Questions For Machine Learning Toolbox:

1. The general trend of the curve is upwards, indicating that the most data used, the higher the accuracy of the test.
2. The curve is most noisy with a lower percentage of the data used. This is because there is more variability in the data selected. Because there is less training data, any outliers in the training data, will make the errors more drastic.
3. As seen in figure 1, the more times you run the accuracy test, the smoother the curve you get. While it is rather subjective what a “smooth curve” is, the data becomes relatively smooth at about 100 trials.
4. Changing the C value will skew your data. With a lower c value, the graph shows that the data is more accurate and becomes more accurate exponentially when compared to the percentage of data used. With a higher c value, the data is flat indicating that the percentage of data used for training has no effect on the data. The accuracy for this however, is very low and highly variable. When C is equal to 10^{-10}, the data becomes more accurate linearly with percentage data used and has an accuracy ranging from about 0.4 to 0.0.

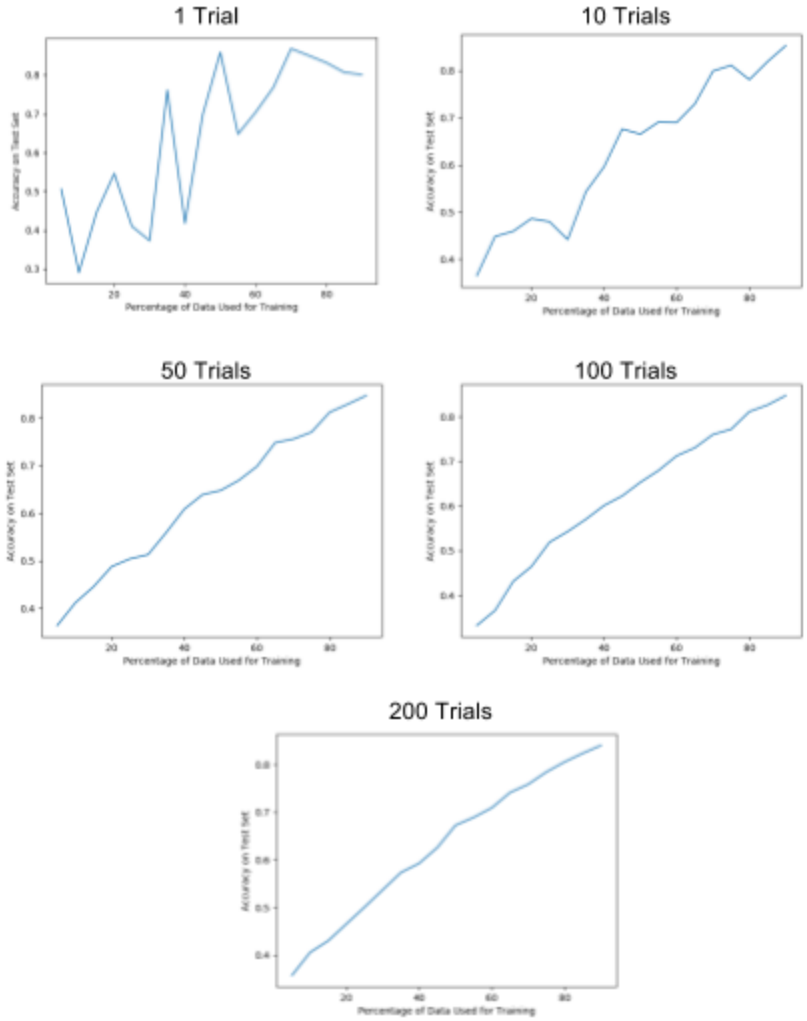

Figure 1: Graphs of accuracy of test with differing numbers of trials

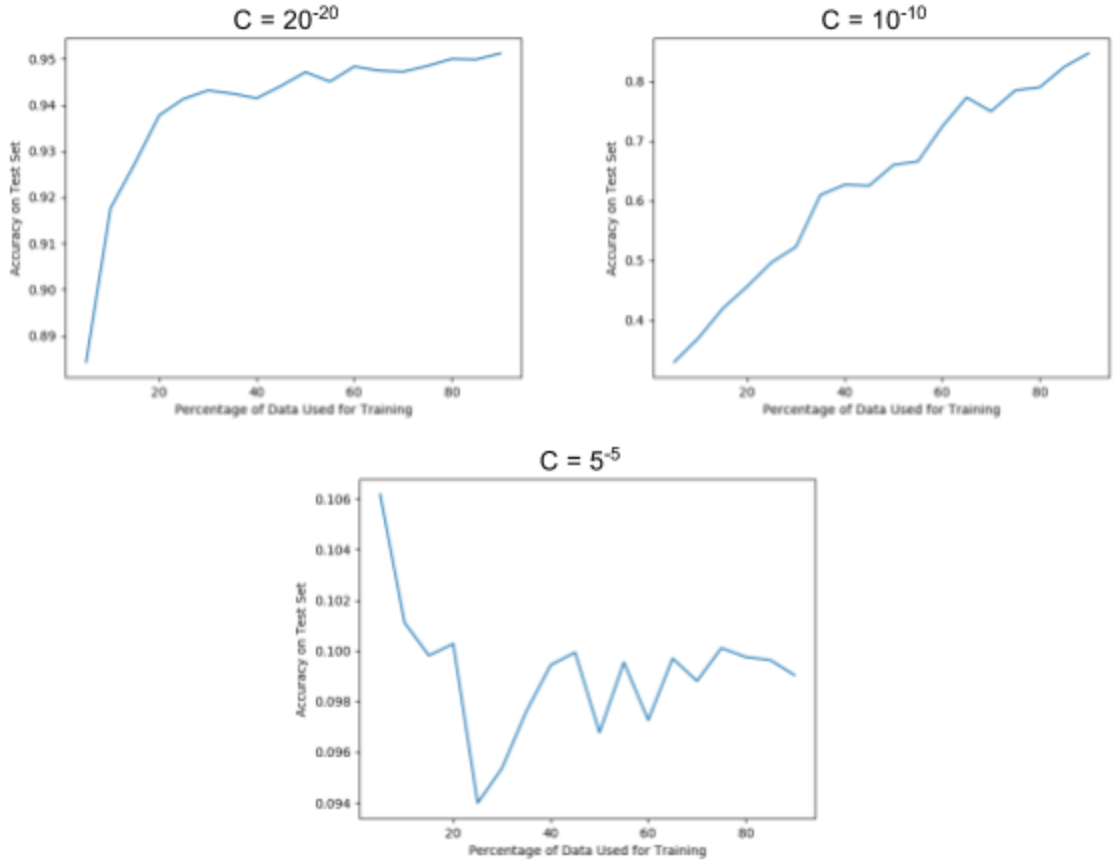

Figure 2: Differing C values.