Maniya Yash Rajeshbhai
B20CS033

ML Lab 10 Report

Question 1 :

1. Methodology :

- I have used the sklearn library for using the SVM model which uses C-Support Vector Classification.
- The implementation is based on libsvm. The fit time scales at least quadratically with the number of samples, hence it takes considerable time to train.
- However our dataset has 4601 data points in total, of which 70% of the dataset (3220 samples) will be used to train the SVM.
- For the linear kernel, I have also used LinearSVM of the same library to train as it takes considerably less time to train.

2. Training SVM and Testing :

a. Varying Kernel :

First we tried to use different kernels to see which one fits best with our dataset and gives best accuracy on the testing set.

We keep C=1 (default) and vary kernels :

- Linear Kernel :

```
Linear Kernel SVM
Training Accuracy : 0.9227
Testing Accuracy : 0.9319
```

- Quadratic Kernel :

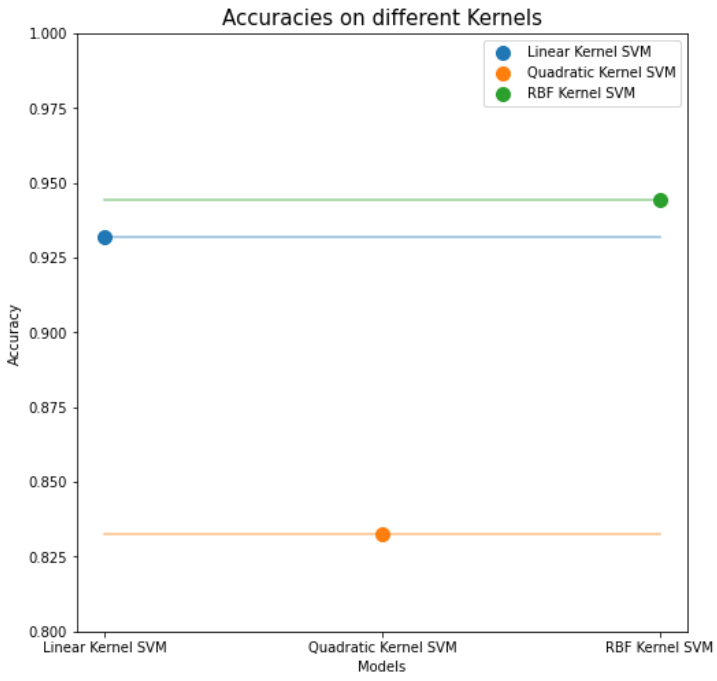
```
Quadratic Kernel SVM
Training Accuracy : 0.8556
Testing Accuracy : 0.8327
```

- RBF Kernel :

```
RBF Kernel SVM
Training Accuracy : 0.9441
Testing Accuracy : 0.9442
```

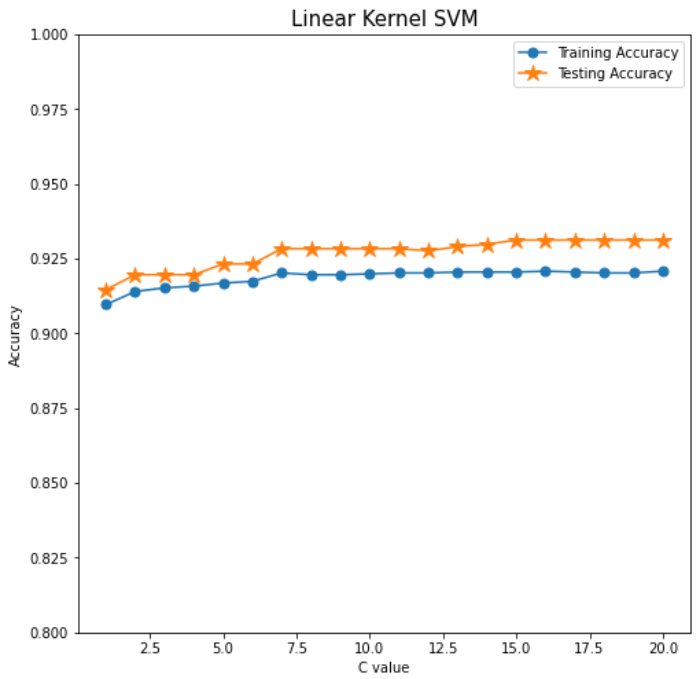
Comparing the three kernels, we find that the RBF Kernel gives maximum accuracy among all the trained models.

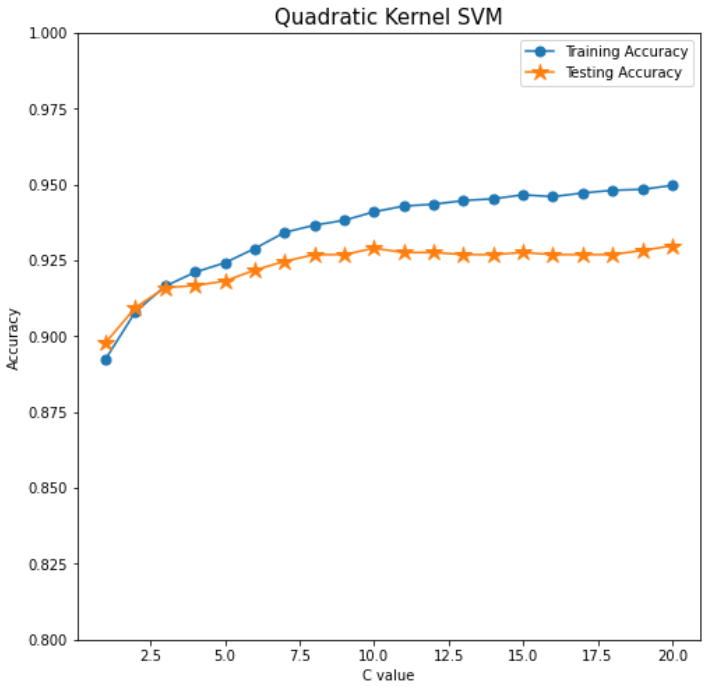
	Models	Training Accuracy	Testing Accuracy
0	Linear Kernel SVM	0.9227	0.9319
1	Quadratic Kernel SVM	0.8556	0.8327
2	RBF Kernel SVM	0.9441	0.9442

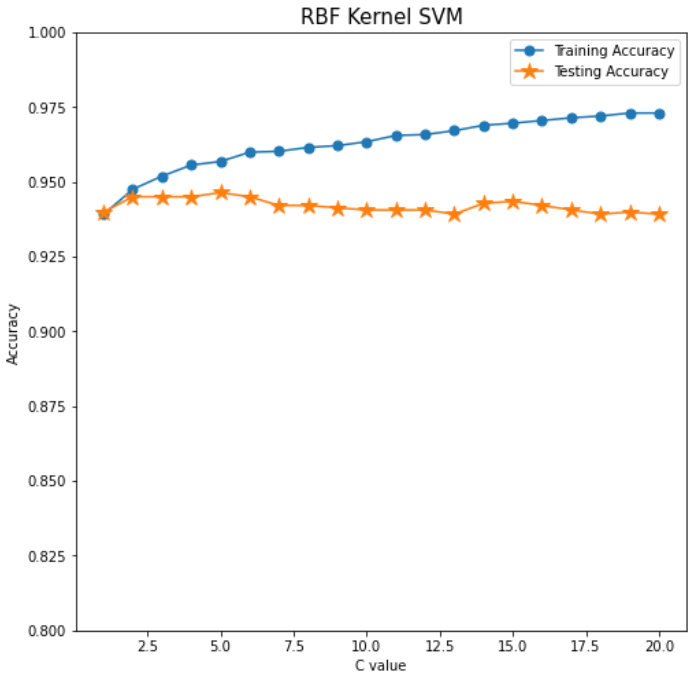


b. Varying C :

- On varying **C** (regularisation parameter), we find that on all the models, Training Accuracy increases however Testing accuracy starts to saturate or decrease.
- So we can say that higher values of **C** can lead to overfitting, especially for RBF kernel SVM.







	C value	Linear Kernel SVM	Quadratic Kernel SVM	RBF Kernel SVM
0	1	0.9146	0.8979	0.9399
1	2	0.9196	0.9095	0.9450
2	3	0.9196	0.9160	0.9450
3	4	0.9196	0.9167	0.9450
4	5	0.9232	0.9182	0.9464
5	6	0.9232	0.9218	0.9450
6	7	0.9283	0.9247	0.9421
7	8	0.9283	0.9269	0.9421
8	9	0.9283	0.9269	0.9413
9	10	0.9283	0.9290	0.9406
10	11	0.9283	0.9276	0.9406
11	12	0.9276	0.9276	0.9406
12	13	0.9290	0.9269	0.9392
13	14	0.9298	0.9269	0.9428
14	15	0.9312	0.9276	0.9435
15	16	0.9312	0.9269	0.9421

We get maximum Testing Accuracy on following **C** values :

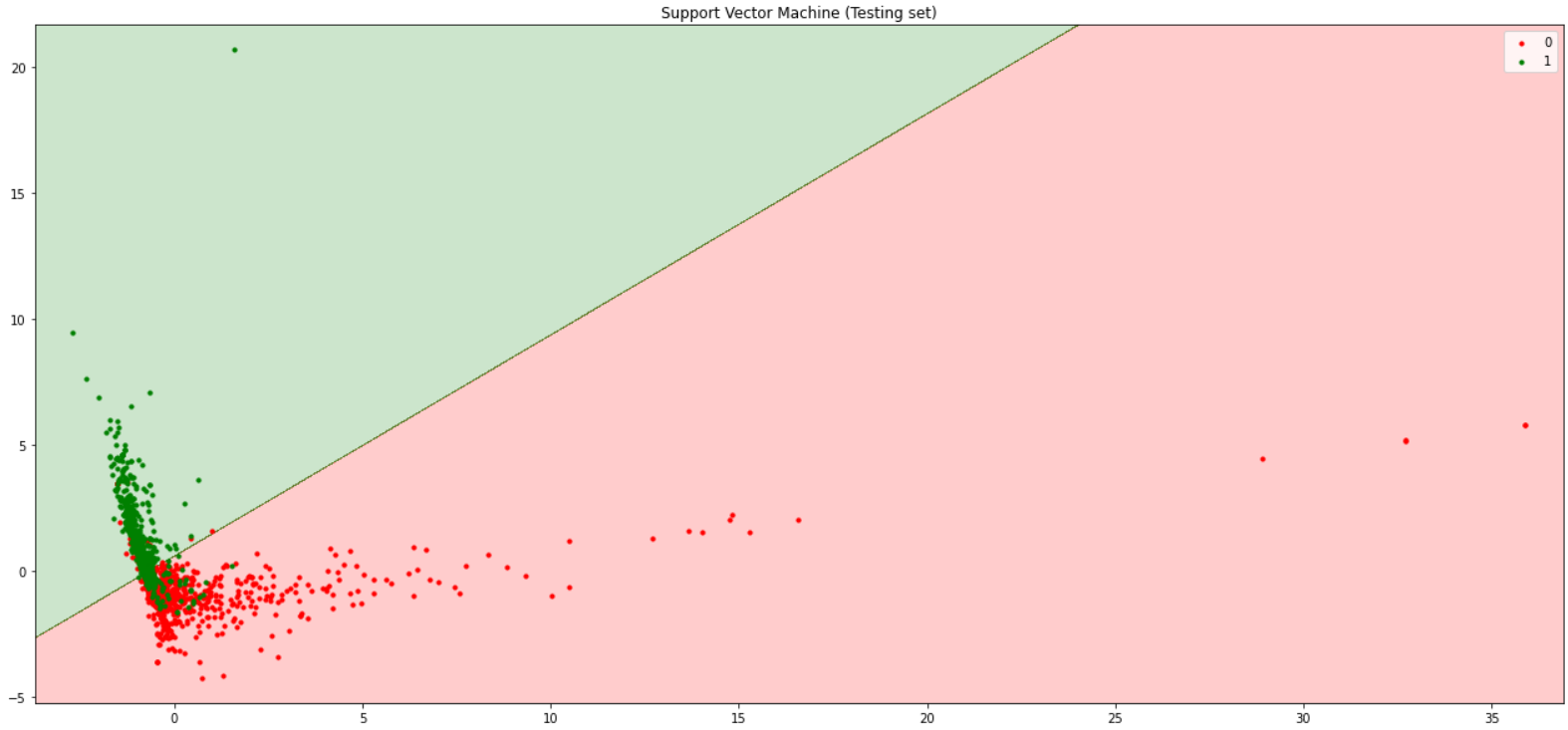
```
Metrics for SVM with different Kernels :

Linear Kernel SVM
Max Testing Accuracy : 0.9312
C value : 15

Quadratic Kernel SVM
Max Testing Accuracy : 0.9298
C value : 20

RBF Kernel SVM
Max Testing Accuracy : 0.9464
C value : 5
```

Here is the separation of the classes by SVM with linear kernel :



End of the Report !