SEng 474 - Assignment 2

Kenil Shah

V00903842

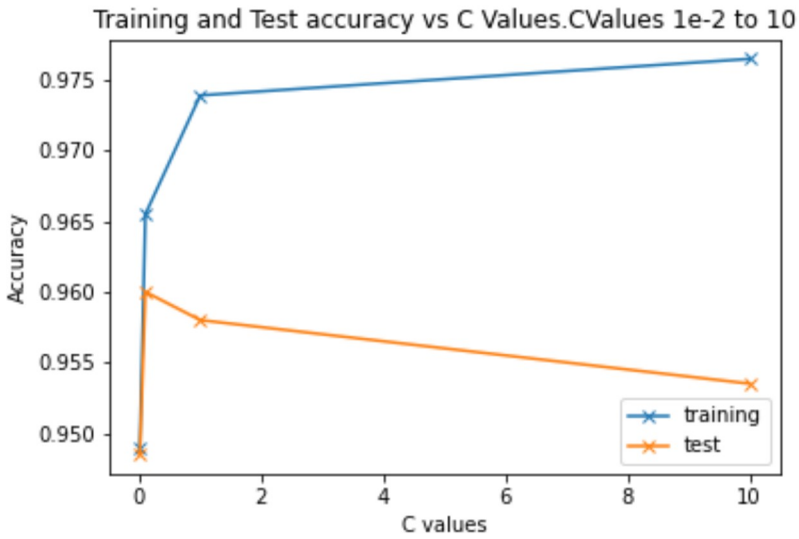
Fashion-MNIST

For all experiments training size was of 12000 examples and testing size was of 2000 examples

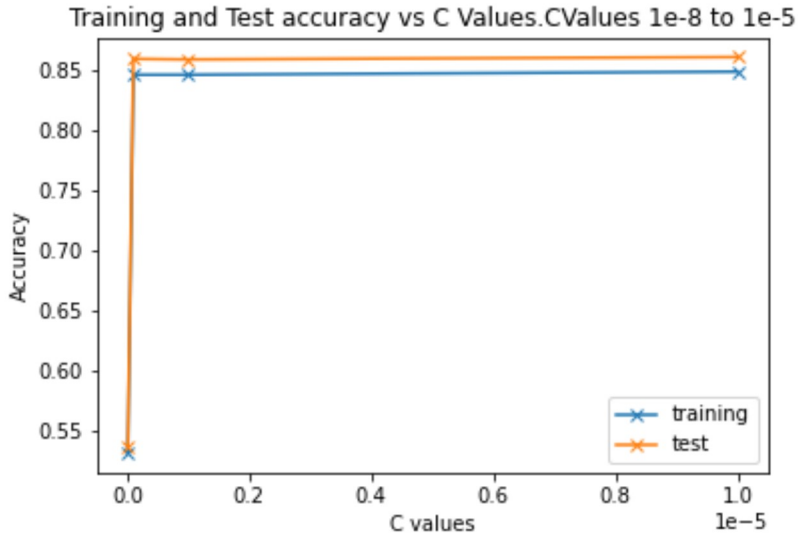
1. Logistic Regression

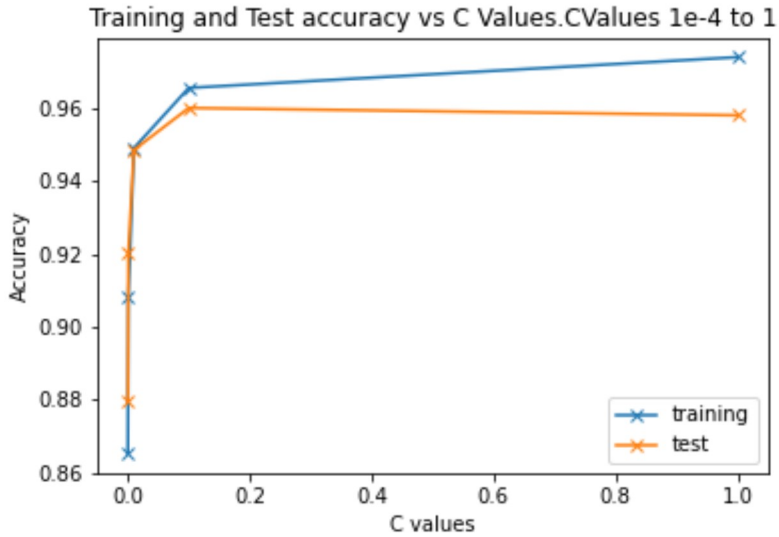
First we are going to use logistic regression. For the analysis of our classification we compare the regularization parameter C with training and test accuracy . We vary the values of c as 1e-8 to 1e4 with c0 = 1e-8, alpha = 10, 0<=i<=12.

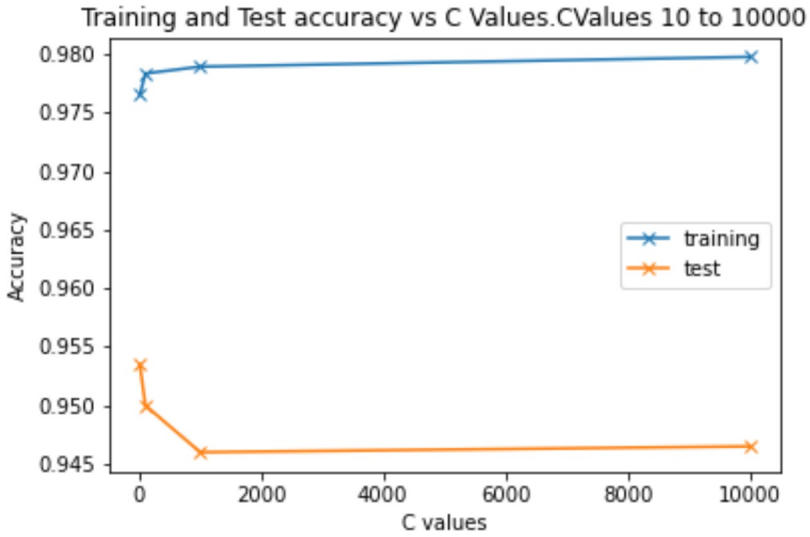
We find that a c-value of 0.1 gives us the best accuracy of 96%.



We show 3 more diagram which show phases of underfitting, overfitting and max accuracy.

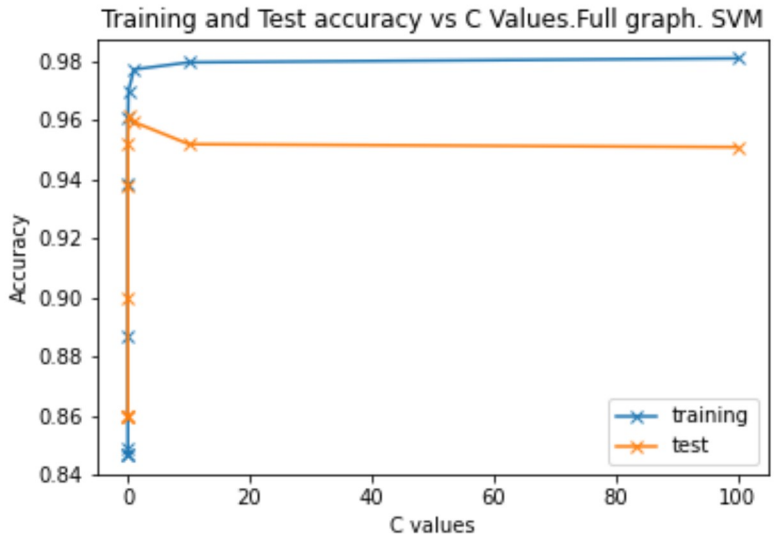




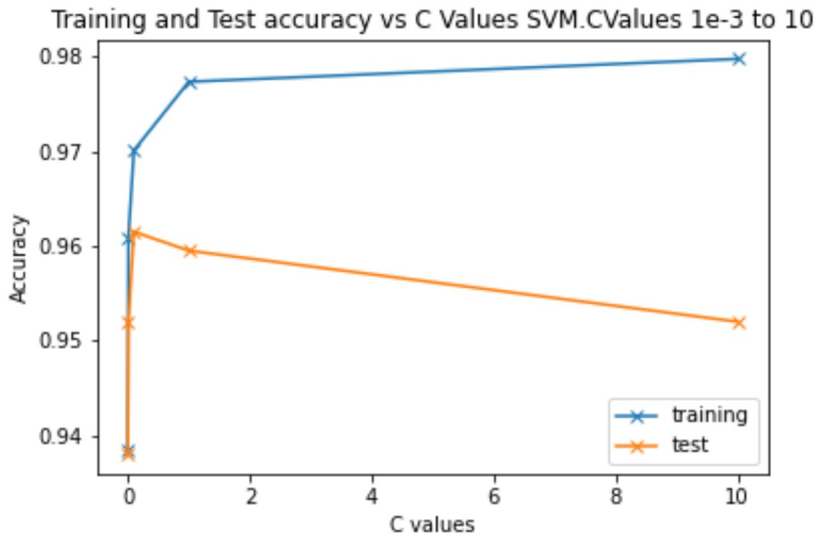


2. Support Vector Machine

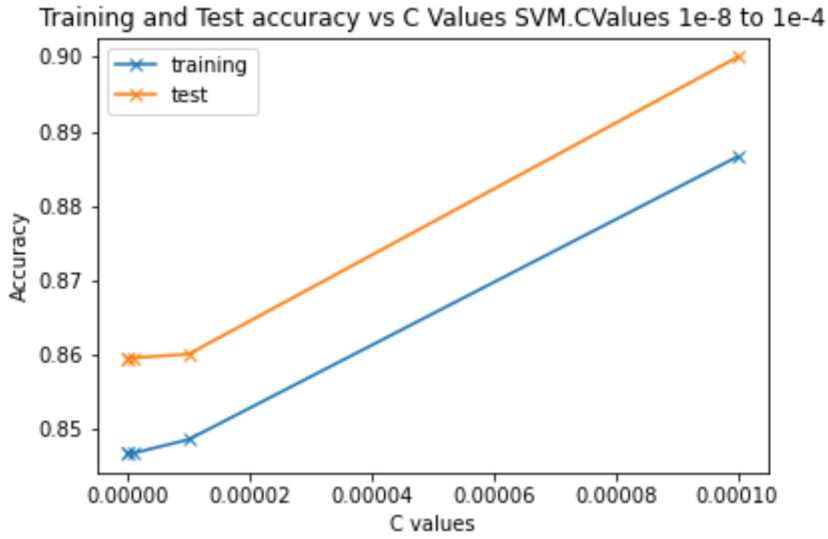
The second classification method we use is the support vectom machines. Similar to logistic regression to do the analysis of our classification we compare the regularization parameter C with training and test accuracy. We keep c0 and alpha the same, but change i to 0<=i<=10. Thus the c values vary from 1e-8 to 100.

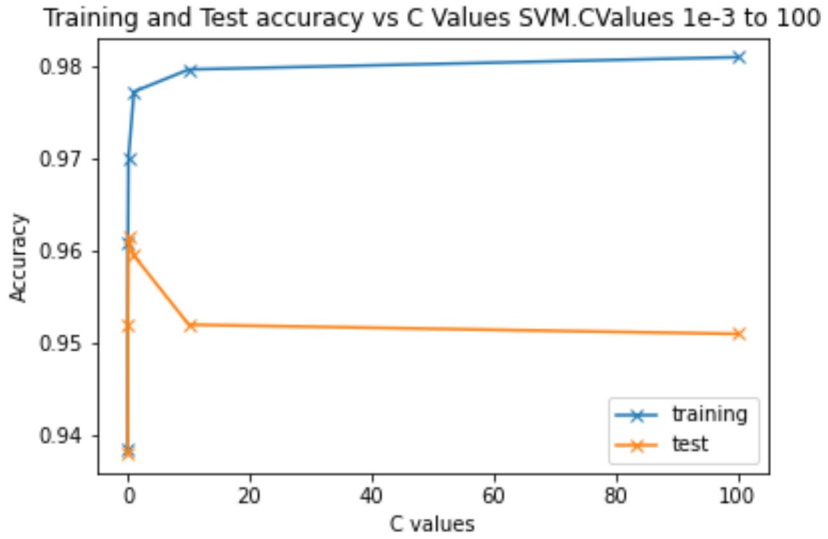


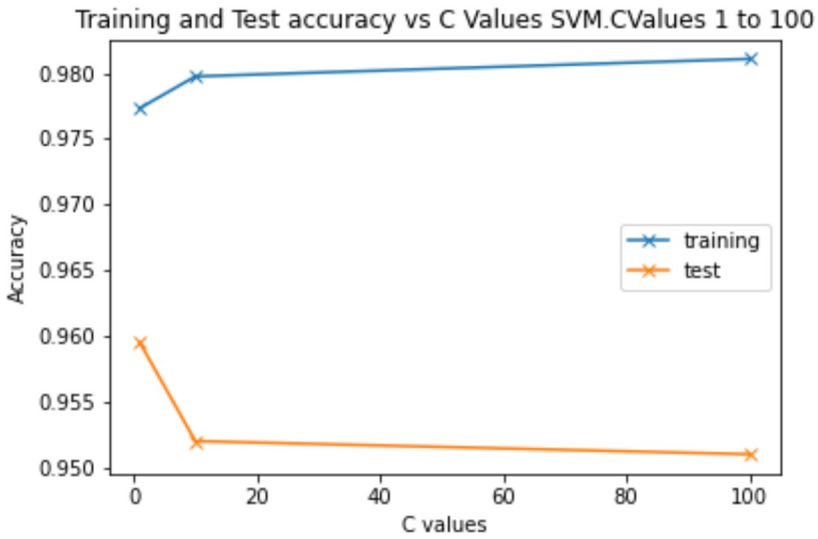
We find that a c-value of 0.1 gives us the best accuracy of 96.5%. This is a very marginal improvement oven logical regression.



We show 3 more diagram which show phases of underfitting, overfitting and max accuracy.

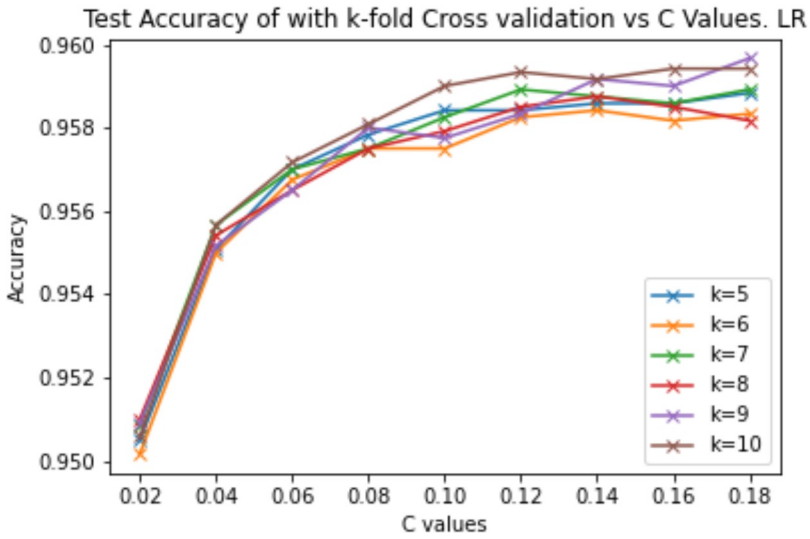


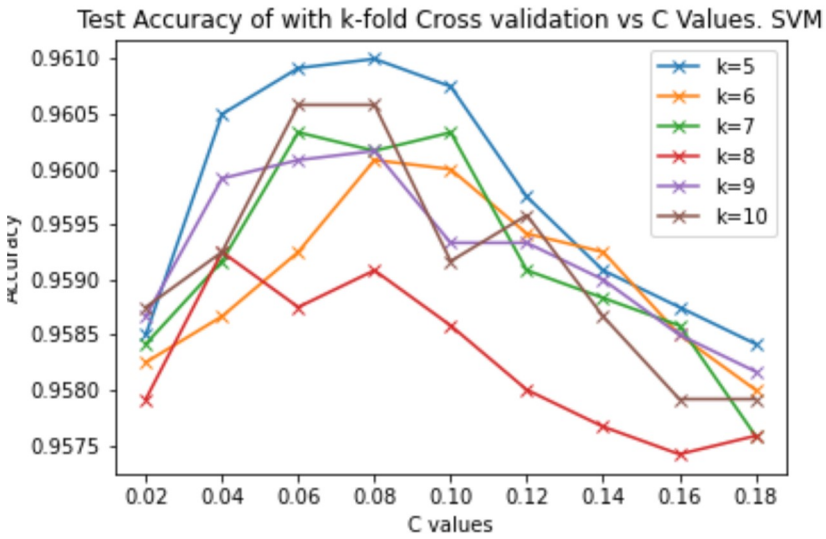




3. K-Fold Cross Validation

Now, for both SVM and Logistic regression we use K-fold cross validation. In our prevous experiment we found that both max accuracy occurs at C=0.1. Thus, we use 9 values of C centered around 0.1 as:- C-values = [0.02,0.04,0.06,0.08,0.1,0.12,0.14,0.16,0.18] Now for K = 5,6,7,8,9,10; we compare average test accuracy against regularisation parameter and obtain the following results for Logistic regression followed by SVM.





For Logistic regression we get the max accuracy of 95.97% for a c value of 0.18 and k value of 9

For SVM we get the max accuracy of 96.1% for a c value of 0.08 and k value of 5

Now as we have found the optimal c values for both LR and SVM, we apply this to the entire training set and find that both Logistic regression and SVM have the same max accuracy of 96.1 and thus there is no difference in classification accuracy for this case in either logistic regression or SVM. This is because of the large dataset which naturally converge to this accuracy value

References

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

https://scikit-learn.org/stable/modules/svm.html#svc

https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html

https://machinelearningmastery.com/k-fold-cross-validation/

https://numpy.org/doc/stable/reference/generated/numpy.split.html

https://stackabuse.com/python-how-to-flatten-list-of-lists/