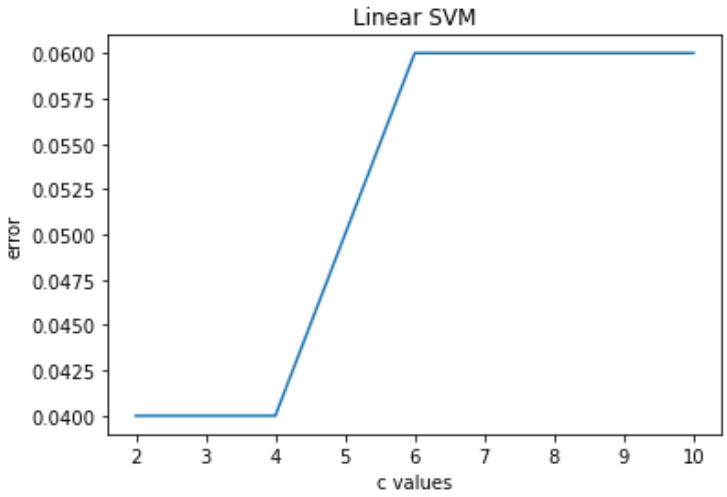
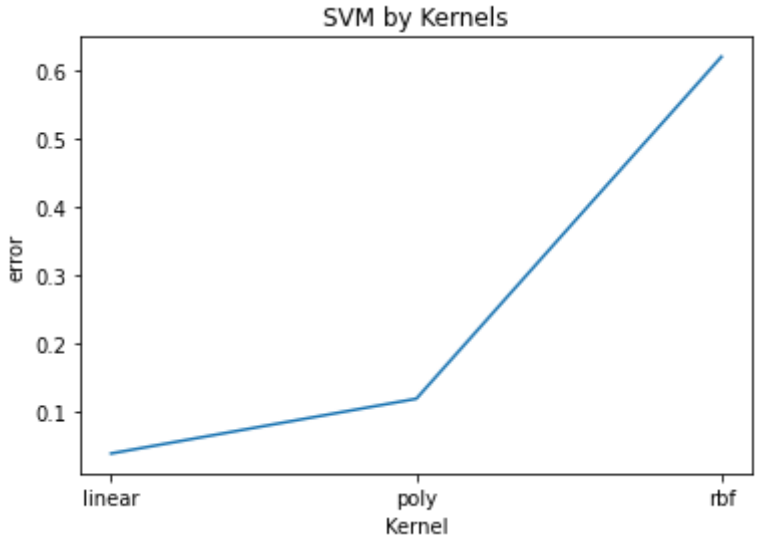
Homework Assignment 4

Due: March 17, 11:59pm, 2023

Using shuffled indices S, the first 50 elements create the training subsets x_train and y_train, while the remaining elements form the validation subsets x_validation and y_validation. Splitting the data helps prevent overfitting and provides better model performance estimates. We train a linear SVM model with various C values, evaluate their validation set performance, and plot the relationship between C values and errors. By choosing the C value with the lowest error (2-4) , we can optimize our SVM model as demonstrated in the figure.



In step 4, we train an SVM models with several kernel types while evaluating how well they perform on the validation set and illustrates the correlation between different kernel types and errors. In this case, our ideal kernel type for the SVM model is linear with the C vale of 2.



This has returned the confusion matrix, in where we obtain the model average accuracy of 94%.

```
Confusion Matrix:
[[42  6]
 [ 0 52]]
Average Accuracy: 0.94
Per-Class Precision: [1.         0.89655172]
Per-Class Recall: [0.875 1.   ]
```

All in all, based on the model's errors, it is evident that most of the errors were classified as false negatives rather than false positives. Below is a data frame containing five successful classifications and five incorrect signals.

	Y	Yhat	Error
37	1.0	-1.0	FALSE NEGATIVE
31	1.0	-1.0	FALSE NEGATIVE
84	1.0	-1.0	FALSE NEGATIVE
14	1.0	-1.0	FALSE NEGATIVE
60	1.0	-1.0	FALSE NEGATIVE
...	...	...	...
28	-1.0	-1.0	NO ERROR
27	1.0	1.0	NO ERROR
26	1.0	1.0	NO ERROR
50	1.0	1.0	NO ERROR
99	-1.0	-1.0	NO ERROR

100 rows × 3 columns