

Question 8

Michael Ball, Alex Caravan, Alec Guertin, Peter Sujan

We chose to optimize the SVM with the radial kernel. For each of parts a, b, and c we performed a grid search over values of C and μ . For C , we tried 2^{-5} , 1, 2^5 , and for μ we tried 2^{-15} , 2^{-10} , 2^{-5} , 1, 2^5 . For efficiency reasons, we used 1/4 (2000) of the 8000-observation training set. After choosing the optimal parameters, we re-trained the SVM using all of the training data, and predicted on the test set. Using just training accuracy to compare models (part a), we achieved a very poor test accuracy of 0.565. Using a 70/30 split of the training data for holdout purposes, we achieved a slightly better 0.581 test set accuracy. Finally, we attempted cross-validation to better predict model accuracy on unseen data, but this method gave no improvement (in fact, it chose exactly the same parameters as method b, $\mu = 1$ and $C = 32$). We suspect that the poor performance on the test data is due to the SVM overfitting the training data. It seems that the holdout set and CV methods aren't enough to combat this effect (and perhaps the training data subset too small). We found that by choosing parameters that gave worse accuracies on the training set during the grid search, we could actually improve our test accuracy to around .62-.63. In conclusion, using grid search, we actually failed to do better than the baseline. However, by picking slightly "worse" parameters to prevent overfitting, we were able to achieve improved accuracy.