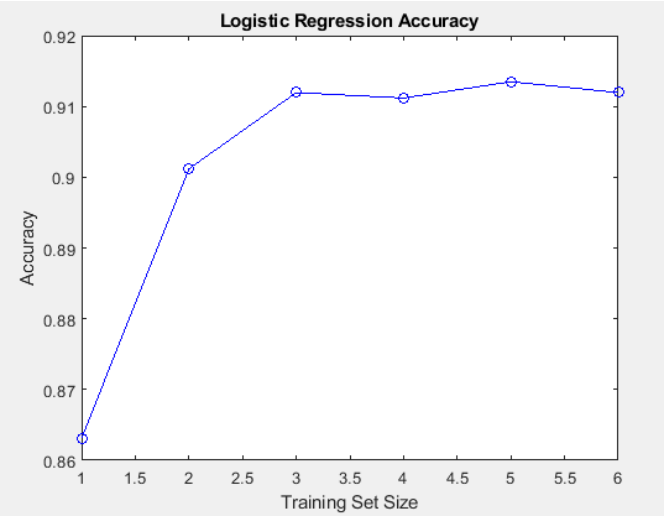
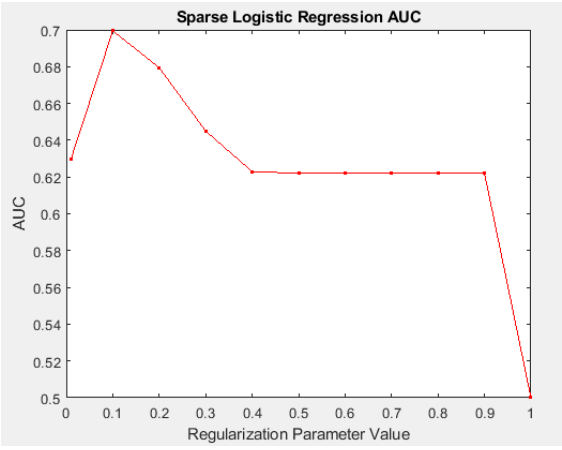
CSE 847: Homework 4

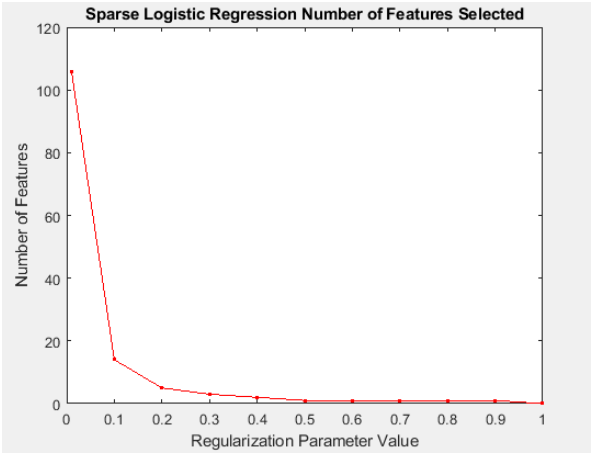
Ikechukwu Uchendu

1. It seems as though the training set does not have too much impact on the accuracy of our logistic regression model. A training set size of 200 already has 86% accuracy. This could mean that the first 200 data points are representative of most of the testing set. Using 1600 data points provided the greatest accuracy of 91.35%.



2. Interestingly, a regularization parameter of 0.1 produces the highest AUC. We then see that the AUC stays fixed at 0.6220 for regularization parameters 0.4 – 0.9. In other words, these regularization parameters give the same accuracy. Indeed, this result matches the number of features selected. At parameter values 0.5 – 0.9, we can see that one feature is selected. Assuming that it is the same feature selected for each parameter, it would explain their similar AUC values. For the regularization parameter value of 1, we can see that the number of features selected goes to zero and the AUC goes to 0.5. With no features we cannot make accurate predictions.





https://github.com/uchendui/CSE847/tree/master/Homework/Homework4