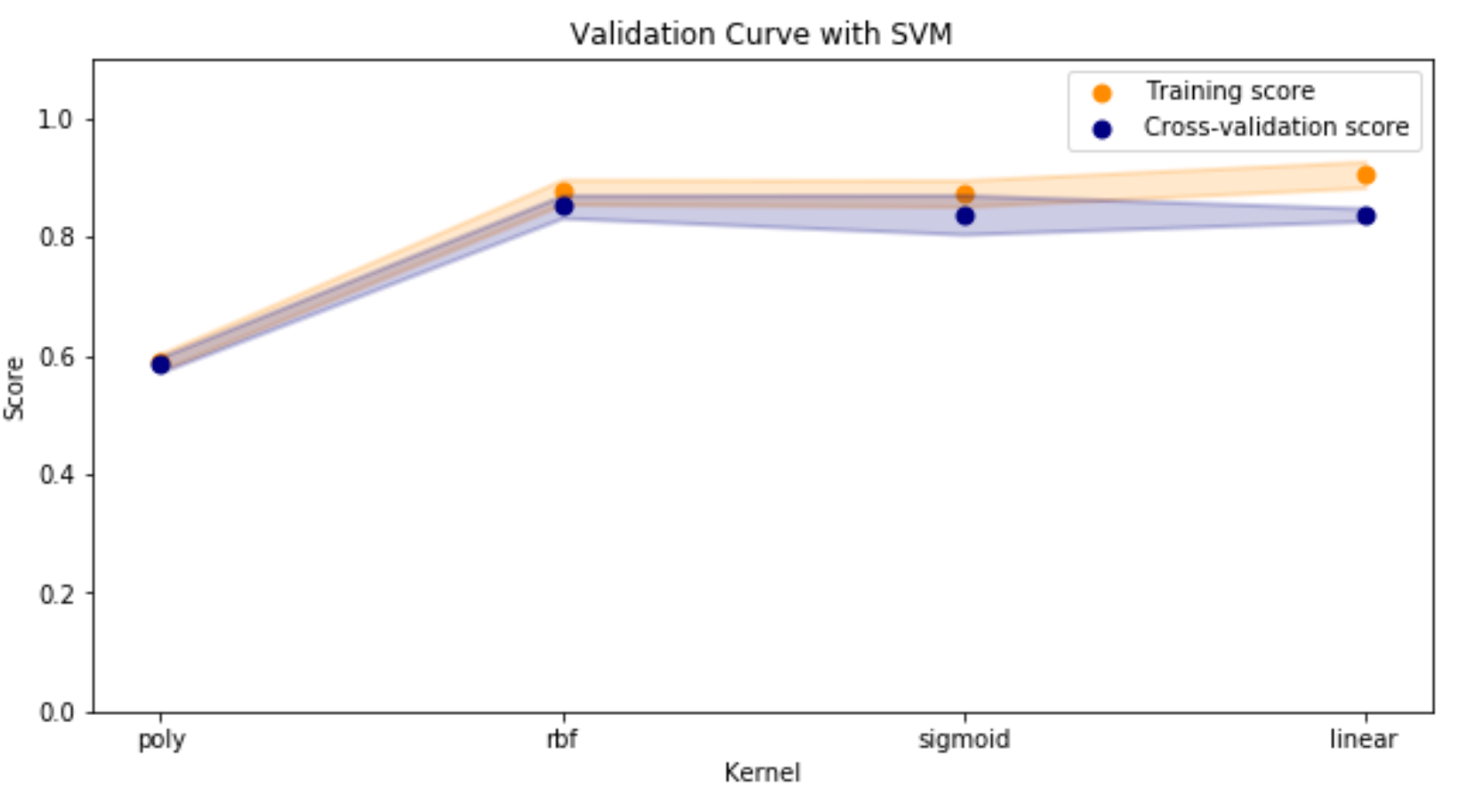
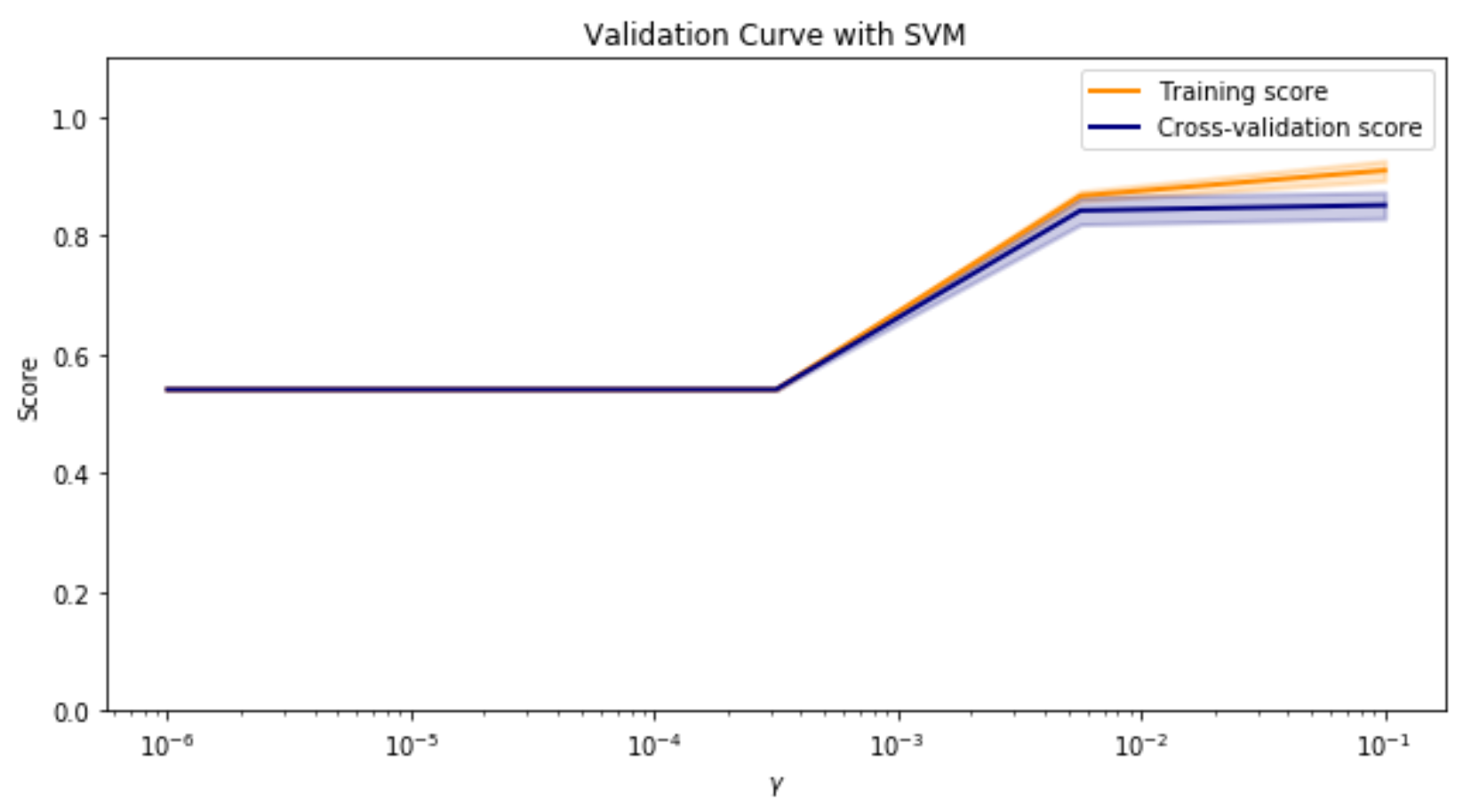
From Scikit learn: <https://scikit-learn.org/stable/modules/svm.html>

**Support vector machines (SVMs)**

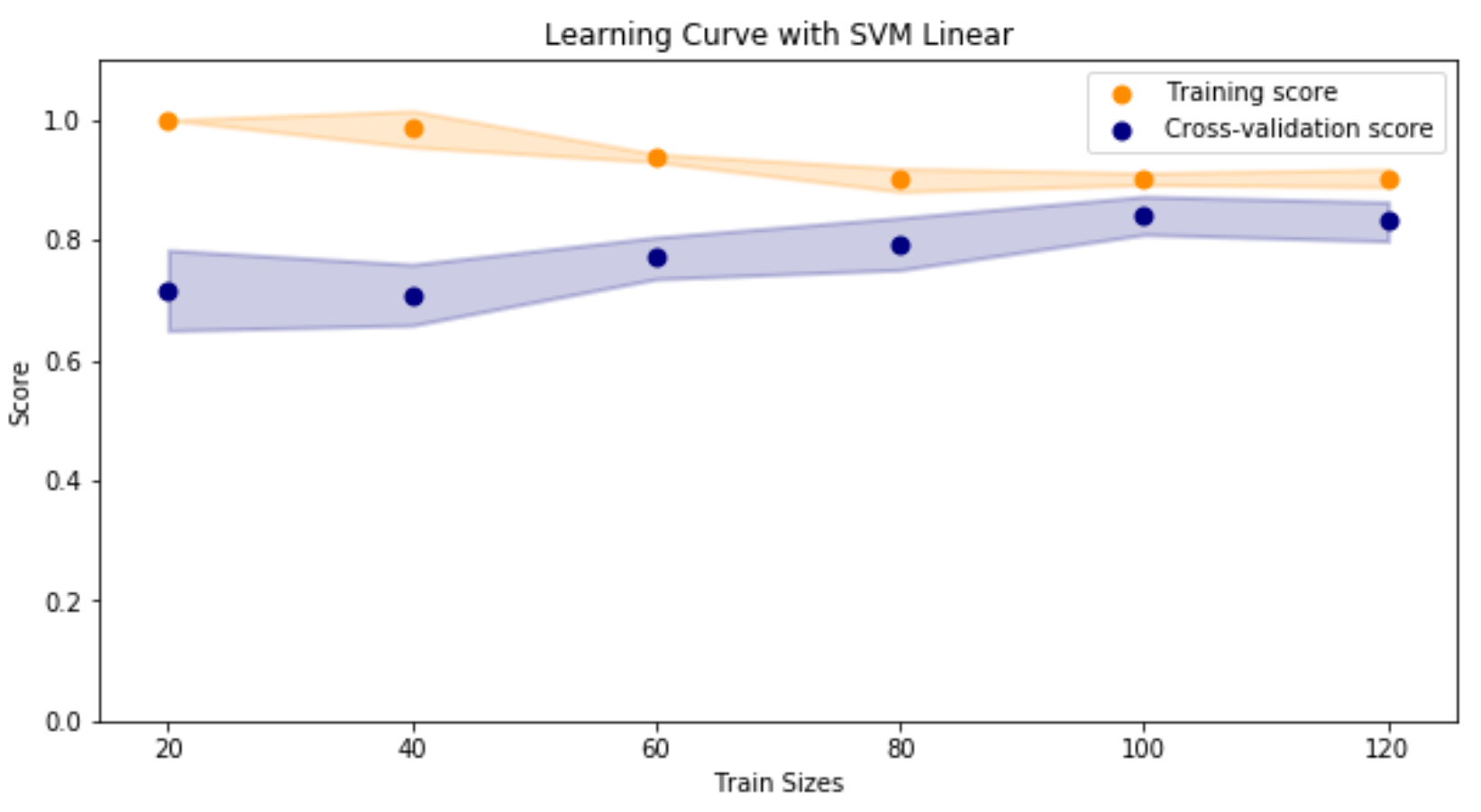
This is a Supervised (correct answer is known) learning method used for classification. It can also be used for regression and outliers. The key advantage for the Heart Disease data set is that SVM is effective for high dimensional spaces. The data set has 13 features (5 variable and 8 categorical). Once the categorical data is reformatted for binary there are 28 features.

Validation curves were used to visualize and ultimately select the input parameters to maximize the accuracy of the model. A linear kernel was selected and a gamma value of 0.1 to maximize the cross-validation score. The decision function type was irrelevant.

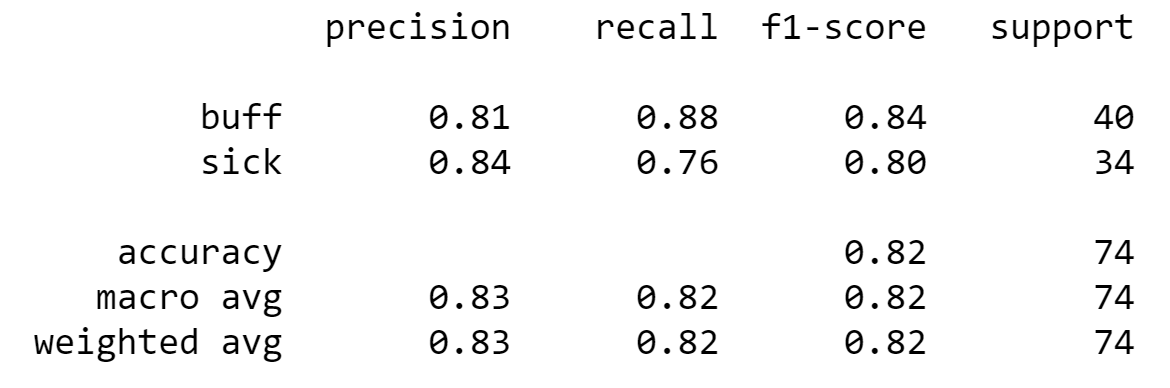




Next, the size of the training set was evaluated using a learning curve. It was confirmed that the default train/test/split method from scikit-learn splits training (120) and testing (74) data sets where the scores converge.



Running the test data set, the following classification report results:



Running this data set on the pre-set train/test split resulted in the following:

* 82% of the time Heart Disease or Healthy was predicted correctly (Accuracy)
* 83% of Heart Disease predictions were Heart Disease (Precision)
* 82% of Heart Disease samples were correctly predicted (Recall)
* Overall 82% harmonic average of precision and recall (F1 Score)