

CSE 474/574 Assignment 3

Kyle Shuttleworth

Calvin Kyi

Yuanyi Yang

Logistic Regression:

Training Set Accuracy: 86.214%

Validation Set Accuracy: 85.31%

Testing Set Accuracy: 85.39%

Training error is the error you get when you run the trained model back on the data that was used to train the model. While testing error is the error you get when you run the trained model on data the model has never seen before. Models will typically have higher accuracy when tested against the data it was trained on.

Training Error Class 1: 2%

Training Error Class 2: 2.1%

Training Error Class 3: 6.2%

Training Error Class 4: 7.5%

Training Error Class 5: 4.4%

Training Error Class 6: 8.3%

Training Error Class 7: 3.3%

Training Error Class 8: 4.3%

Training Error Class 9: 10.1%

Training Error Class 10: 9.7%

Testing Error Class 1: .1%

Testing Error Class 2: .17%

Testing Error Class 3: 4.7%

Testing Error Class 4: 4.1%

Testing Error Class 5: 1.55%

Testing Error Class 6: 5.55%

Testing Error Class 7: 1%

Testing Error Class 8: 2%

Testing Error Class 9: 8.3%

Testing Error Class 10: 7.3%

SVM:

Using linear kernel (all other parameters are kept default):

Training Set Accuracy: 97.286%

Validation Set Accuracy: 93.64%

Testing Set Accuracy: 93.78%

Using a radial basis function with the value of gamma setting set to 1 (all other parameters are kept default).

Training Set Accuracy: 100%

Validation Set Accuracy: 15.48%

Testing Set Accuracy: 17.14%

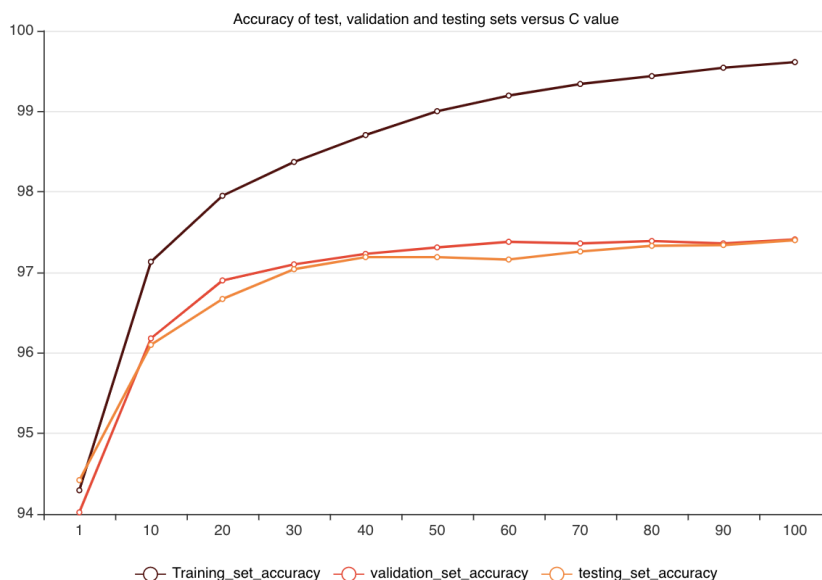
Using radial basis function with value of gamma setting to default (all other parameters are kept default):

Training Set Accuracy: 94.294%

Validation Set Accuracy: 94.02%

Testing Set Accuracy: 94.42%

Using radial basis function with value of gamma setting to default and varying value of C (1, 10, 20, 30, . . . , 100) and plot the graph of accuracy with respect to values of C in the report:



The radial basis function with value of gamma setting to 1 model was way overfitted to the training data and due to this it had by far the worst results out of the models tested. The radial basis function with the value of gamma setting set to default was more accurate on the validation and test sets while the linear kernel model was more accurate on the training set. Due to the divergences in accuracies between the training, validation, testing sets accuracies we can deduce that the linear kernel model is slightly overfitted and that the radial basis function with the value of gamma setting set to default is the better choice since it doesn't have these large divergences and therefore we can deduce that this model is best fitted. There is an improvement in all of the accuracies when we set the c attribute on the radial basis function with the value of gamma setting set to default model.

The best choice of parameters after these experiments is the radial basis function with the value of gamma setting set to default model and the value of c set to 100:

Training Set Accuracy: 99.612%

Validation Set Accuracy: 97.41%

Testing Set Accuracy: 97.4%