using a median filter and a 3rd order low pass Butterworth filter with a corner frequency of 20 Hz to remove noise. 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).

Out of the total 561 features, data has been subsetted down to 79 features which contain the words Mean and STD in them. The objective of the exercise is to produce a tidy data set with average of the mean and standard deviation of all variable across 30 subjects and 6 activities.

Following list corresponds to the 79 features that remain from the different signals processed and filtered with the criteria of mean and standard deviation.

1 tBodyAcc.mean...X

2 tBodyAcc.mean...Y

3 tBodyAcc.mean...Z

4 tGravityAcc.mean...X

5 tGravityAcc.mean...Y

6 tGravityAcc.mean...Z

7 tBodyAccJerk.mean...X

8 tBodyAccJerk.mean...Y

9 tBodyAccJerk.mean...Z

10 tBodyGyro.mean...X

11 tBodyGyro.mean...Y

12 tBodyGyro.mean...Z

13 tBodyGyroJerk.mean...X

14 tBodyGyroJerk.mean...Y

15 tBodyGyroJerk.mean...Z

16 tBodyAccMag.mean..

17 tGravityAccMag.mean..

18 tBodyAccJerkMag.mean..

19 tBodyGyroMag.mean..

20 tBodyGyroJerkMag.mean..

21 fBodyAcc.mean...X

22 fBodyAcc.mean...Y

23 fBodyAcc.mean...Z

24 fBodyAcc.meanFreq...X

25 fBodyAcc.meanFreq...Y

26 fBodyAcc.meanFreq...Z

27 fBodyAccJerk.mean...X

28 fBodyAccJerk.mean...Y

29 fBodyAccJerk.mean...Z

30 fBodyAccJerk.meanFreq...X

31 fBodyAccJerk.meanFreq...Y

32 fBodyAccJerk.meanFreq...Z

33 fBodyGyro.mean...X

34 fBodyGyro.mean...Y

35 fBodyGyro.mean...Z

36 fBodyGyro.meanFreq...X

37 fBodyGyro.meanFreq...Y

38 fBodyGyro.meanFreq...Z

39 fBodyAccMag.mean..

40 fBodyAccMag.meanFreq..

41 fBodyBodyAccJerkMag.mean..

42 fBodyBodyAccJerkMag.meanFreq..

43 fBodyBodyGyroMag.mean..

44 fBodyBodyGyroMag.meanFreq..

45 fBodyBodyGyroJerkMag.mean..

46 fBodyBodyGyroJerkMag.meanFreq..

47 tBodyAcc.std...X

48 tBodyAcc.std...Y

49 tBodyAcc.std...Z

50 tGravityAcc.std...X

51 tGravityAcc.std...Y

52 tGravityAcc.std...Z

53 tBodyAccJerk.std...X

54 tBodyAccJerk.std...Y

55 tBodyAccJerk.std...Z

56 tBodyGyro.std...X

57 tBodyGyro.std...Y

58 tBodyGyro.std...Z

59 tBodyGyroJerk.std...X

60 tBodyGyroJerk.std...Y

61 tBodyGyroJerk.std...Z

62 tBodyAccMag.std..

63 tGravityAccMag.std..

64 tBodyAccJerkMag.std..

65 tBodyGyroMag.std..

66 tBodyGyroJerkMag.std..

67 fBodyAcc.std...X

68 fBodyAcc.std...Y

69 fBodyAcc.std...Z

70 fBodyAccJerk.std...X

71 fBodyAccJerk.std...Y

72 fBodyAccJerk.std...Z

73 fBodyGyro.std...X

74 fBodyGyro.std...Y

75 fBodyGyro.std...Z

76 fBodyAccMag.std..

77 fBodyBodyAccJerkMag.std..

78 fBodyBodyGyroMag.std..

79 fBodyBodyGyroJerkMag.std..

Explanation of Features

- t or f at the start of the name denotes "time domain" or "frequency domain".
- The term "body" or "Gravity" denotes the origin of acceleration into the sensor.
- The following term "Acc" or "Gyro" denotes the sensor which picked up the information
- All the terms following this denotes derived quantity
    - Jerk denotes the acceleration due to Jerk from body (in mm/s^2)
    - Mag. denotes magnitude of acceleration derived from the X,Y and Z components, in mm/s^2(normalized)
    - mean and std denotes calculated mean and standard deviation of the data
    - XYZ denotes the direction from which the acceleration or velocity is calculated
    - Freq refers to frequency estimated from the data, in Hz(normalized)

Other estimates have been removed for the purpose of this excercise.