Code Book

Stephen Aboagye-Ntow

9/25/2020

Methodology

Thirty (30) subjects volunteered to participate in an experiment which involves each subject performing six **activities** each while wearing a smartphone (Samsung Galaxy S II) on the waist

Activities

```
##                    . X1.6
## 1            walking    1
## 2   walking_upstairs    2
## 3 walking_downstairs    3
## 4            sitting    4
## 5           standing    5
## 6             laying    6
```

Signals

Using its embedded accelerometer and gyroscope, the samsung galaxy SII was used to captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz and the following **signals** were generated

```
##                    .
## 1       tBodyAcc-XYZ
## 2    tGravityAcc-XYZ
## 3   tBodyAccJerk-XYZ
## 4      tBodyGyro-XYZ
## 5  tBodyGyroJerk-XYZ
## 6        tBodyAccMag
## 7     tGravityAccMag
## 8    tBodyAccJerkMag
## 9       tBodyGyroMag
## 10  tBodyGyroJerkMag
## 11      fBodyAcc-XYZ
## 12  fBodyAccJerk-XYZ
## 13     fBodyGyro-XYZ
## 14       fBodyAccMag
## 15   fBodyAccJerkMag
## 16      fBodyGyroMag
## 17  fBodyGyroJerkMag
```

Variables

The set of variables that were estimated from the **signals** are:

```
##                                                                                 variables
## 1                                                                      mean(): Mean value
## 2                                                                  std(): Standard deviation
## 3                                                          mad(): Median absolute deviation
## 4                                                             max(): Largest value in array
## 5                                                           min(): Smallest value in array
## 6                                                             sma(): Signal magnitude area
## 7                 energy(): Energy measure. Sum of the squares divided by the number of values
## 8                                                                 iqr(): Interquartile range
## 9                                                                  entropy(): Signal entropy
## 10                           arCoeff(): Autorregresion coefficients with Burg order equal to 4
## 11                                  correlation(): correlation coefficient between two signals
## 12                              maxInds(): index of the frequency component with largest magnitude
## 13         meanFreq(): Weighted average of the frequency components to obtain a mean frequency
## 14                                           skewness(): skewness of the frequency domain signal
## 15                                           kurtosis(): kurtosis of the frequency domain signal
## 16 bandsEnergy(): Energy of a frequency interval within the 64 bins of the FFT of each window
## 17                                                  angle(): Angle between to vectors
```

features

For each **Signal**, **Variables** was generated. These are the features or measurements of each activity for each subject.

First 20 elements of features

```
##  [1] "1 tBodyAcc-mean()-X"    "2 tBodyAcc-mean()-Y"    "3 tBodyAcc-mean()-Z"
##  [4] "4 tBodyAcc-std()-X"     "5 tBodyAcc-std()-Y"     "6 tBodyAcc-std()-Z"
##  [7] "7 tBodyAcc-mad()-X"     "8 tBodyAcc-mad()-Y"     "9 tBodyAcc-mad()-Z"
## [10] "10 tBodyAcc-max()-X"    "11 tBodyAcc-max()-Y"    "12 tBodyAcc-max()-Z"
## [13] "13 tBodyAcc-min()-X"    "14 tBodyAcc-min()-Y"    "15 tBodyAcc-min()-Z"
## [16] "16 tBodyAcc-sma()"      "17 tBodyAcc-energy()-X" "18 tBodyAcc-energy()-Y"
## [19] "19 tBodyAcc-energy()-Z" "20 tBodyAcc-iqr()-X"
```

Data from Experiment

The obtained dataset from the excersice has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% for the test data

Train Data

First 20 rows and first 5 columns

```
##    Subjects Activity tBodyAcc-mean()-X tBodyAcc-mean()-Y tBodyAcc-mean()-Z
## 1         1        5             0.289          -0.02030           -0.1330
## 2         1        5             0.278          -0.01640           -0.1240
## 3         1        5             0.280          -0.01950           -0.1130
```

```
## 4         1        5             0.279          -0.02620          -0.1230
## 5         1        5             0.277          -0.01660          -0.1150
## 6         1        5             0.277          -0.01010          -0.1050
## 7         1        5             0.279          -0.01960          -0.1100
## 8         1        5             0.277          -0.03050          -0.1250
## 9         1        5             0.277          -0.02180          -0.1210
## 10        1        5             0.281          -0.00996          -0.1060
## 11        1        5             0.277          -0.01270          -0.1030
## 12        1        5             0.276          -0.02140          -0.1080
## 13        1        5             0.278          -0.02040          -0.1130
## 14        1        5             0.277          -0.01470          -0.1070
## 15        1        5             0.298           0.02710          -0.0617
## 16        1        5             0.279          -0.02300          -0.1220
## 17        1        5             0.279          -0.01480          -0.1170
## 18        1        5             0.280          -0.01390          -0.1060
## 19        1        5             0.278          -0.01820          -0.1090
## 20        1        5             0.276          -0.01700          -0.1110
```

Test Data

First 20 rows and first 5 columns

```
##    Subjects Activity tBodyAcc-mean()-X tBodyAcc-mean()-Y tBodyAcc-mean()-Z
## 1         2        5             0.257          -0.02330          -0.0147
## 2         2        5             0.286          -0.01320          -0.1190
## 3         2        5             0.275          -0.02610          -0.1180
## 4         2        5             0.270          -0.03260          -0.1180
## 5         2        5             0.275          -0.02780          -0.1300
## 6         2        5             0.279          -0.01860          -0.1140
## 7         2        5             0.280          -0.01830          -0.1040
## 8         2        5             0.275          -0.02500          -0.1170
## 9         2        5             0.273          -0.02100          -0.1140
## 10        2        5             0.276          -0.01040          -0.0998
## 11        2        5             0.279          -0.01520          -0.0989
## 12        2        5             0.279          -0.02190          -0.1100
## 13        2        5             0.275          -0.02310          -0.1130
## 14        2        5             0.269          -0.02770          -0.1100
## 15        2        5             0.276          -0.01890          -0.0974
## 16        2        5             0.282          -0.00488          -0.0861
## 17        2        5             0.311          -0.01940          -0.1020
## 18        2        5             0.262          -0.02330          -0.1260
## 19        2        5             0.288          -0.00349          -0.0838
## 20        2        5             0.271          -0.02600          -0.0949
```

To have a full table of observations from the excercise, both test and train sets must be merged into a one.

First 20 rows and first 5 columns

```
##    Subjects Activity tBodyAcc-mean()-X tBodyAcc-mean()-Y tBodyAcc-mean()-Z
## 1         1        5             0.289          -0.02030          -0.1330
## 2         1        5             0.278          -0.01640          -0.1240
## 3         1        5             0.280          -0.01950          -0.1130
## 4         1        5             0.279          -0.02620          -0.1230
```

```
## 5         1        5              0.277          -0.01660          -0.1150
## 6         1        5              0.277          -0.01010          -0.1050
## 7         1        5              0.279          -0.01960          -0.1100
## 8         1        5              0.277          -0.03050          -0.1250
## 9         1        5              0.277          -0.02180          -0.1210
## 10        1        5              0.281          -0.00996          -0.1060
## 11        1        5              0.277          -0.01270          -0.1030
## 12        1        5              0.276          -0.02140          -0.1080
## 13        1        5              0.278          -0.02040          -0.1130
## 14        1        5              0.277          -0.01470          -0.1070
## 15        1        5              0.298           0.02710          -0.0617
## 16        1        5              0.279          -0.02300          -0.1220
## 17        1        5              0.279          -0.01480          -0.1170
## 18        1        5              0.280          -0.01390          -0.1060
## 19        1        5              0.278          -0.01820          -0.1090
## 20        1        5              0.276          -0.01700          -0.1110
```

To communicate more meaning, the numeric values of the **Activity** column must be replace with their corresponding character variables as indicated in the vector: walking, walking_upstairs, walking_downstairs, sitting, standing, laying

Now we replace the numerics with characters from the above code

First 20 rows and first 5 columns

```
##    Subjects Activity tBodyAcc-mean()-X tBodyAcc-mean()-Y tBodyAcc-mean()-Z
## 1         1 standing              0.289          -0.02030           -0.1330
## 2         1 standing              0.278          -0.01640           -0.1240
## 3         1 standing              0.280          -0.01950           -0.1130
## 4         1 standing              0.279          -0.02620           -0.1230
## 5         1 standing              0.277          -0.01660           -0.1150
## 6         1 standing              0.277          -0.01010           -0.1050
## 7         1 standing              0.279          -0.01960           -0.1100
## 8         1 standing              0.277          -0.03050           -0.1250
## 9         1 standing              0.277          -0.02180           -0.1210
## 10        1 standing              0.281          -0.00996           -0.1060
## 11        1 standing              0.277          -0.01270           -0.1030
## 12        1 standing              0.276          -0.02140           -0.1080
## 13        1 standing              0.278          -0.02040           -0.1130
## 14        1 standing              0.277          -0.01470           -0.1070
## 15        1 standing              0.298           0.02710           -0.0617
## 16        1 standing              0.279          -0.02300           -0.1220
## 17        1 standing              0.279          -0.01480           -0.1170
## 18        1 standing              0.280          -0.01390           -0.1060
## 19        1 standing              0.278          -0.01820           -0.1090
## 20        1 standing              0.276          -0.01700           -0.1110
```

Since the estimated variables of interest are only mean and std, we subset only features or measurements with mean() and std()

First 20 rows and first 5 columns

```
##    Subjects Activity tBodyAcc.mean...X tBodyAcc.mean...Y tBodyAcc.mean...Z
## 1         1 standing              0.289          -0.02030           -0.1330
## 2         1 standing              0.278          -0.01640           -0.1240
```

```
## 3         1 standing             0.280          -0.01950           -0.1130
## 4         1 standing             0.279          -0.02620           -0.1230
## 5         1 standing             0.277          -0.01660           -0.1150
## 6         1 standing             0.277          -0.01010           -0.1050
## 7         1 standing             0.279          -0.01960           -0.1100
## 8         1 standing             0.277          -0.03050           -0.1250
## 9         1 standing             0.277          -0.02180           -0.1210
## 10        1 standing             0.281          -0.00996           -0.1060
## 11        1 standing             0.277          -0.01270           -0.1030
## 12        1 standing             0.276          -0.02140           -0.1080
## 13        1 standing             0.278          -0.02040           -0.1130
## 14        1 standing             0.277          -0.01470           -0.1070
## 15        1 standing             0.298           0.02710           -0.0617
## 16        1 standing             0.279          -0.02300           -0.1220
## 17        1 standing             0.279          -0.01480           -0.1170
## 18        1 standing             0.280          -0.01390           -0.1060
## 19        1 standing             0.278          -0.01820           -0.1090
## 20        1 standing             0.276          -0.01700           -0.1110
```

Ordering

The data is then ordered by the subjects variable in a descending order of magnitude

First 5 rows and columns and Last 5 rows and columns

```
##   Subjects Activity tBodyAcc.mean...X tBodyAcc.mean...Y tBodyAcc.mean...Z
## 1        1 standing             0.289           -0.0203            -0.133
## 2        1 standing             0.278           -0.0164            -0.124
## 3        1 standing             0.280           -0.0195            -0.113
## 4        1 standing             0.279           -0.0262            -0.123
## 5        1 standing             0.277           -0.0166            -0.115
```

```
##      Subjects         Activity tBodyAcc.mean...X tBodyAcc.mean...Y
## 7347       30 Walking_upstairs             0.238          -0.00109
## 7348       30 Walking_upstairs             0.300          -0.05720
## 7349       30 Walking_upstairs             0.274          -0.00775
## 7350       30 Walking_upstairs             0.273          -0.01700
## 7351       30 Walking_upstairs             0.290          -0.01880
## 7352       30 Walking_upstairs             0.352          -0.01240
##      tBodyAcc.mean...Z
## 7347            -0.148
## 7348            -0.181
## 7349            -0.147
## 7350            -0.045
## 7351            -0.158
## 7352            -0.204
```

The data was then structured such that each variable is in its own column and each observation in a row to conform to the "so called tidy data" philosophy introduced by Hadley Wickham.

First 20 rows

```
##    Subjects Activity       measurement value
```

```
## 1         1 standing tBodyAcc.mean...X 0.289
## 2         1 standing tBodyAcc.mean...X 0.278
## 3         1 standing tBodyAcc.mean...X 0.280
## 4         1 standing tBodyAcc.mean...X 0.279
## 5         1 standing tBodyAcc.mean...X 0.277
## 6         1 standing tBodyAcc.mean...X 0.277
## 7         1 standing tBodyAcc.mean...X 0.279
## 8         1 standing tBodyAcc.mean...X 0.277
## 9         1 standing tBodyAcc.mean...X 0.277
## 10        1 standing tBodyAcc.mean...X 0.281
## 11        1 standing tBodyAcc.mean...X 0.277
## 12        1 standing tBodyAcc.mean...X 0.276
## 13        1 standing tBodyAcc.mean...X 0.278
## 14        1 standing tBodyAcc.mean...X 0.277
## 15        1 standing tBodyAcc.mean...X 0.298
## 16        1 standing tBodyAcc.mean...X 0.279
## 17        1 standing tBodyAcc.mean...X 0.279
## 18        1 standing tBodyAcc.mean...X 0.280
## 19        1 standing tBodyAcc.mean...X 0.278
## 20        1 standing tBodyAcc.mean...X 0.276
```

The dataset was then divided into groups, thus into separate dataframes based on the unique variable amongst the measurements which are:

- mean()
- std()

First and last 5 rows of mean() inherent measurements

```
##   Subjects Activity       measurement value
## 1        1 standing tBodyAcc.mean...X 0.289
## 2        1 standing tBodyAcc.mean...X 0.278
## 3        1 standing tBodyAcc.mean...X 0.280
## 4        1 standing tBodyAcc.mean...X 0.279
## 5        1 standing tBodyAcc.mean...X 0.277
```

```
##        Subjects         Activity                 measurement  value
## 669430       30 Walking_upstairs fBodyBodyGyroJerkMag.mean.. -0.706
## 669431       30 Walking_upstairs fBodyBodyGyroJerkMag.mean.. -0.681
## 669432       30 Walking_upstairs fBodyBodyGyroJerkMag.mean.. -0.683
## 669433       30 Walking_upstairs fBodyBodyGyroJerkMag.mean.. -0.686
## 669434       30 Walking_upstairs fBodyBodyGyroJerkMag.mean.. -0.712
## 669435       30 Walking_upstairs fBodyBodyGyroJerkMag.mean.. -0.716
```

First and Last 5 rows of std() inherent measurements

```
##       Subjects Activity      measurement  value
## 30898        1 standing tBodyAcc.std...X -0.995
## 30899        1 standing tBodyAcc.std...X -0.998
## 30900        1 standing tBodyAcc.std...X -0.995
## 30901        1 standing tBodyAcc.std...X -0.996
## 30902        1 standing tBodyAcc.std...X -0.998
```

```
##        Subjects         Activity                measurement  value
## 679729       30 Walking_upstairs fBodyBodyGyroJerkMag.std.. -0.755
## 679730       30 Walking_upstairs fBodyBodyGyroJerkMag.std.. -0.724
## 679731       30 Walking_upstairs fBodyBodyGyroJerkMag.std.. -0.771
## 679732       30 Walking_upstairs fBodyBodyGyroJerkMag.std.. -0.726
## 679733       30 Walking_upstairs fBodyBodyGyroJerkMag.std.. -0.689
## 679734       30 Walking_upstairs fBodyBodyGyroJerkMag.std.. -0.745
```

Grouped analysis

The average of the mean() measurements as well as the average of the std() measurements for every activity of each subject in the immediate-above dataframes.

First 5 rows of Grouped average of mean() inherent measurements

```
##   Subjects Activity       value
## 1        1  walking -0.08385119
## 2        2  walking -0.17936829
## 3        3  walking -0.18490108
## 4        4  walking -0.22743638
## 5        5  walking -0.15513715
```

First 5 rows of Grouped average of std() inherent measurements

```
##     Subjects Activity      value
## 176       26   laying -0.9806527
## 177       27   laying -0.9850152
## 178       28   laying -0.9642508
## 179       29   laying -0.9885946
## 180       30   laying -0.9713173
```

Resultant dataframes from immediate-above are merged ones again by row to achieve the data format which adheres to the so called "tidy data" philosophy and ordered afterwards using the subject variable or column to communicate more meaning and create ease for further analysis on the data.

First 10 rows and Last 10 rows

```
##    Subjects           Activity Measurement     Average
## 1         1            walking     mean_av -0.08385119
## 2         1   Walking_upstairs     mean_av -0.20225106
## 3         1 walking_downstairs     mean_av -0.03499685
## 4         1            sitting     mean_av -0.48749654
## 5         1           standing     mean_av -0.51696667
## 6         1             laying     mean_av -0.46646191
## 7         1            walking      std_av -0.30255112
## 8         1   Walking_upstairs      std_av -0.42843716
## 9         1 walking_downstairs      std_av -0.26371548
## 10        1            sitting      std_av -0.96253256
```

```
##     Subjects           Activity Measurement    Average
## 351       30 walking_downstairs     mean_av -0.1024438
## 352       30            sitting     mean_av -0.4924508
## 353       30           standing     mean_av -0.4871227
```

```
## 354       30             laying     mean_av -0.4961378
## 355       30            walking      std_av -0.3881271
## 356       30   Walking_upstairs      std_av -0.4565593
## 357       30 walking_downstairs      std_av -0.3325921
## 358       30            sitting      std_av -0.9716808
## 359       30           standing      std_av -0.9414391
## 360       30             laying      std_av -0.9713173
```

Data Dictionary

Subjects

Volunteered persons for the study :

- 1 - subject1
- 2 - subject2
- 3 - subject3
- 4 - subject4
- 5 - subject5
- 6 - subject6
- 7 - subject7
- 8 - subject8
- 9 - subject9
- 10 - subject10
- 11 - subject11
- 12 - subject12
- 13 - subject13
- 14 - subject14
- 15 - subject15
- 16 - subject16
- 17 - subject17
- 18 - subject18
- 19 - subject19

- 20 - subject20
- 21 - subject21
- 22 - subject22
- 23 - subject23
- 24 - subject24
- 25 - subject25
- 26 - subject26
- 27 - subject27
- 28 - subject28
- 29 - subject29
- 30 - subject30

Activities

Each Subject undertook six activities:

- 1.walking
- 2.walking_upstairs
- 3.walking_downstairs
- 4.sitting
- 5.standing
- 6.laying

Measurements

Of the 17 variables estimated from the signals, the analysis only required the computation of the average of only the measurements or features with mean() and std() variables inherent in them for each **Activity** and each **Subject**.

```
mean_av - representing all mean() inherent measurements for each activity and
std_av - representing all std() inherent measurements for each activity and
```

Average

Average value of mean() or std() inherent measurements for each activity and for each subject