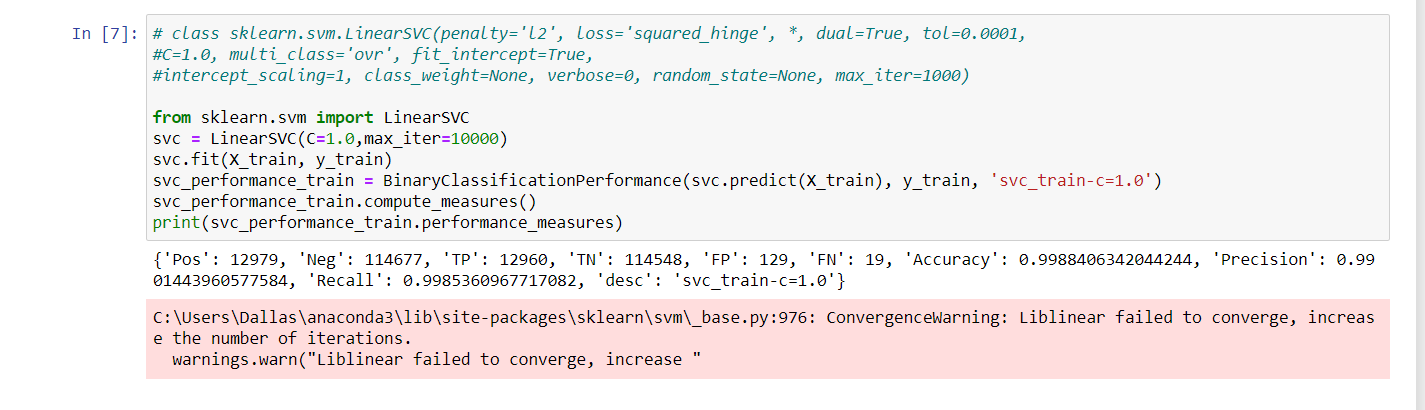
Dallas Alberti 11/8/2020

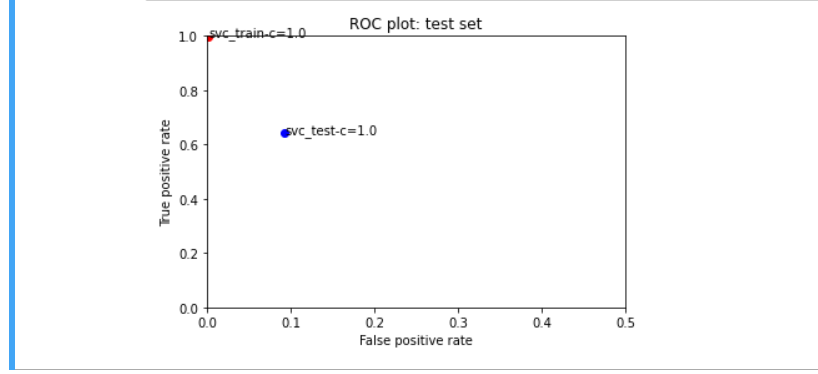
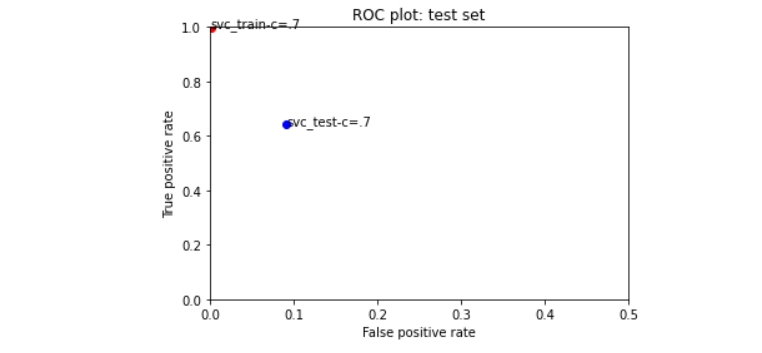
Machine Learning Discussion Assignment 3

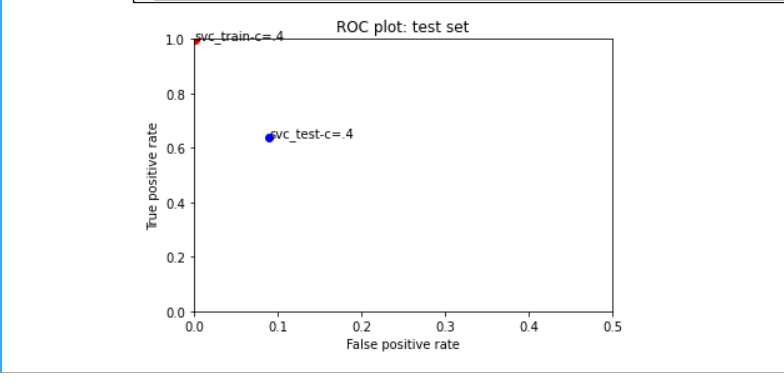
Feature Set:



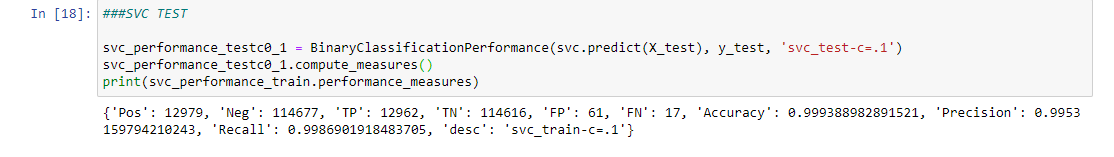
Error encountered: Failure to converge. Iterations increased from 1000 to 10,000 to 100,000. 

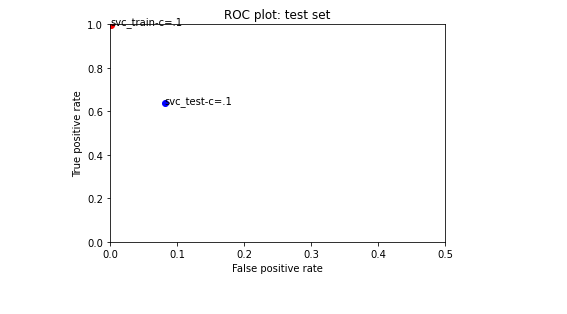
Tests: C=1, 0.7, 0.4, 0.1, 0.001, 0.0001

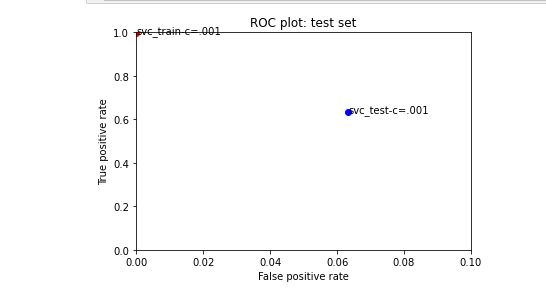
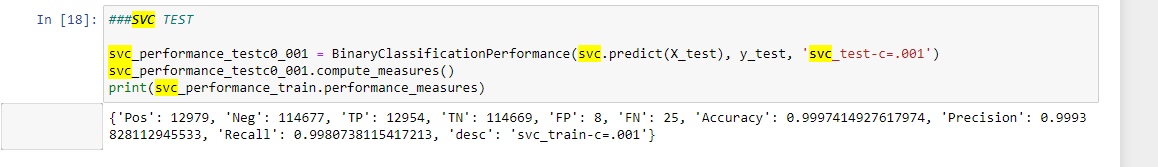
1. SVC(C=1.0) 
2. SVC(C=0.7)
3. SVC(C=0.4)

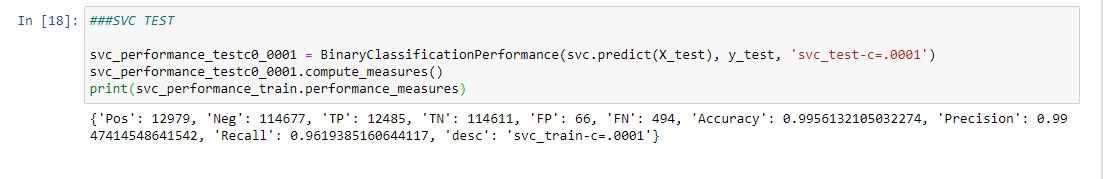
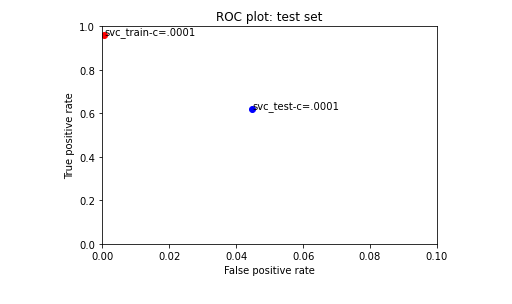


1. SVC(C=0.1)





1. SVC(C=0.001)
2. SVC(C=0.0001)



Writeup:

For discussion assignment 3 I introduce the ‘Linear Support Vector Classification’ as outlined in: [scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html#sklearn.svm.LinearSVC](https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html#sklearn.svm.LinearSVC). The documentation for the C-Support Vector Classification suggests to use this LinearSVC when samples are beyond tens of thousands of examples. I introduce two new blocks of code. The first is placed below the ‘MODEL:SVM, linear’ in the starter code. Here I ‘comment in’ the full example of a LinearSVC in order to reference different parameters that might be changed however for the purpose of this assignment only C values and max iterations are altered.

The code block was written in similar format to the other models that are tested. Before discussing the effects of C value manipulations it is important to note that every time I ran the code I received the error “ConvergenceWarning: Maximum number of iterations reached before convergence. Consider increasing max\_iter to improve fit.” This is shown in the second image of this document. I attempted to change the max iterations from 1000(Default) to 10,000 to 100,000 however no significant changes in performance were detected and my success measures did not change. The features used for this experimentation are outlined in image one.

I chose to begin the test with a default C value of C=1 and then work my way down to a value of .0001. One may notice that the scale of the graphs changes as I iterate through different values of C and that is because I started to realize that changes to this value were very small and often undetectable with an X-Scale of 0-1.0. My initial reaction was that as the C value was decreased the false positive rate would decrease. However upon looking at the 3 performance measures images for C=.1, C=.001, and C=.0001 we can see that FP initially decreases (.1) before increasing again (.0001). In addition to an increase in FP for the C=.0001 test, the FN rate greatly increases from 25(c=.001) to 494(c=.0001). The results of this testing would suggest that, for this feature set, an optimal C value would be closer to .001. In other words, regularization has a strong positive impact on the models accuracy until a certain threshold is met.