

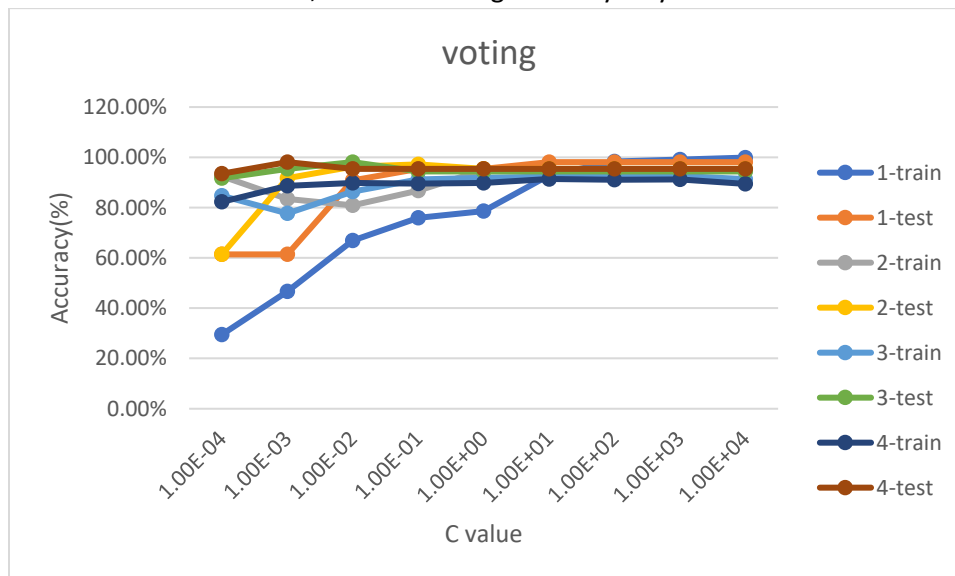
Empirical Analysis 1 (EA1)

Bagging

From my observations I noticed that as the depth of the tree increases the number of correctly classified instances. At a depth of 1 only 82.6% of the instances were correctly classified versus a depth of 10 where 91.1% of instances were correctly classified. Another observation that I discovered was the impact that number of trees had on the classification rate. At 10 trees used the correctly classified instances was at 92.8% where 108 a classes were correctly classified and 218 b classes were correctly classified, however when 100 trees were used even though a classification rate of 92.8% was observed, the a class only correctly classified 107 instances and the b class classified 219 instances.

Classification

1. My observations that I found when looking at the voting data show several interesting things. One of those observations is a general upward trend as C value increases. The lowest accuracy that occurred was on the linear kernel training data with a percentage of 29.5% and c value of $1.0E-4$ and a highest accuracy of 92.5% on the dual kernel training data at the same C value. Another observation that is revealed by looking at the graph of the voting data is the comparison between training accuracy and testing accuracy. As testing accuracy increases training accuracy generally increases as well. On this data set the training accuracy and testing accuracy both went up at the same time, but the data did not always increase by the same proportion at each value of C. The largest jump between training data was between C values of $1.0E-4$ and $1.0E-3$ when the accuracy went from 29.9% to 46.7%, and the testing accuracy stayed the same at 61.4%.



2. Looking at the Ionosphere data the graph looks very similar to the voting data with only a few differences. As with the voting data there is a general upward trend for all testing and training data as the C value increases. From the lowest C value to the highest C value there is a general increase in accuracy of 9-10% for the test data and an increase between -9-20% for the training data. Also like the voting data there is a general trend across the degree of polynomials where accuracy across all categories increases between 3-4% for each increase in polynomial degree. Overall, between all

data sets looked at it seems like both increasing the kernel degree and increasing the C size results in higher accuracy across all category.

