COURSERA – Data Science Specialization – John Hopkins University

Getting and Cleaning Data Course Project

CODEBOOK

INTRODUCTION

The University of California, Irvine (UCI) conducted Human Activity Recognition (HAR) experiments using a specific type of Samsung smartphone, and published the resultant dataset in their Machine Learning Repository. Their dataset, namely https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip, is well described in a set of text files (specifically README.txt, feature_info.txt and feature.txt) in the root folder (UCI HAR dataset) of the .zip file, as well as here: http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones.

This CODEBOOK describes the tidy dataset that has been created from the dataset mentioned in the paragraph above, for the purposes of the project mentioned in the header above.

CODEBOOK

The tidy dataset contains information for 180 unique combinations of 6 activities and 30 subjects. The activities are WALKING, WALKING UPSTAIRS, WALKING DOWNSTAIRS, SITTING, STANDING and LAYING. The subjects (volunteers who participated in the experiment) are numbered from 1 through to 30.

The information for each unique combination of the 2 key fields comprises 66 means of 10,299 observations of means and standard deviations that have been extracted from the original UCI HAR dataset. The total of 10,299 observations represents a merge that has been done of 7,352 (70%) training observations and 2,947 (30%) test observations (subjects were randomly split 70% / 30%).

Specifically, each activity and subject combination is represented by a vector of the following:

Key fields

Note there are 180 = 6 X 30 unique combinations:

1. **Activity** (WALKING, WALKING UPSTAIRS, WALKING DOWNSTAIRS, SITTING, STANDING and LAYING)
2. **Subject** (range 1 to 30)

Time domain variables (with units of measurement in brackets)

Note the mean and standard deviations mentioned in this section were originally calculated based on raw signal vectors provided under "test/Inertial Signals" and "train/Inertial Signals" of the UCI HAR dataset. The raw signal data had been obtained using the embedded accelerometer and gyroscope of the Samsung Galaxy SII. The accelerometer measures body acceleration in standard gravity units "g" (gravity of earth -> 9.80665 m/seg) while the gyroscope measures angular velocity in radians/second. X, Y and Z denote tri-axial directions of measurement.

3. **tBodyAcc-mean**-X (g)
4. tBodyAcc-mean-Y (g)
5. tBodyAcc-mean-Z (g)
6. **tBodyAcc-stdev**-X (g)
7. tBodyAcc-stdev-Y (g)
8. tBodyAcc-stdev-Z (g)

9. **tGravityAcc-mean**-X (g)
10. tGravityAcc-mean-Y (g)
11. tGravityAcc-mean-Z (g)
12. **tGravityAcc-stdev**-X (g)
13. tGravityAcc-stdev-Y (g)
14. tGravityAcc-stdev-Z (g)

15. **tBodyAccJerk-mean**-X (g)
16. tBodyAccJerk-mean-Y (g)
17. tBodyAccJerk-mean-Z (g)
18. **tBodyAccJerk-stdev**-X (g)
19. tBodyAccJerk-stdev-Y (g)
20. tBodyAccJerk-stdev-Z (g)

21. **tBodyGyro-mean**-X (radians/second)
22. tBodyGyro-mean-Y (radians/second)
23. tBodyGyro-mean-Z (radians/second)
24. **tBodyGyro-stdev**-X (radians/second)
25. tBodyGyro-stdev-Y (radians/second)
26. tBodyGyro-stdev-Z (radians/second)

27. **tBodyGyroJerk-mean**-X (radians/second)
28. tBodyGyroJerk-mean-Y (radians/second)
29. tBodyGyroJerk-mean-Z (radians/second)
30. **tBodyGyroJerk-stdev**-X (radians/second)
31. tBodyGyroJerk-stdev-Y (radians/second)
32. tBodyGyroJerk-stdev-Z (radians/second)

33. **tBodyAccMag-mean** (g)
34. **tBodyAccMag-stdev** (g)

35. **tGravityAccMag-mean** (g)
36. **tGravityAccMag-stdev** (g)

37. **tBodyAccJerkMag-mean** (g)
38. **tBodyAccJerkMag-stdev** (g)

39. **tBodyGyroMag-mean** (radians/second)
40. **tBodyGyroMag-stdev** (radians/second)

41. **tBodyGyroJerkMag-mean** (radians/second)
42. **tBodyGyroJerkMag-stdev** (radians/second)

<u>Frequency domain variables (with units of measurement in brackets)</u>

A Fast Fourier Transform (FFT) was applied to some of the raw signals. Regarding the units after FFT, refer https://dsp.stackexchange.com/questions/35992/what-are-the-units-of-my-data-after-an-fft – specifically, "The units after FFT remain the same as for the signal".

43. **fBodyAcc-mean**-X (g)
44. fBodyAcc-mean-Y (g)
45. fBodyAcc-mean-Z (g)
46. **fBodyAcc-stdev**-X (g)
47. fBodyAcc-stdev-Y (g)
48. fBodyAcc-stdev-Z (g)

49. **fBodyAccJerk-mean**-X (g)
50. fBodyAccJerk-mean-Y (g)
51. fBodyAccJerk-mean-Z (g)
52. **fBodyAccJerk-stdev**-X (g)
53. fBodyAccJerk-stdev-Y (g)
54. fBodyAccJerk-stdev-Z (g)

55. **fBodyGyro-mean**-X (radians/second)
56. fBodyGyro-mean-Y (radians/second)
57. fBodyGyro-mean-Z (radians/second)
58. **fBodyGyro-stdev**-X (radians/second)
59. fBodyGyro-stdev-Y (radians/second)
60. fBodyGyro-stdev-Z (radians/second)

61. **fBodyAccMag-mean** (g)
62. **fBodyAccMag-stdev** (g)

63. **fBodyAccJerkMag-mean** (g)
64. **fBodyAccJerkMag-stdev** (g)

65. **fBodyGyroMag-mean** (radians/second)
66. **fBodyGyroMag-stdev** (radians/second)

67. **tBodyGyroJerkMag-mean** (radians/second)
68. **tBodyGyroJerkMag-stdev** (radians/second)

THE END.