CodeBook

Study Design (per original features\_info.txt)

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 in the raw data) 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 in the raw data).

Compilation of Data

Per project instruction, the training data and test data are combined. First, the X pair, the Y pair and the subject pair are combined among themselves. There are 10299 observations, 561 features (X), 1 response variable Y and 1 subject variable. Feature and activity labels are then extracted and inserted into the combined X set and Y set, respectively. Then regular expression is used to select features that appear to be mean and standard deviation of measurements (excluding the ones related to angle.) 79 features are selected. The selected features, y set and subject set are then combined. Feature names are modified to clarify their meanings. Finally, the measurements are grouped by activity and subject jointly, creating 180 observations (from 30 participants, each showing 6 different activities.)

List of Variables

There are 81 variables. The first 79 variables are features. They are selected by using regular expression to look for variable names with the word “mean” and “std.” They are continuous, normalized and bounded within [-1,1]. For time related features, the unit is second. For frequency related features, the unit is Hz. For this project, these variables are renamed in a number of ways to add clarity:

1. The leading “f” is renamed “frequency.”
2. The leading “t” is renamed “time.”
3. “Acc” is renamed “Accelerometer.”
4. “Gyro” is renamed “Gyroscope.”
5. “Mag” is renamed “Magnitude.”
6. “BodyBody” is renamed “Body”
7. Any parentheses are removed.
8. Any dash is replaced with underscore.

The “Activity” variable has six values: WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTARS, SITTING and LAYING. The “Subject\_id” variable ranges from 1 to 30 and represents the thirty participants of the study.

The list of variables:

"TimeBodyAccelerometer\_mean\_X"

"TimeBodyAccelerometer\_mean\_Y"

"TimeBodyAccelerometer\_mean\_Z"

"TimeBodyAccelerometer\_std\_X"

"TimeBodyAccelerometer\_std\_Y"

"TimeBodyAccelerometer\_std\_Z"

"TimeGravityAccelerometer\_mean\_X"

"TimeGravityAccelerometer\_mean\_Y"

"TimeGravityAccelerometer\_mean\_Z"

"TimeGravityAccelerometer\_std\_X"

"TimeGravityAccelerometer\_std\_Y"

"TimeGravityAccelerometer\_std\_Z"

"TimeBodyAccelerometerJerk\_mean\_X"

"TimeBodyAccelerometerJerk\_mean\_Y"

"TimeBodyAccelerometerJerk\_mean\_Z"

"TimeBodyAccelerometerJerk\_std\_X"

"TimeBodyAccelerometerJerk\_std\_Y"

"TimeBodyAccelerometerJerk\_std\_Z"

"TimeBodyGyroscope\_mean\_X"

"TimeBodyGyroscope\_mean\_Y"

"TimeBodyGyroscope\_mean\_Z"

"TimeBodyGyroscope\_std\_X"

"TimeBodyGyroscope\_std\_Y"

"TimeBodyGyroscope\_std\_Z"

"TimeBodyGyroscopeJerk\_mean\_X"

"TimeBodyGyroscopeJerk\_mean\_Y"

"TimeBodyGyroscopeJerk\_mean\_Z"

"TimeBodyGyroscopeJerk\_std\_X"

"TimeBodyGyroscopeJerk\_std\_Y"

"TimeBodyGyroscopeJerk\_std\_Z"

"TimeBodyAccelerometerMagnitude\_mean"

"TimeBodyAccelerometerMagnitude\_std"

"TimeGravityAccelerometerMagnitude\_mean"

"TimeGravityAccelerometerMagnitude\_std"

"TimeBodyAccelerometerJerkMagnitude\_mean"

"TimeBodyAccelerometerJerkMagnitude\_std"

"TimeBodyGyroscopeMagnitude\_mean"

"TimeBodyGyroscopeMagnitude\_std"

"TimeBodyGyroscopeJerkMagnitude\_mean"

"TimeBodyGyroscopeJerkMagnitude\_std"

"FrequencyBodyAccelerometer\_mean\_X"

"FrequencyBodyAccelerometer\_mean\_Y"

"FrequencyBodyAccelerometer\_mean\_Z"

"FrequencyBodyAccelerometer\_std\_X"

"FrequencyBodyAccelerometer\_std\_Y"

"FrequencyBodyAccelerometer\_std\_Z"

"FrequencyBodyAccelerometer\_meanFreq\_X"

"FrequencyBodyAccelerometer\_meanFreq\_Y"

"FrequencyBodyAccelerometer\_meanFreq\_Z"

"FrequencyBodyAccelerometerJerk\_mean\_X"

"FrequencyBodyAccelerometerJerk\_mean\_Y"

"FrequencyBodyAccelerometerJerk\_mean\_Z"

"FrequencyBodyAccelerometerJerk\_std\_X"

"FrequencyBodyAccelerometerJerk\_std\_Y"

"FrequencyBodyAccelerometerJerk\_std\_Z"

"FrequencyBodyAccelerometerJerk\_meanFreq\_X"

"FrequencyBodyAccelerometerJerk\_meanFreq\_Y"

"FrequencyBodyAccelerometerJerk\_meanFreq\_Z"

"FrequencyBodyGyroscope\_mean\_X"

"FrequencyBodyGyroscope\_mean\_Y"

"FrequencyBodyGyroscope\_mean\_Z"

"FrequencyBodyGyroscope\_std\_X"

"FrequencyBodyGyroscope\_std\_Y"

"FrequencyBodyGyroscope\_std\_Z"

"FrequencyBodyGyroscope\_meanFreq\_X"

"FrequencyBodyGyroscope\_meanFreq\_Y"

"FrequencyBodyGyroscope\_meanFreq\_Z"

"FrequencyBodyAccelerometerMagnitude\_mean"

"FrequencyBodyAccelerometerMagnitude\_std"

"FrequencyBodyAccelerometerMagnitude\_meanFreq"

"FrequencyBodyAccelerometerJerkMagnitude\_mean"

"FrequencyBodyAccelerometerJerkMagnitude\_std"

"FrequencyBodyAccelerometerJerkMagnitude\_meanFreq"

"FrequencyBodyGyroscopeMagnitude\_mean"

"FrequencyBodyGyroscopeMagnitude\_std"

"FrequencyBodyGyroscopeMagnitude\_meanFreq"

"FrequencyBodyGyroscopeJerkMagnitude\_mean"

"FrequencyBodyGyroscopeJerkMagnitude\_std"

"FrequencyBodyGyroscopeJerkMagnitude\_meanFreq"

"Activity"

"Subject\_id"