Devon James
CSI 431
Prof. Petkov

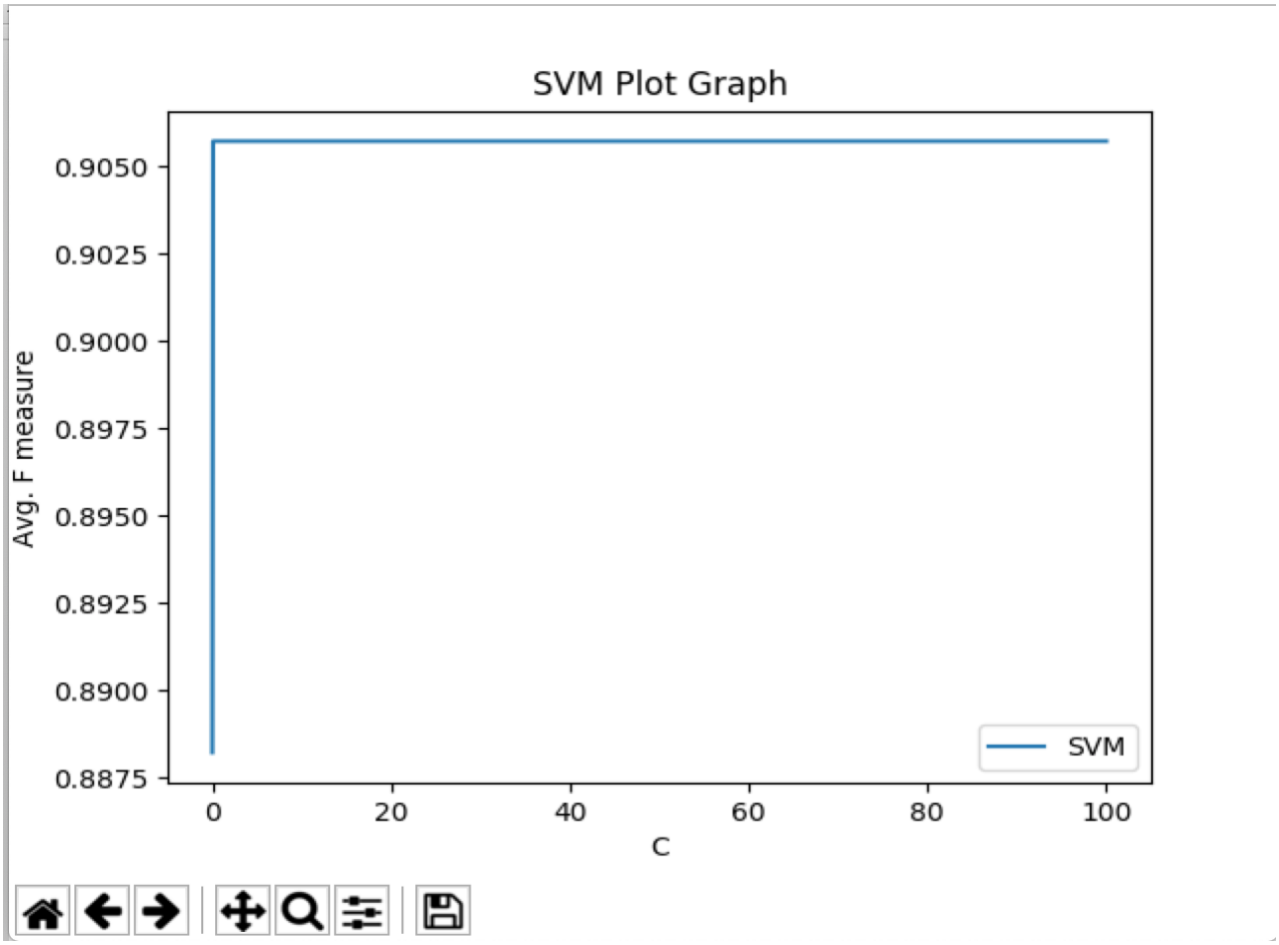
November 19th, 2018

HW #3 Solution

Part A

Question: Is smaller or larger margin better for this dataset (need to explain which C values are likely to produce smaller vs. larger values and then which end up being better in cross validation.)

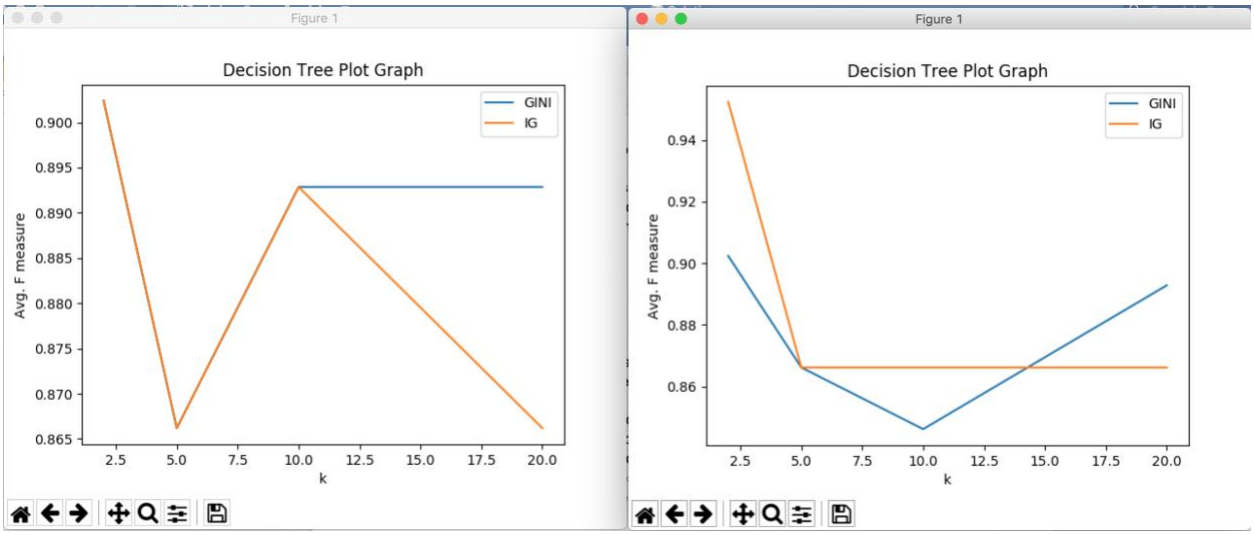
Answer: As shown in my results, after doing cross validation, the C values do not affect the average F measure of the dataset with the exception of C=.01, which has the lowest (and only non-equivalent) average. With that being said, the smaller margin (larger C values) may be better for this dataset due to the fact that when C is smaller, the dataset is at its lowest point.



Part B

Question: Does larger tree mean better F-measure? Which criterion is better?

Answer: After a series of instances of trial and error, I was unable to get this part of the program working properly. When plotting Decision Trees, the graphs looked different from each other each time I reran that part of the program as shown below:

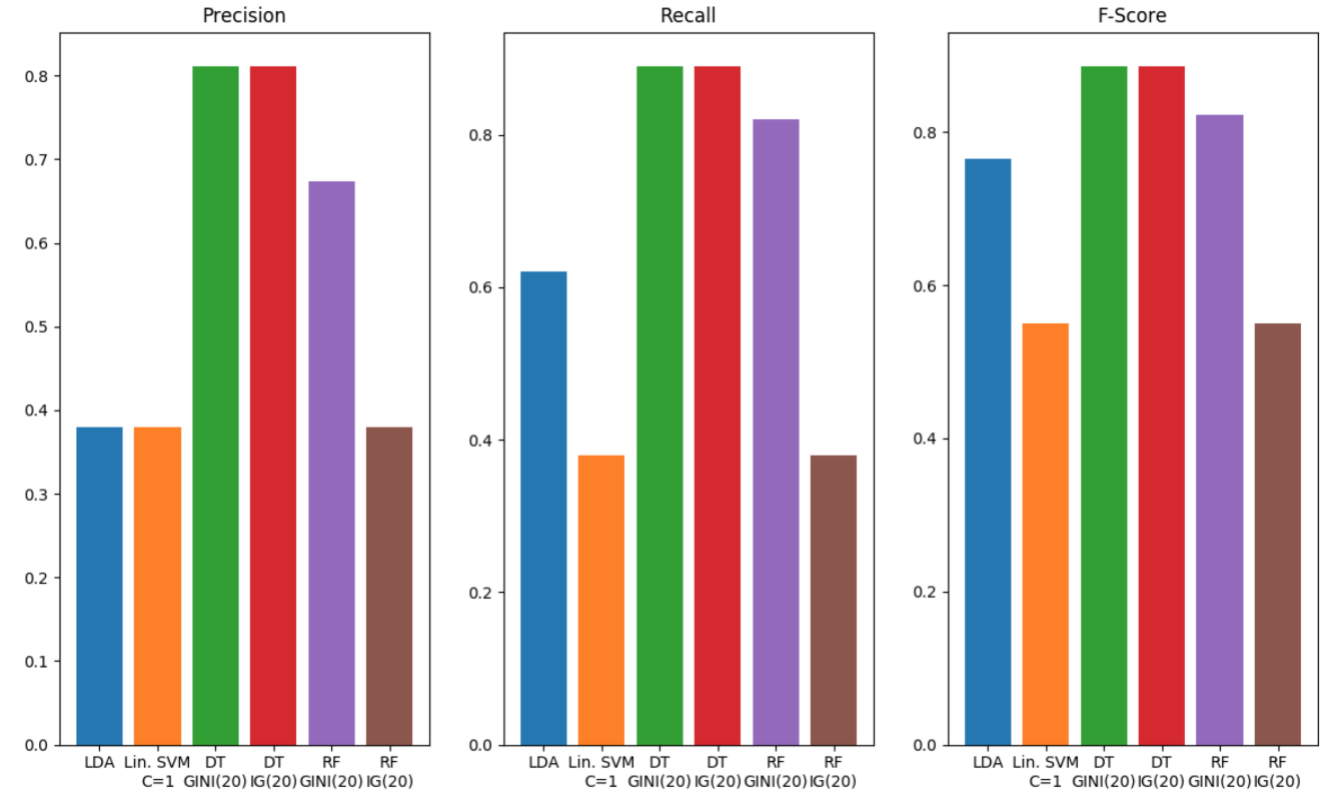


However, through all iterations, k=2 had the highest average F measure for the IG and k=20 had the highest average F measure for the GINI Index.

Part C

Question: Which are the best classifiers when you consider the different metrics? Is there a single winner for this dataset?

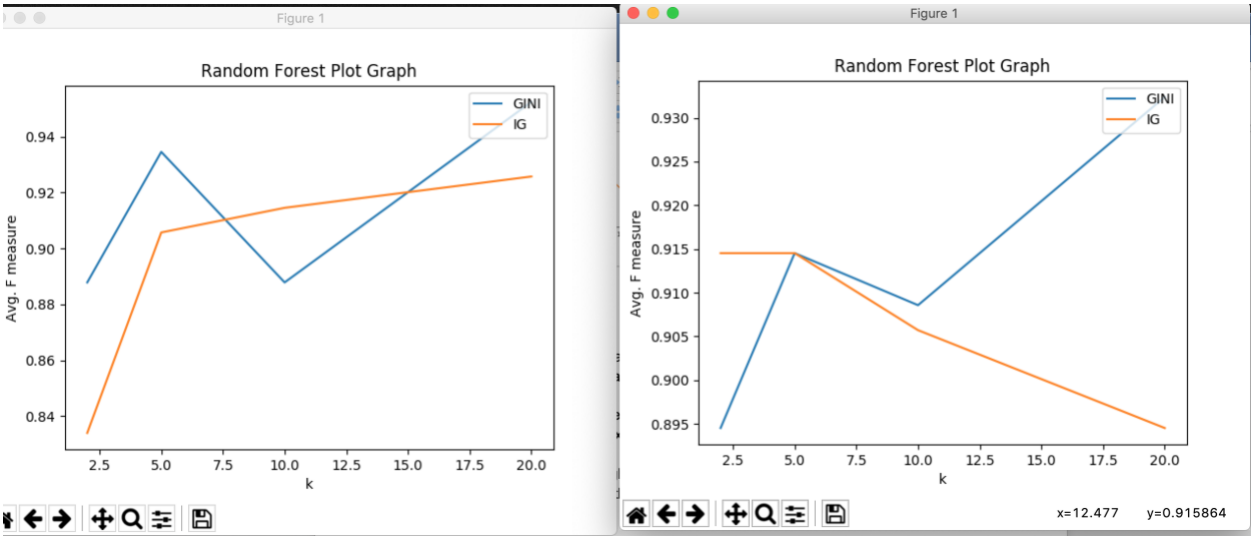
Answer: When looking at the metrics provided through my program, I can say that both Decision Tree (IG and GINI Index) classifiers are the best classifiers, with Random Forest (GINI Index) coming in as a close second due to their high F-scores. These three classes consistently have the highest Precision and Recall compared to the other classifiers. With that being said, there is no single winner for this dataset.



Part D

Question: Explain results (of Random Forest Classifier)

Answer: After a series of instances of trial and error, I was unable to get this part of the program working properly. When plotting Decision Trees, the graphs looked different from each other each time I reran that part of the program as shown below:



However, through all iterations, k=5 had the highest average F measure for the IG and k=20 had the highest average F measure for the GINI Index.