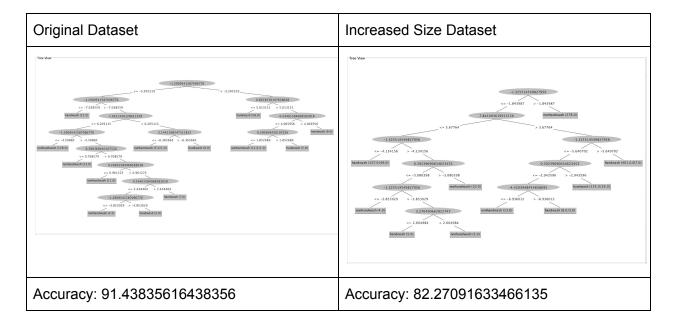
CS 6501Fall 2019 **SP, ML, FC**

Programming Assignment 2
Due: Oct. 22, 2019

Part 1: Using More Data

Originally, we had 20 non-hand washing samples and 20 hand washing samples. We collected an additional 20 non-hand washing samples and 40 hand washing samples. When we ran this through our program, it resulted in a significant decrease in accuracy, down to near 60%. We then visually inspected the data and found that many of the original hand washing recordings greatly differed from the new hand washing data we collected. To our recollection, we did not change the procedure for hand washing. However, with this discrepancy, we elected to not include the original handwashing data and instead only include the 40 new hand washing samples. This increased the accuracy to around 80%. Although this is worse than our original 90% accuracy, we feel that this data is more representative. It could be that the original 90% accuracy was due to the smaller sample size, thus making it easier to fit the data.

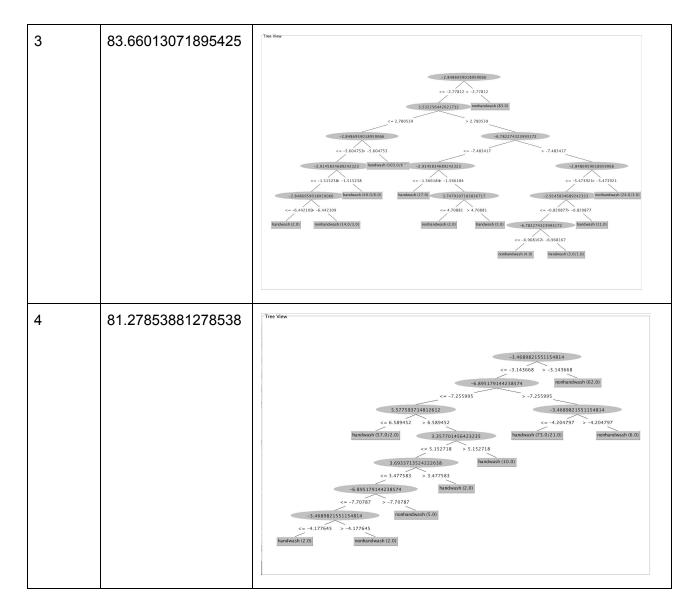


Part 2: Changing Sliding Window

The results are reported below. We found that the sliding window of 1 and 2 seconds had very similar accuracies, differing by .01 percentage points. Sliding window 3 provided increased accuracy, going up by 1.4 percentage points. However, Sliding window 4 was the worst of all, with accuracy 2.4 percentage points lower than window 3. We note that interestingly, although window 1 and window 2 provide very similar accuracy, the tree for window 2 is much simpler.

Results:

Sliding Window Size (Sec)	Accuracy	Tree
1	82.27091633466135	Tree View 1.125514549827956 <1.843987 7.441093615932558
2	82.26804123711341	-2.378031027326425 <= -2.997957 > -2.997957 handwash (345.0/79.0) nonhandwash (140.0)

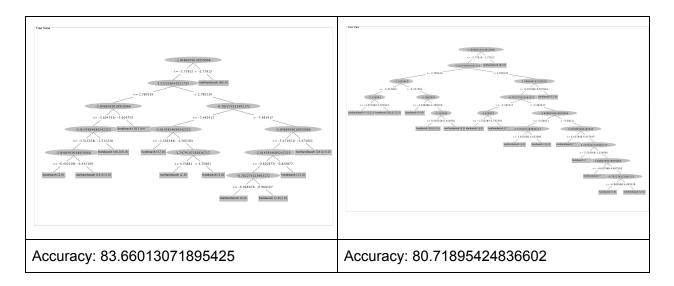


Part 3 - Using All Features

The results are reported below. The accuracy of using all 12 features is 3 percentage points lower than the original. This may be because we need to perform some feature selection in order to determine the best set of features to predict our labels (handwash or not).

Results:

Using 6 features (and sliding window 3)	Using 12 features (and sliding window 3)
---	--



Part 4 - Feature Selection

The feature selection results for the decision tree are reported below:

Individual Feature Accuracy:

Feature Name	Accuracy		
X Mean	82.6797385620915		
X Standard Deviation	70.91503267973856		
X Median	83.00653594771242		
X Root Mean Square	71.56862745098039		
Y Mean	74.83660130718954		
Y Standard Deviation	70.26143790849673		
Y Median	71.56862745098039		
Y Root Mean Square	70.91503267973856		
Z Mean	76.47058823529412		
Z Standard Deviation	70.58823529411765		
Z Median	77.12418300653594		
Z Root Mean Square	70.91503267973856		

The following are the set of features selected and the given accuracy for each feature:

Current feature set: Size: 1 [X Median] Current accuracy: 83.00653594771242

Current feature set: Size: 2 [X Mean, X Median]

Current accuracy: 83.00653594771242

Part 5 - Best Features and Different Classifiers

Random Forest Feature Selection:

Individual Feature Accuracy

Feature Name	Accuracy
X Mean	76.14379084967321
X Standard Deviation	65.359477124183
X Median	78.10457516339869
X Root Mean Square	68.62745098039215
Y Mean	68.62745098039215
Y Standard Deviation	57.51633986928105
Y Median	72.22222222223
Y Root Mean Square	66.01307189542484
Z Mean	67.3202614379085
Z Standard Deviation	61.43790849673203
Z Median	72.22222222223
Z Root Mean Square	63.071895424836605

The following are the set of features selected and the given accuracy for each feature:

Current feature set: Size: 1 [X Median] Current accuracy: 78.10457516339869

Current feature set: Size: 2 [X Median, Z Mean]

Current accuracy: 82.6797385620915

Current feature set: Size: 3 [X Standard Deviation, X Median, Z Mean]

Current accuracy: 83.333333333333333

Individual Feature Accuracy:

Feature Name	Accuracy
X Mean	83.00653594771242
X Standard Deviation	56.209150326797385
X Median	82.6797385620915
X Root Mean Square	72.54901960784314
Y Mean	72.22222222223
Y Standard Deviation	56.209150326797385
Y Median	72.22222222223
Y Root Mean Square	67.97385620915033
Z Mean	76.79738562091504
Z Standard Deviation	71.56862745098039
Z Median	75.16339869281046
Z Root Mean Square	71.89542483660131

The following are the set of features selected and the given accuracy for each feature:

Current feature set: Size: 1 [X Mean]
Current accuracy: 83.00653594771242

Current feature set: Size: 2 [X Mean, Z Median]

Current accuracy: 83.333333333333333

Best Classifier:

The best classifier is a tie between the Random Forest and SVM classifiers. They both provide an accuracy of 83.33%. However, they each do select a different set of features: Random Forest uses the three features of X Standard Deviation, X Median, Z Mean and the SVM uses two features of X Mean and Z Median.

Description of What Each Team Member Did:

Josephine and Trey worked together for this entire assignment. Both Josephine and Trey acted as subjects and collected data for handwashing and non-hand washing for this project. Trey did the pre-processing of the data, and Josephine created the features file. Both Trey and Josephine wrote the classification code and used Weka to create their classifiers. Finally, both Josephine and Trey wrote the report together.