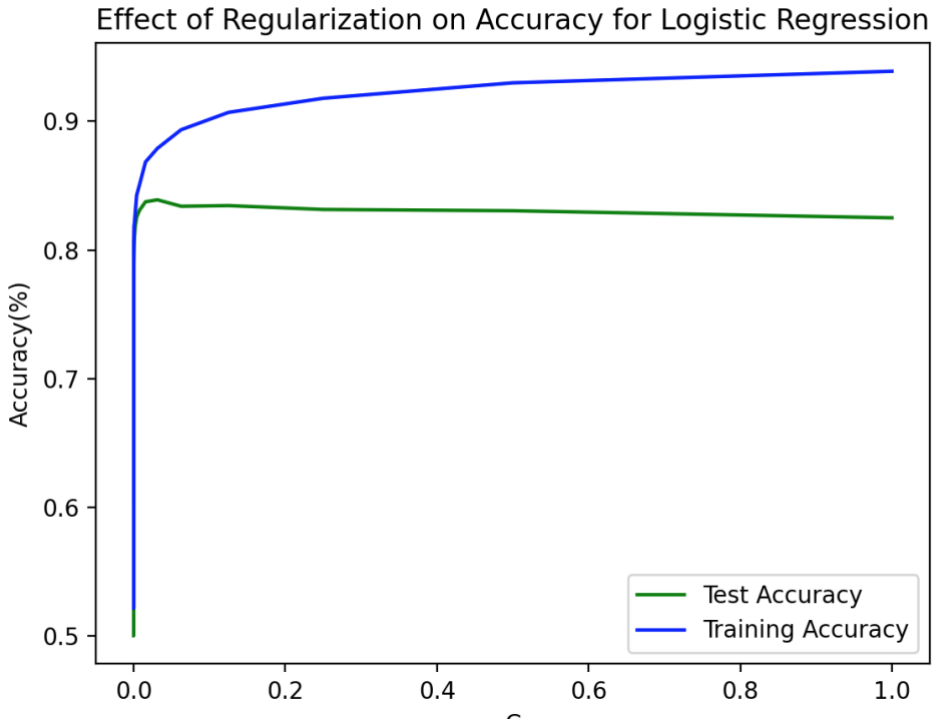
Nick Wurzer
V00958568
February 27, 2023
Seng 474

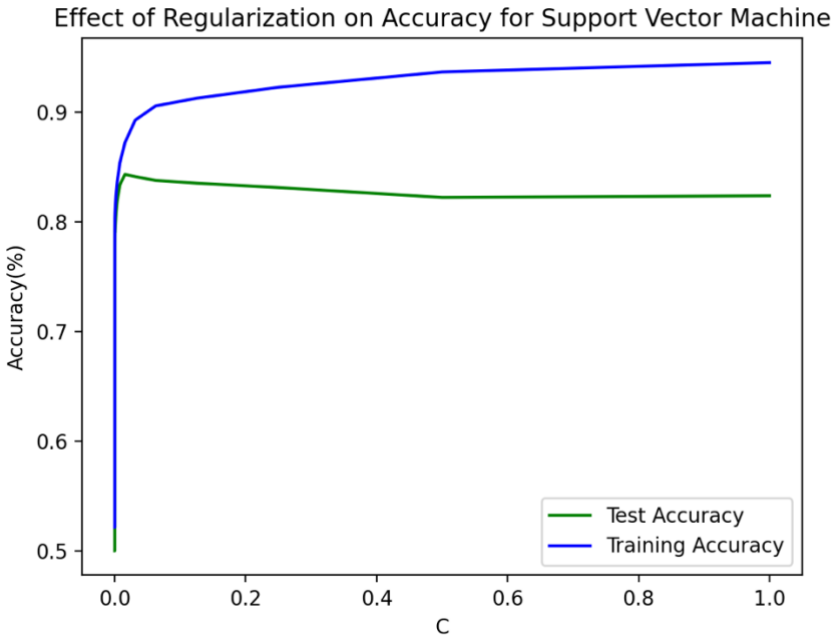
A2 Report

Since testing with SVM was taking quite a long time, I reduced my training size to 2000.

For the logistic regression model the Value of C which obtained the best results (highest test accuracy) was C = .00390625. From the graph below, the model underfits when 0 < C < .00390625 and the model overfits when C > .00390625.



For the linear support vector machine model, the Value of C which obtained the best results (highest test accuracy) was C = .001953125. From the graph below, the model underfits when 0 < C < .001953125 and the model overfits when C > .001953125.



After implementing and running k-fold cross-validation on both the logistic regression and SVM models, I obtained different values for the optimal C. I then trained both models on the full training set with the optimal value for C and got 95% confidence intervals of .8204 +- .0136 and .8254 +-.0136 for logistic regression and SVM respectively. Since these confidence intervals overlap, we can say with 95% confidence that there is not a significant difference between these two models for this application.

Unfortunately, I did not have time to reach the fourth part of the assignment.