# Assignment 1 – SENG 474

## **Logistic Regression**

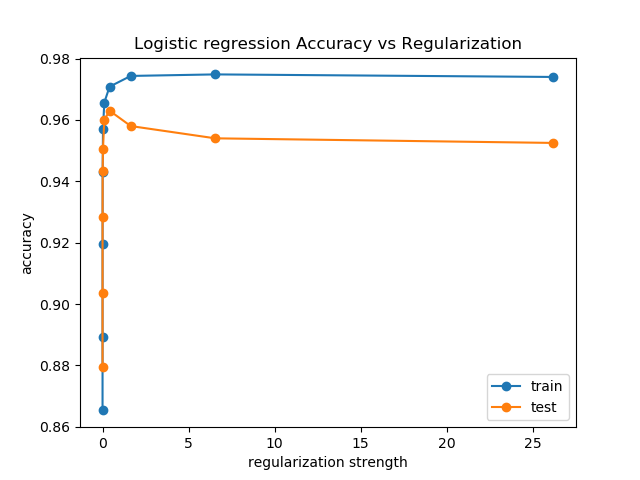
The task for logistic regression was to find a regularization strength *C* the has the best test accuracy/least test error. This was done using the *l2* penalty where training is penalized according to . To do this the following formula was used:

****

Where *C*0 = 0.0001; α = 4; and 0 ≤ *i* < 10

This produced values of *C* ranging from 0.0001 to 26.2. *C*=0.4096 (*I* = 6) was the most accurate with 96.3% test accuracy or 3.7% test error (97.5% training accuracy / 2.5% training error). See figure 1 below for a comparison of *C* and the respective accuracy.

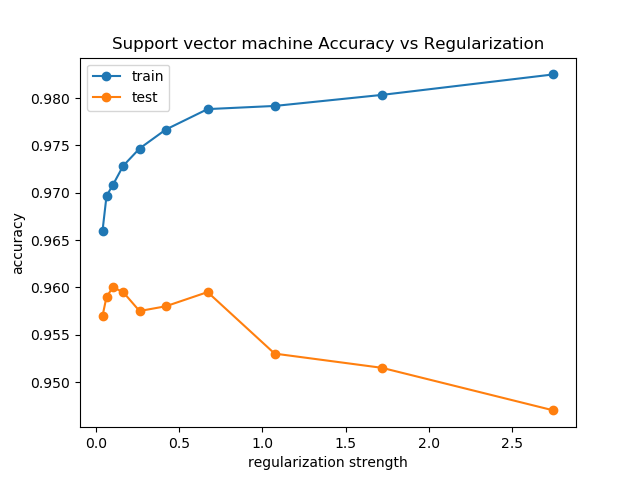
Figure 1 – Logistic Regression Test Accuracy



## **Support Vector Machines**

The task for training the support vector machine (SVM) was the same as logistic regression, determining a value of *C* that has the least amount of error on the test data. This was done on an SVM with a linear kernel. The same formula was used to determine the best value of *C.* In this case *C*0 = 0.004; α = 1.6; and 0 ≤ *i* < 10. The least error obtained on the test set was 4.0% or 96.0% accuracy for both *C*= 0.164 and 0.671. Figure 2 below illustrates the training and test error for all values of *C* that were tested.

Figure 1 – Logistic Regression Test Accuracy



## **K-Fold Cross-Validation**

## **Gaussian Kernel**