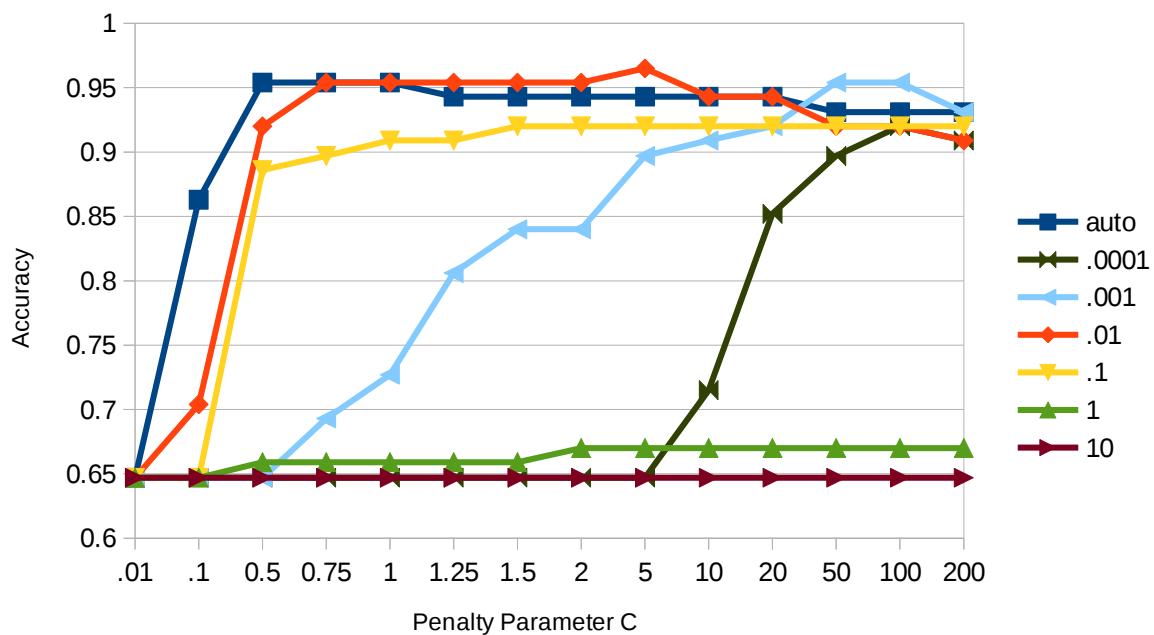


## Ionosphere

To begin I first needed to standardize my ionosphere data. I did so by first changing the b's to 0's and the g's to 1's. I then shuffled and split my data into training, validation, and testing sets with half in training, one fourth in validation, and one fourth in testing. Next, I fitted the StandardScaler tool in its default state from the scikit-learn library on the training data and transformed each dataset. When adjusting the hyper parameters C and Gamma by hand, I obtained the following results from the SVC:



The C parameter is displayed on the x-axis. The gamma parameter is displayed as separate lines with value denoted in the legend.

From the above data I found that hyper-parameters  $C=5$  and  $\gamma=.01$  gave the best accuracy. When I used the GridSearchCV tool provided by sklearn on the same parameter set, I found that the best parameters were  $C=1$  and  $\gamma=0.1$ . This difference is likely due to the GridSearchCV evaluating the estimator using more than just the accuracy of its predictions. I then used GridSearchCV to do the same operation but using large parameter sets that were randomly generated. The best parameters given for this method was  $C=0.955$  and  $\gamma=0.0959$  which is extremely close to the previous answer given. Finally when I ran the optimized SVC on my test set, I got an accuracy of 90.9%.

### **vowel-context**

To begin I first needed to standardize my vowel-context data. I did so by first removing the first three columns in the data that were irrelevant. I then shuffled and split my data into training, validation, and testing sets with half in training, one fourth in validation, and one fourth in testing. Next, I fitted the StandardScaler tool in its default state from the scikit-learn library on the training data and transformed each dataset. When I used the GridSearchCV tool provided by sklearn on the parameter set that I had used in the ionosphere example, I found that the best parameters were  $C=10$  and  $\gamma=0.1$ . I then used GridSearchCV to do the same operation but using large parameter sets that were randomly generated. The best parameters given for this method was  $C=16.45$  and  $\gamma=0.0435$  which is nearby but substantially further apart from the previous answer given compared to the ionosphere example. Finally when I ran the optimized SVC on my test set, I got an accuracy of 92.7%.

### **sat**

To begin I first needed to standardize my vowel-context data. Since the data was already split, I simply shuffled the data. Then, I fitted the StandardScaler tool in its default state from the scikit-learn library on the training data and transformed both datasets. When I used the GridSearchCV tool

provided by sklearn on the parameter set that I had used in the ionosphere example, I found that the best parameters were  $C=1$  and  $\gamma=0.1$ . I then used GridSearchCV to do the same operation but using large parameter sets that were randomly generated. The best parameters given for this method was  $C=1.135$  and  $\gamma=0.0946$  which is near to the previous answer given. Finally when I ran the optimized SVC on my test set, I got an accuracy of 91.1%.