CS 6762
SP, ML, FC
Programming Assignment 2

General: The accuracy results you obtained in programming assignment 1 may not have been very high (somewhat dependent on how many confounding gestures you tested with). The second assignment *builds upon the first assignment* to demonstrate the art of tuning parameters, adding features, investigating alternative classifiers, and performing comparison evaluations to improve accuracy. To be more efficient, instead of using the Wada tool and the Weka GUI (which are quite manual), you will write a program (with our help) to convert from *csv* to *arrf* and to perform classification. Details, including example code, are provided in the tutorial. However, you should continue to use the Wada tool to extract the accelerometer data from the .wada files (exactly as you did in assignment 1).

Assignment 2: This assignment has five parts.

Part One: Collect additional data. Accuracy often improves with more training data. Retrain using the data from both assignments using a decision tree. However, instead of using Wada and the Weka GUI, your program should perform the creation of the arrf files and access Weka via code in a library (see the tutorial for help with these 2 aspects). This will set up your program for the remaining parts of this assignment. Show the difference of using more data as your first result. Did the accuracy improve?

Part Two: Adjust the time slices used from 1 sec to also consider 2, 3 and 4 seconds. Use a sliding window of 1 sec for the 2, 3 and 4 second tests. Rerun the decision tree solution. Use the increased training set from part one. Report the results. Did the accuracy improve?

Part Three: Add additional features (median and root mean square) to the feature set. Now, there are four features per axis, resulting in 12 features in total. In this step choose the best time interval found from Part Two. Rerun the decision tree solution and report the results. Did the accuracy improve?

Part Four: Choose the best time interval from Part Two. Now, find the best feature set using the Sequential Feature Selection method (described in the associated tutorial) and a decision tree classifier. Which features are selected? Does accuracy improve compared to Part Three? Also report how accuracy is changed as features are added.

Part Five: Repeat Part Four for Random Forest and SVM classification algorithms (see the tutorial on how to use these Weka classifiers). For each classifier, report the set of features selected, and the change of accuracy as a feature is selected. Out of the three classifiers, which was the best?

Submission

Submit all the followings within a zipped folder named "Assignment 2 StudentName."
Use same name as used for assignment 1.

1. Features.csv file
2. A report containing the results of each of the 5 parts as requested above.