Assignment – 4

Sai Krishna Prasad Kurapati

Sxk190015@utdallas.edu

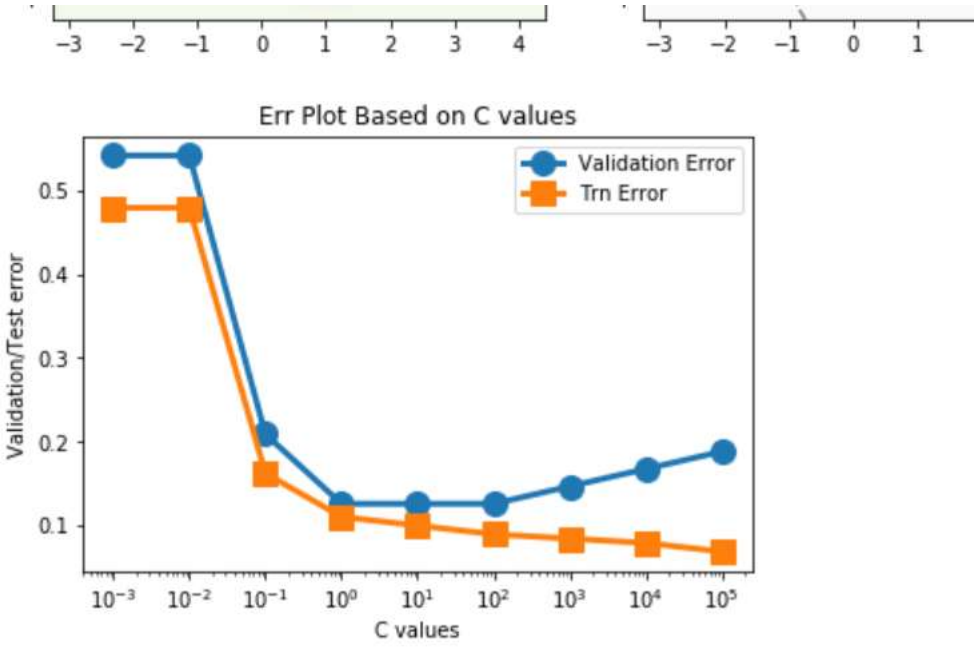
1st Question

With synthetic paraments - Here there are two features

a) Changing of C parameter

Best value of gamma is at 1.0 and the accuracy on the test set is 83.33

Plotting of errors is as follows



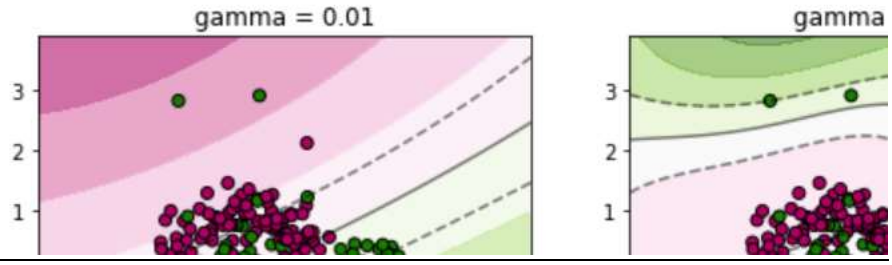
At C value = 1, we could see that the error difference is minimum over here

Discussion part (a): As we can see that here at (10pow0) i.e at "1" the error difference is less over here. And by on changing the "C" value we could see that the error is decreasing on the training set whereas on the validation set it is increasing. By visualizing the above models, we could see that the model is getting more complex as C value is increasing. The C value has to be stopped at 10pow0 even though the training error is decreasing, the validation error is increasing.

b) Changing the gamma value:

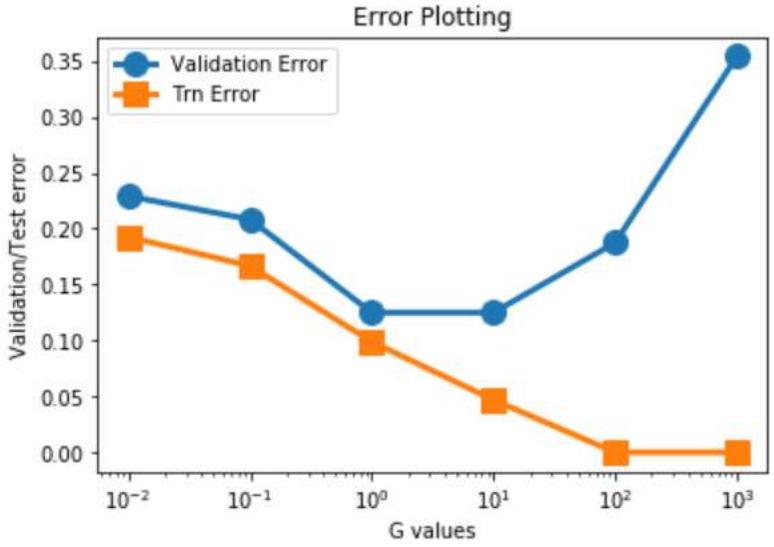
```
print("Accuracy on Test Set is:",acc*100)
```

```
Best Value of Gamma: 1.0
Accuracy on Test Set is: 83.33333333333334
```



Here also from the above result we can see that the gamma value of 1.0 performs better with the accuracy of 83.33%

Plotting of errors is as follows



Discussion part (b): As we can see that here at (10pow0) i.e at "1" the error difference is less over here. And on changing the gamma value ,we could see that the error is decreasing on the training set whereas on the validation set it is increasing.

By visualizing the above models, we could see that the model is getting more complex as gamma value is increasing. The gamma value has to be stopped at 10pow0 even though the training error is decreasing, the validation error is **drastically** increasing.

2) Breast Cancer Diagnosis with SVM

Have extracted the data from the given files and result is as follows

```
Loading Training DataSet
Y_trn : (339,)
X_trn: (339, 30)
Loading Testing DataSet
Y_tst : (115,)
X_tst: (115, 30)
Loading Validation DataSet
Y_val : (115,)
X_val: (115, 30)
```

Training Error Table

	C-G	Error Values
0	(0.01, 0.001)	0.371681
1	(0.01, 0.01)	0.371681
2	(0.01, 0.1)	0.371681
3	(0.01, 1.0)	0.371681
4	(0.01, 10.0)	0.371681
5	(0.01, 100.0)	0.371681
6	(0.1, 0.001)	0.306785
7	(0.1, 0.01)	0.050147
8	(0.1, 0.1)	0.035398
9	(0.1, 1.0)	0.371681
10	(0.1, 10.0)	0.371681
11	(0.1, 100.0)	0.371681
12	(1.0, 0.001)	0.047198
13	(1.0, 0.01)	0.029499
14	(1.0, 0.1)	0.011799
15	(1.0, 1.0)	0.000000
16	(1.0, 10.0)	0.000000
17	(1.0, 100.0)	0.000000
18	(10.0, 0.001)	0.026549
19	(10.0, 0.01)	0.011799
20	(10.0, 0.1)	0.000000
21	(10.0, 1.0)	0.000000
22	(10.0, 10.0)	0.000000
23	(10.0, 100.0)	0.000000
24	(100.0, 0.001)	0.014749
25	(100.0, 0.01)	0.002950
26	(100.0, 0.1)	0.000000
27	(100.0, 1.0)	0.000000
28	(100.0, 10.0)	0.000000
29	(100.0, 100.0)	0.000000
30	(1000.0, 0.001)	0.005900
31	(1000.0, 0.01)	0.000000
32	(1000.0, 0.1)	0.000000
33	(1000.0, 1.0)	0.000000
34	(1000.0, 10.0)	0.000000
35	(1000.0, 100.0)	0.000000
36	(10000.0, 0.001)	0.000000
37	(10000.0, 0.01)	0.000000
38	(10000.0, 0.1)	0.000000
39	(10000.0, 1.0)	0.000000
40	(10000.0, 10.0)	0.000000
41	(10000.0, 100.0)	0.000000

Validation Error Table

	C-G	ErrorValues
0	(0.01, 0.001)	0.373913
1	(0.01, 0.01)	0.373913
2	(0.01, 0.1)	0.373913
3	(0.01, 1.0)	0.373913
4	(0.01, 10.0)	0.373913
5	(0.01, 100.0)	0.373913
6	(0.1, 0.001)	0.304348
7	(0.1, 0.01)	0.069565
8	(0.1, 0.1)	0.078261
9	(0.1, 1.0)	0.373913
10	(0.1, 10.0)	0.373913
11	(0.1, 100.0)	0.373913
12	(1.0, 0.001)	0.060870
13	(1.0, 0.01)	0.060870
14	(1.0, 0.1)	0.043478
15	(1.0, 1.0)	0.373913
16	(1.0, 10.0)	0.373913
17	(1.0, 100.0)	0.373913
18	(10.0, 0.001)	0.034783
19	(10.0, 0.01)	0.043478
20	(10.0, 0.1)	0.034783
21	(10.0, 1.0)	0.373913
22	(10.0, 10.0)	0.373913
23	(10.0, 100.0)	0.373913
24	(100.0, 0.001)	0.034783
25	(100.0, 0.01)	0.026087
26	(100.0, 0.1)	0.034783
27	(100.0, 1.0)	0.373913
28	(100.0, 10.0)	0.373913
29	(100.0, 100.0)	0.373913
30	(1000.0, 0.001)	0.034783
31	(1000.0, 0.01)	0.026087
32	(1000.0, 0.1)	0.034783
33	(1000.0, 1.0)	0.373913
34	(1000.0, 10.0)	0.373913
35	(1000.0, 100.0)	0.373913
36	(10000.0, 0.001)	0.026087
37	(10000.0, 0.01)	0.026087
38	(10000.0, 0.1)	0.034783
39	(10000.0, 1.0)	0.373913
40	(10000.0, 10.0)	0.373913
41	(10000.0, 100.0)	0.373913

And from the tables after calculating the absolute difference between the errors, the best C-gamma combination(0.01,0.001) has been selected

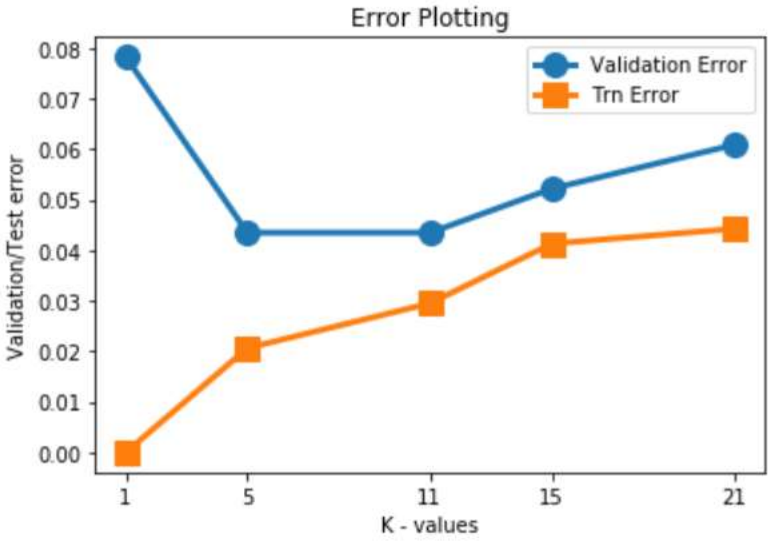
And the accuracy on test set for the selected model is as follows

```
Best combination of C-Gamma: (0.01, 0.001)
Accuracy on Test Set is: 62.60869565217392
```

We could see that we got the less accuracy

3) Breast Cancer Diagnosis with KNN

After extracting the data from the given files, the model has been trained for different values and the Training Error and the validation Error are calculated are plotted as below



The best k- value has been selected from the given k values by calculating the absolute difference between the errors and the k value is 15

```
Best K value: 15
Accuracy on Test Set is: 94.78260869565217
```

We got an accuracy of 94.78%

Discussion: Here we got good accuracy in the KNN about 94% when compared to that of the SVM, it indicates that the given data is not easily separable using the decision planes. Basically it depends the type of the data set provided