GDC Project Data Set – Codebook

**Frank**

**09/13/2014**

========================================== **THE FIRST TWO VARIABLES / COLUMNS**

SUBJECT

Integer A numeric ID that indicates the subject who participated in the experiment. These were taken from the files subject\_test.txt and subject\_train.txt in the original UCI data set.

ACTIVITIES

Factor with 6 levels The activity performed by the SUBJECT. One of the following six activities: WALKING WALKING\_UPSTAIRS WALKING\_DOWNSTAIRS SITTING STANDING LAYING

The data for this variable was built using data from the files y\_test.txt, y\_train.txt and activity\_labels.txt int he original UCI data set.

==========================================

**THE REMAINING VARIABLES IN THE GDC PROJECT DATA SET**

The remaining variables in the GDC PROJECT DATA SET utilize the following measurements from the UCI experiment (a description of these measurements can be found at the end of this document in the section, "ABOUT THE MEASUREMENTS FROM THE ORIGINAL UCI DATA SET"):

tBodyAcc-XYZ  
tGravityAcc-XYZ  
tBodyAccJerk-XYZ  
tBodyGyro-XYZ  
tBodyGyroJerk-XYZ  
tBodyAccMag  
tGravityAccMag  
tBodyAccJerkMag  
tBodyGyroMag  
tBodyGyroJerkMag  
fBodyAcc-XYZ  
fBodyAccJerk-XYZ  
fBodyGyro-XYZ  
fBodyAccMag  
fBodyAccJerkMag  
fBodyGyroMag  
fBodyGyroJerkMag

For each of the above measurement with the suffix "-XYZ," three measurements were taken, one for each axis (X, Y, Z). The variables in the original UCI data set were generated by performing the following calculations on each measurement:

mean(): Mean value  
std(): Standard deviation

This data set contains the AVERAGE of all of the mean and standard deviation measurements from the original UCI data set.

==========================================

**COMPLETE LIST OF VARIABLES IN THE GDC PROJECT DATA SET**

Each of the following variables appear in the GDC PROJECT DATA SET. With the exception of "SUBJECT" and "ACTIVITIES," all are numeric and contain the AVERAGE of each variable for each SUBJECT and ACTIVITY.

SUBJECT

ACTIVITIES

tBodyAcc-mean()-X

tBodyAcc-mean()-Y

tBodyAcc-mean()-Z

tGravityAcc-mean()-X

tGravityAcc-mean()-Y

tGravityAcc-mean()-Z

tBodyAccJerk-mean()-X

tBodyAccJerk-mean()-Y

tBodyAccJerk-mean()-Z

tBodyGyro-mean()-X

tBodyGyro-mean()-Y

tBodyGyro-mean()-Z

tBodyGyroJerk-mean()-X

tBodyGyroJerk-mean()-Y

tBodyGyroJerk-mean()-Z

tBodyAccMag-mean()

tGravityAccMag-mean()

tBodyAccJerkMag-mean()

tBodyGyroMag-mean()

tBodyGyroJerkMag-mean()

fBodyAcc-mean()-X

fBodyAcc-mean()-Y

fBodyAcc-mean()-Z

fBodyAccJerk-mean()-X

fBodyAccJerk-mean()-Y

fBodyAccJerk-mean()-Z

fBodyGyro-mean()-X

fBodyGyro-mean()-Y

fBodyGyro-mean()-Z

fBodyAccMag-mean()

fBodyBodyAccJerkMag-mean()

fBodyBodyGyroMag-mean()

fBodyBodyGyroJerkMag-mean()

tBodyAcc-std()-X

tBodyAcc-std()-Y

tBodyAcc-std()-Z

tGravityAcc-std()-X

tGravityAcc-std()-Y

tGravityAcc-std()-Z

tBodyAccJerk-std()-X

tBodyAccJerk-std()-Y

tBodyAccJerk-std()-Z

tBodyGyro-std()-X

tBodyGyro-std()-Y

tBodyGyro-std()-Z

tBodyGyroJerk-std()-X

tBodyGyroJerk-std()-Y

tBodyGyroJerk-std()-Z

tBodyAccMag-std()

tGravityAccMag-std()

tBodyAccJerkMag-std()

tBodyGyroMag-std()

tBodyGyroJerkMag-std()

fBodyAcc-std()-X

fBodyAcc-std()-Y

fBodyAcc-std()-Z

fBodyAccJerk-std()-X

fBodyAccJerk-std()-Y

fBodyAccJerk-std()-Z

fBodyGyro-std()-X

fBodyGyro-std()-Y

fBodyGyro-std()-Z

fBodyAccMag-std()

fBodyBodyAccJerkMag-std()

fBodyBodyGyroMag-std()

fBodyBodyGyroJerkMag-std()

==========================================

**ABOUT THE MEASUREMENTS FROM THE ORIGINAL UCI DATA SET**

From the file "features\_info.txt" in the original UCI data set:

"The features [variables] 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).

These signals were used to estimate variables of the feature vector for each pattern:  
'-XYZ' is used to denote 3-axial signals in the X, Y and Z directions."