ECS308/658: Assignment 3 Report

DEEP POOJA
17074

1 Objective

1. To experiment with the use of SVMs for a multiclass classification problem, and understand the effects of varying various parameters therein.

2 Binary Classification

2.1 choosing the binary classes to be [0,1] out of 10 classes 0 to 9

2.1.1 Considering First 10 Features of Data Only

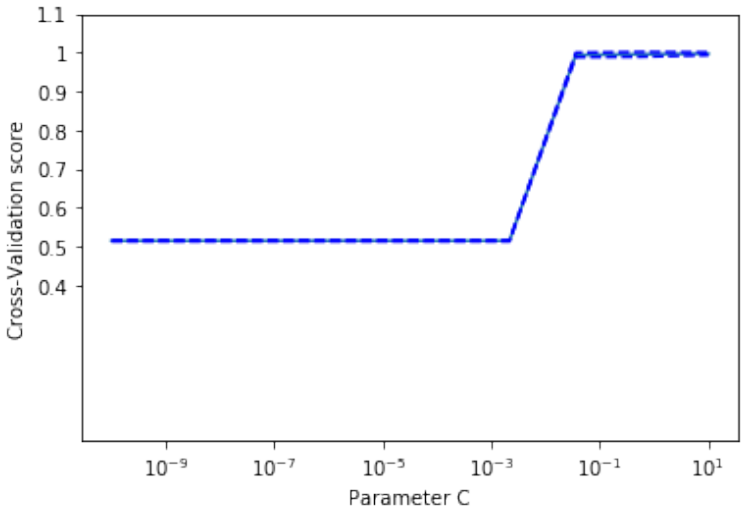
- **Studying the effect of parameter C**

  The C parameter trades off correct classification of training examples against maximization of the decision function's margin. For larger values of C, a smaller margin will be accepted if the decision function is better at classifying all training points correctly. A lower C will encourage a larger margin, therefore a simpler decision function, at the cost of training accuracy. In other words"C" behaves as a regularization parameter in the SVM

  Large Value of parameter C => small margin
  Small Value of paramerter C => Large margin

  The larger the value of C, the smaller the regularization. Therefore, the larger the value of C, the stronger chances of overfitting.

  Fixing **Kernel = RBF**, changing kernel to LINEAR or POLYNOMIAL does not show any difference in plot generated below by varying the parameter C.



  In this setting, the optimal value of C found out to be r between $0.1 - 10$

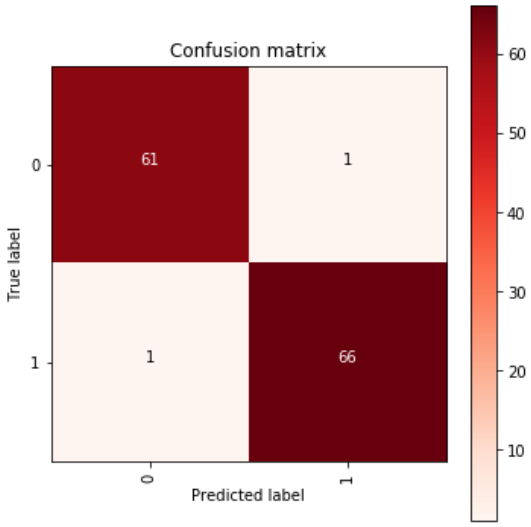
- Studying the effect of Kernels

  Linear Classifier

	precision	recall	f1-score	support
0	0.98	0.98	0.98	62
1	0.99	0.99	0.99	67
accuracy			0.98	129
macro avg	0.98	0.98	0.98	129
weighted avg	0.98	0.98	0.98	129

Confusion Matrix

```
Confusion matrix, without normalization
[[61  1]
 [ 1 66]]
```

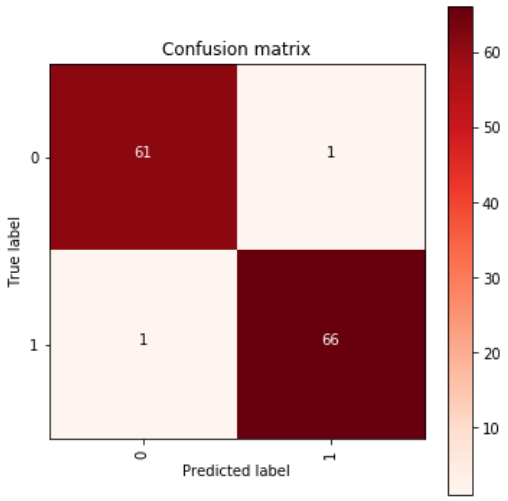


Polynomial(deg=2 or 3) classifier

	precision	recall	f1-score	support
0	0.98	0.98	0.98	62
1	0.99	0.99	0.99	67
accuracy			0.98	129
macro avg	0.98	0.98	0.98	129
weighted avg	0.98	0.98	0.98	129

Confusion Matrix

```
Confusion matrix, without normalization
[[61  1]
 [ 1 66]]
```



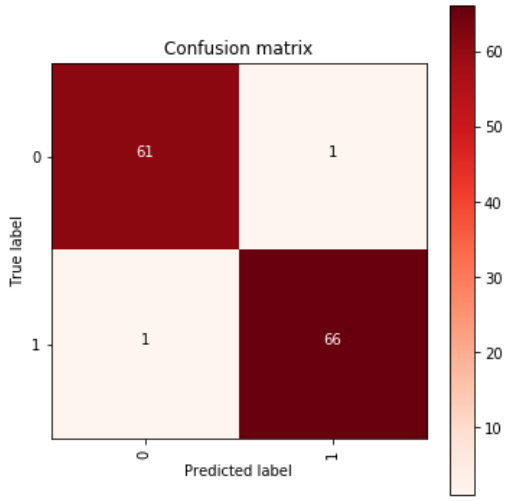
Note: For higher degree > 3, accuracy or score decreases, hence suggesting not a good model for this setting

RBF Classifier

	precision	recall	f1-score	support
0	0.98	0.98	0.98	62
1	0.99	0.99	0.99	67
accuracy			0.98	129
macro avg	0.98	0.98	0.98	129
weighted avg	0.98	0.98	0.98	129

Confusion Matrix

```
Confusion matrix, without normalization
[[61  1]
 [ 1 66]]
```

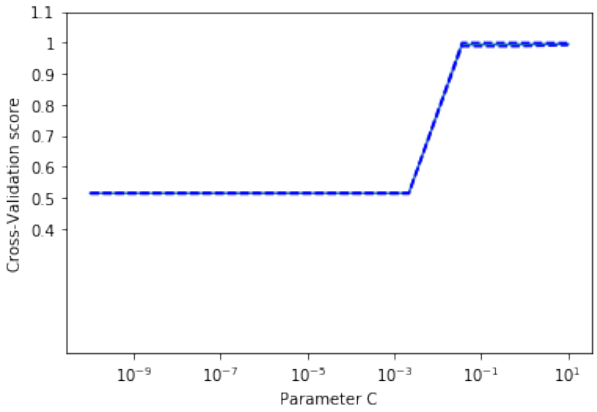


- **Observation:** All Kernels are equally good in classify in this setting.

S.No.	Kernel	Score/Accuracy
1.	Linear	0.9844961240310077
2.	Polynomial(deg$deg = 3$)	0.9844961240310077
3.	RBF	0.9844961240310077

2.1.2 Considering All 25 features

- The effect of parameter C
  The choice of C remains the same.

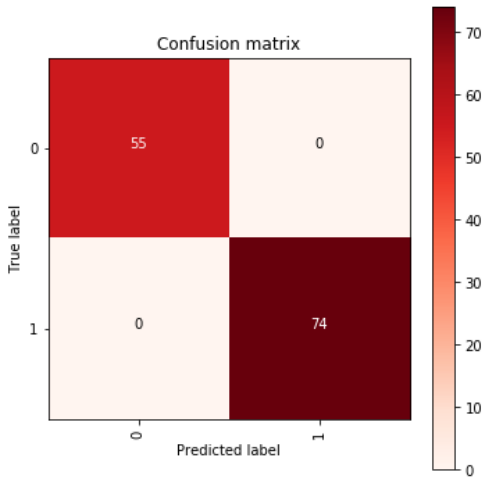


- Effect of Kernels

  Linear/Poly(deg=2)

```
Linear Classifier report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        55
           1       1.00      1.00      1.00        74

    accuracy                           1.00       129
   macro avg       1.00      1.00      1.00       129
weighted avg       1.00      1.00      1.00       129

Polynomial Classifier report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        55
           1       1.00      1.00      1.00        74

    accuracy                           1.00       129
   macro avg       1.00      1.00      1.00       129
weighted avg       1.00      1.00      1.00       129
```

  Confusion Matrix

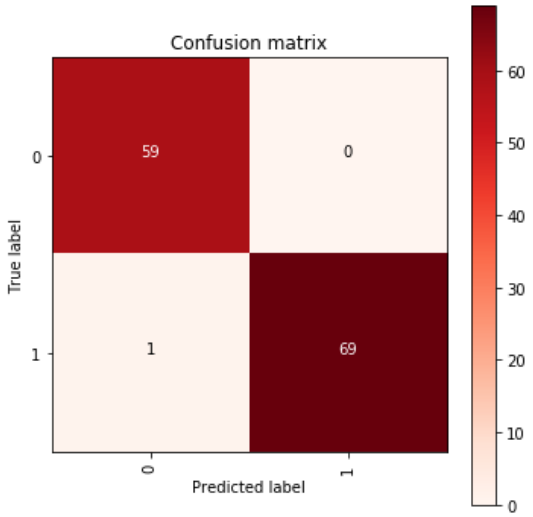


  Note that for higher degree ($deg > 2$)of polynomial accuracy decreases steeply
  rbf

```
Polynomial Classifier report:
              precision    recall  f1-score   support

           0       0.98      1.00      0.99        59
           1       1.00      0.99      0.99        70

    accuracy                           0.99       129
   macro avg       0.99      0.99      0.99       129
weighted avg       0.99      0.99      0.99       129
```

  Confusion Matrix

```
Confusion matrix, without normalization
[[59  0]
 [ 1 69]]
```



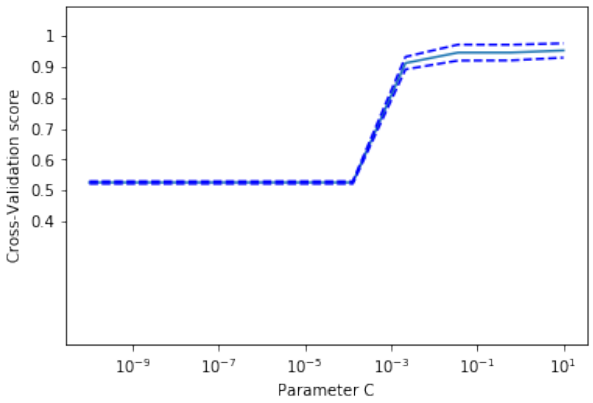
- **Observation:** Linear and polynomial with degree 2 has classified with 100% accuracy when we included all the features(25)and RBF has had one misclassification in binary classification as [0 or 1] classes. With polynomial of degree greater than 2, score decreases steeply, hence not a good model in this setting.

S.No.	Kernel	Score/Accuracy
1.	Linear	1.0
2.	Polynomial(deg$deg = 2$)	1.0
3.	RBF	0.9922480620155039

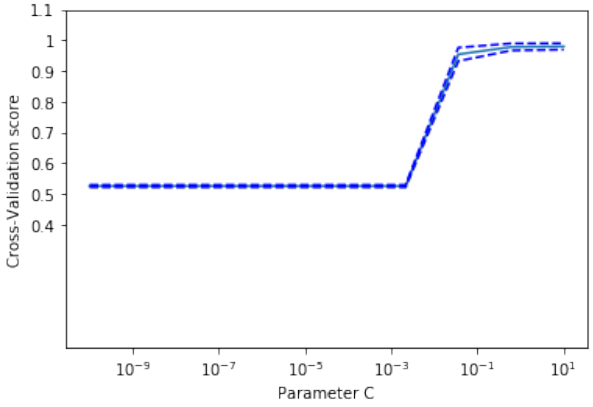
2.2 choosing the classes to be [8,9] out of 10 classes 0 to 9

2.2.1 First 10 features only

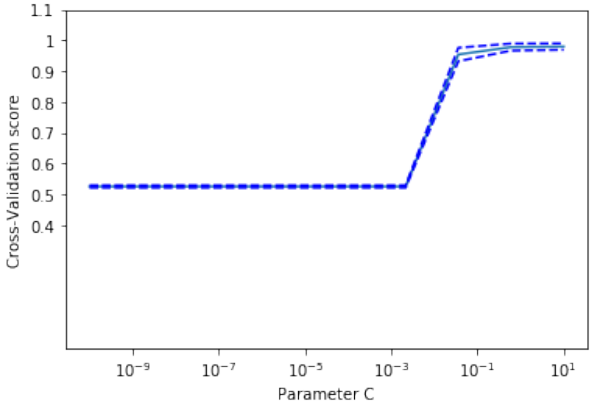
- effect of parameter C
  Linear



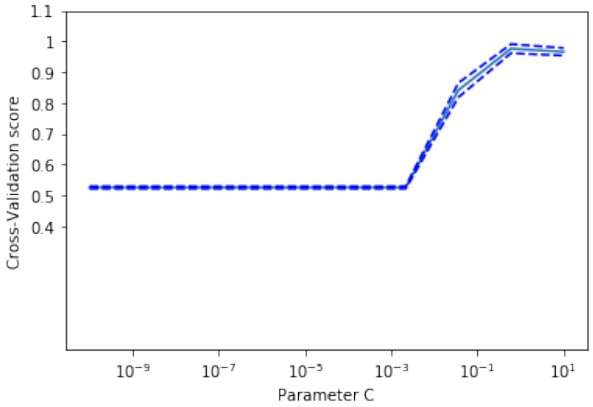
RBF



Poly (deg = 2)



poly (deg 3)



Observation: There is little variation in parametric plot of C w.r.t kernel, but we can safely choose the optimum value of C between 1 and 10.

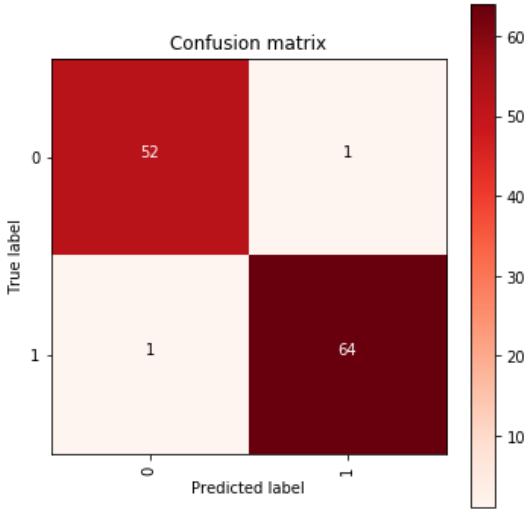
- fixing the value of C= 10 and studying the effect of Kernel

Linear/RBF

Confusion Matrix

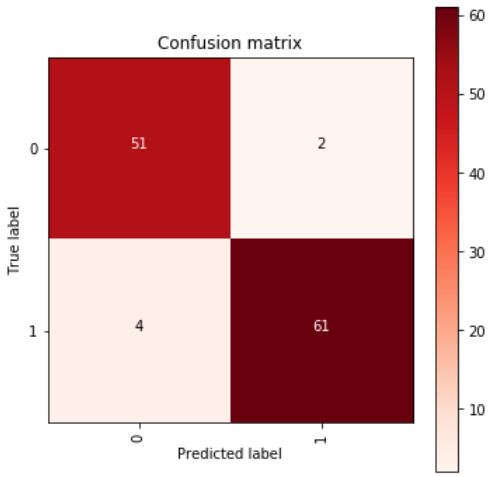
```
Linear Classifier report:
              precision    recall  f1-score   support

           8       0.98      0.98      0.98        53
           9       0.98      0.98      0.98        65

    accuracy                           0.98       118
   macro avg       0.98      0.98      0.98       118
weighted avg       0.98      0.98      0.98       118

RBF Classifier report:
              precision    recall  f1-score   support

           8       0.98      0.98      0.98        53
           9       0.98      0.98      0.98        65

    accuracy                           0.98       118
   macro avg       0.98      0.98      0.98       118
weighted avg       0.98      0.98      0.98       118
```

```
Confusion matrix, without normalization
[[52  1]
 [ 1 64]]
```



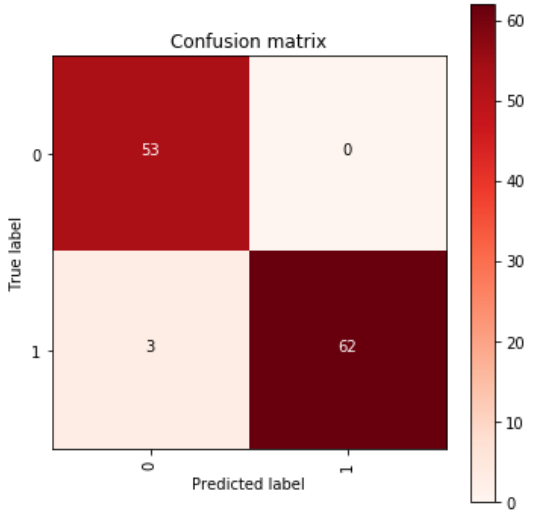
Polynomial

```
Confusion matrix, without normalization
[[51  2]
 [ 4 61]]
```



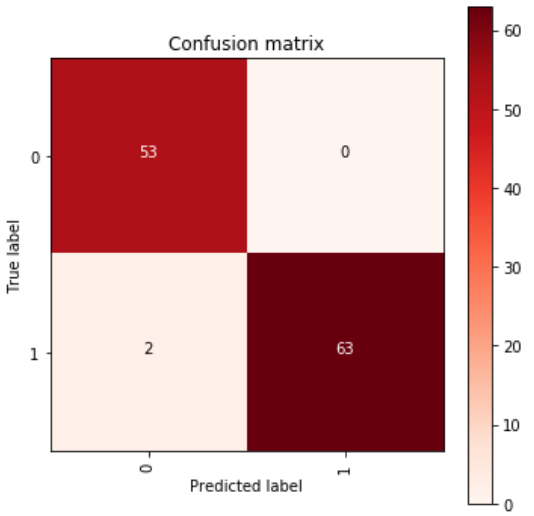
deg= 2

```
Confusion matrix, without normalization
[[53  0]
 [ 3 62]]
```



deg= 3

```
Confusion matrix, without normalization
[[53  0]
 [ 2 63]]
```



deg= 5

- Observations: Linear and RBF kernel has same performance as polynomial with degree = 5, and it's the best performance we could get on binary classification where labels are 8 and 9.

S.No.	Kernel	Score/Accuracy
1.	Linear	0.9830508474576272
2.	Polynomial(deg$deg = 2$)	0.9491525423728814
3.	Polynomial(deg$deg = 3$)	0.9745762711864406
4.	Polynomial(deg$deg = 4$)	0.9491525423728814
5.	Polynomial(deg$deg = 5$)	0.9830508474576272
6.	Polynomial(deg$deg = 7$)	0.8898305084745762
7.	RBF	0.9830508474576272

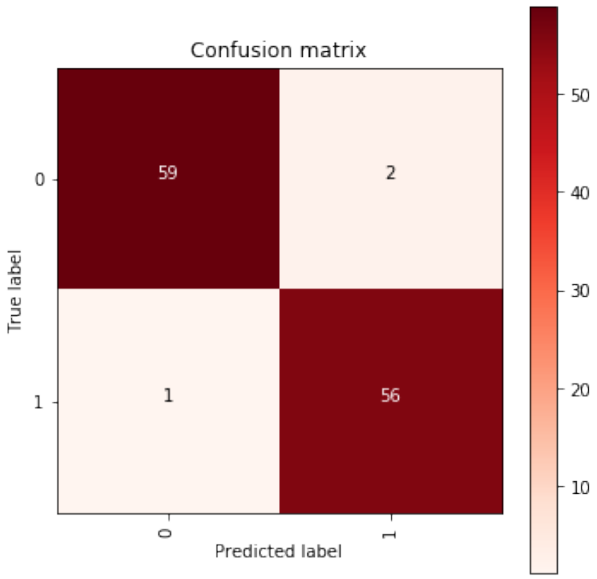
2.2.2 Considering all Features (25)

- Effect of parameter C
  - for linear kernel C= 0.1 is optimum choice
  - for RBF kernel, we found C= 1.0 as optimum value

- Effect of kernel
  Linear/RBF

```
Linear Classifier report:
              precision    recall  f1-score   support

           8       0.98      0.97      0.98        61
           9       0.97      0.98      0.97        57

    accuracy                           0.97       118
   macro avg       0.97      0.97      0.97       118
weighted avg       0.97      0.97      0.97       118

RBF Classifier report:
              precision    recall  f1-score   support

           8       0.98      0.97      0.98        61
           9       0.97      0.98      0.97        57

    accuracy                           0.97       118
   macro avg       0.97      0.97      0.97       118
weighted avg       0.97      0.97      0.97       118
```

Cofusion Matrix



S.No.	Kernel	Score/Accuracy
1.	Linear	0.9745762711864406
6.	RBF	0.9745762711864406

Polynomial Kernel

S.No.	Kernel	C	Score/Accuracy
2.	Polynomial(deg$deg = 2$)	1	0.9915254237288136
3.	Polynomial(deg$deg = 3$)	1	0.9745762711864406
4.	Polynomial(deg$deg = 4$)	10	0.9830508474576272
5.	Polynomial(deg$deg = 5$)	10	0.9661016949152542

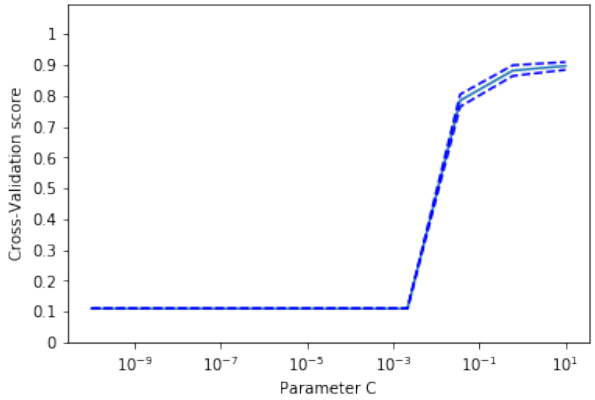
- **Observations:**
  - Changing the classes from [0 or 1] to [8 or 9], there is only slight variations in parameter settings i.e it does not varies a lot.
  - If we try to fit the data with higher degree of polynomial, we observe that we constantly need to increase the regularization parameter C to avoid overfitting.
  - Adding more features does affect the accuracy, in Zero-one classification accuracy increased by 0.1 to 0.2%, where as in eight-nine classification setting accuracy drop by 0.1%.
  - On the line of above observation, we can conclude that some features are more prominent in certain classification settings than other.

3 Multiclass classification

3.1 with 10 Features only

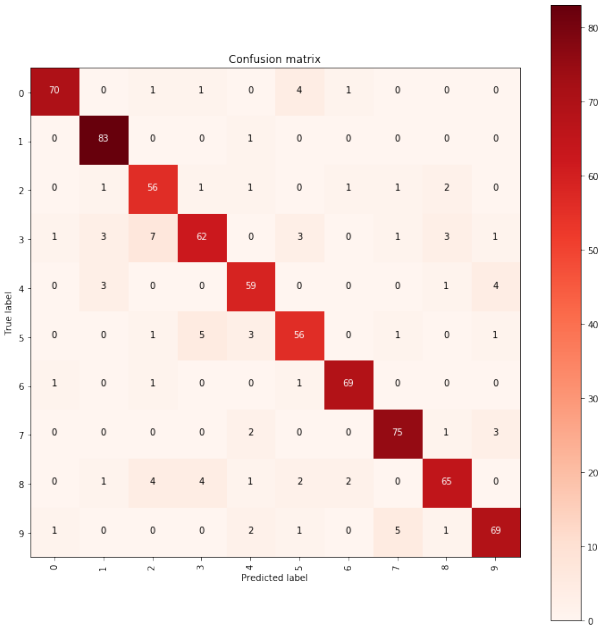
RBF Kernel

- Effect of parameter C



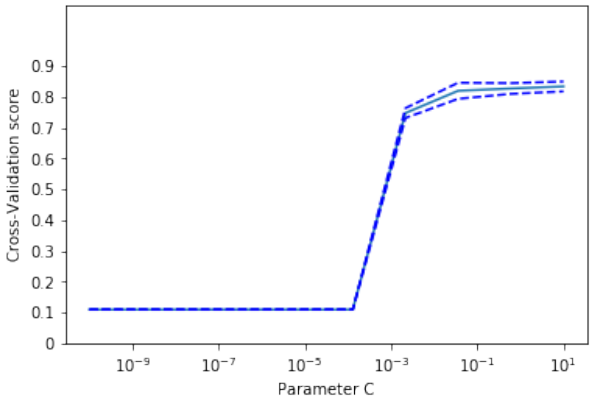
choosing the value of C to be 22 in this setting (by tinkering)
Confusion Matrix

	precision	recall	f1-score	support
0	0.96	0.91	0.93	77
1	0.91	0.99	0.95	84
2	0.80	0.89	0.84	63
3	0.85	0.77	0.81	81
4	0.86	0.88	0.87	67
5	0.84	0.84	0.84	67
6	0.95	0.96	0.95	72
7	0.90	0.93	0.91	81
8	0.89	0.82	0.86	79
9	0.88	0.87	0.88	79
accuracy			0.89	750
macro avg	0.88	0.88	0.88	750
weighted avg	0.89	0.89	0.88	750



Linear Kernel

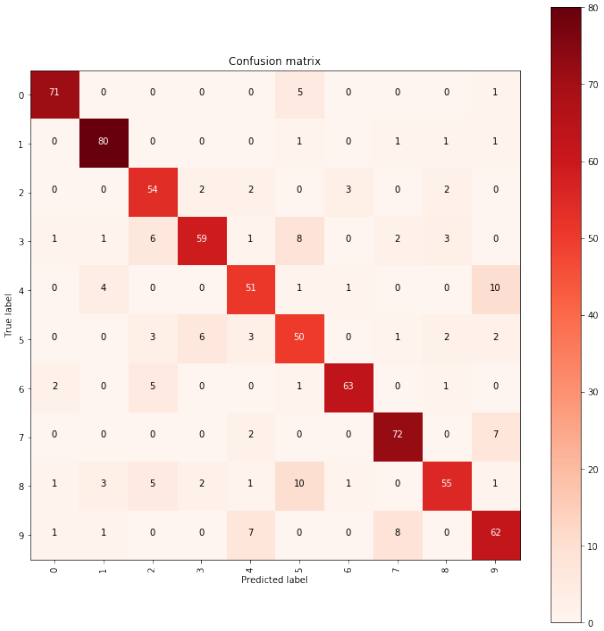
Effect of parameter C



choosing the value of C to be 10 in this setting (by tinkering)

Confusion Matrix

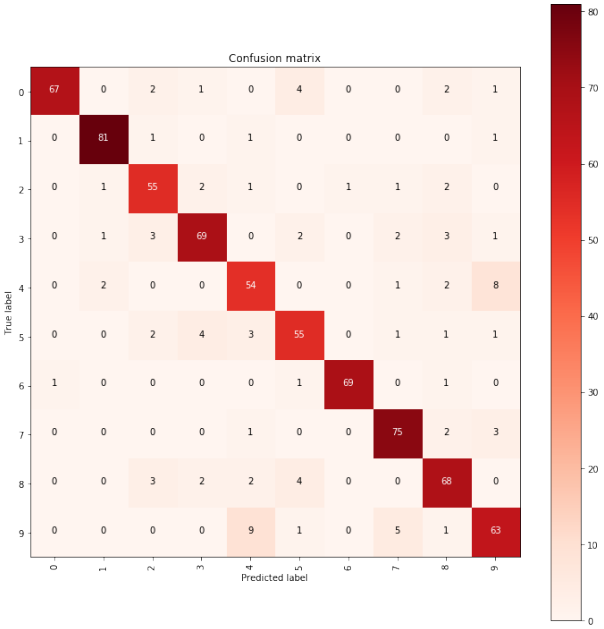
	precision	recall	f1-score	support
0	0.95	0.95	0.95	84
1	0.94	1.00	0.97	85
2	0.86	0.93	0.90	69
3	0.87	0.74	0.80	84
4	0.93	0.89	0.91	71
5	0.79	0.88	0.84	69
6	0.94	0.89	0.91	66
7	0.91	0.91	0.91	68
8	0.84	0.83	0.84	65
9	0.87	0.89	0.88	89
accuracy			0.89	750
macro avg	0.89	0.89	0.89	750
weighted avg	0.89	0.89	0.89	750



Polynomial Kernel

Confusion Matrix

	precision	recall	f1-score	support
0	0.99	0.87	0.92	77
1	0.95	0.96	0.96	84
2	0.83	0.87	0.85	63
3	0.88	0.85	0.87	81
4	0.76	0.81	0.78	67
5	0.82	0.82	0.82	67
6	0.99	0.96	0.97	72
7	0.88	0.93	0.90	81
8	0.83	0.86	0.84	79
9	0.81	0.80	0.80	79
accuracy			0.87	750
macro avg	0.87	0.87	0.87	750
weighted avg	0.88	0.87	0.88	750



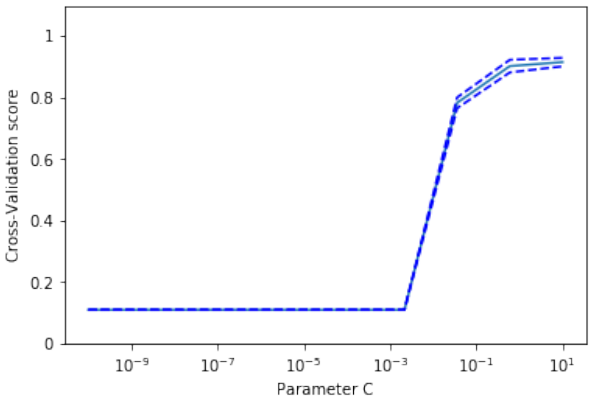
S.No.	Kernel	C	Score/Accuracy
1.	RBF	22	0.8853333333333333
3.	Linear	10	0.8226666666666667
4.	Polynomial(deg$deg = 2$)	10	0.8346666666666667
4.	Polynomial(deg$deg = 3$)	10	0.8746666666666667
5.	Polynomial(deg$deg = 5$)	50	0.8466666666666667

- **Observation:**

3.2 Considering all Features

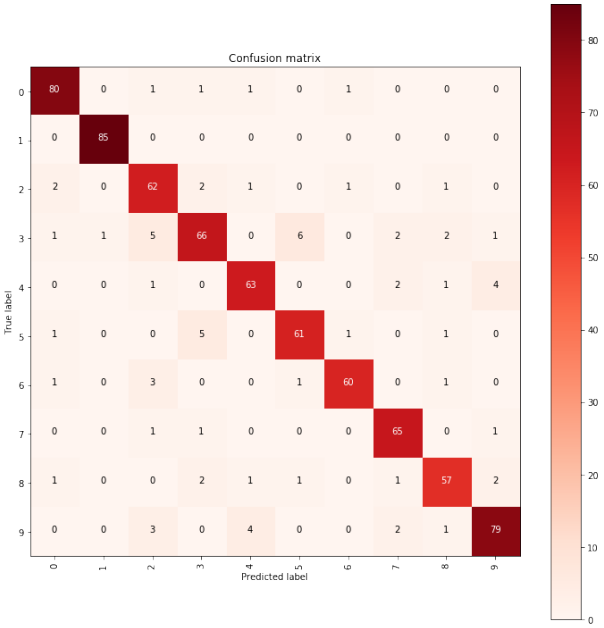
RBF

Effect of C



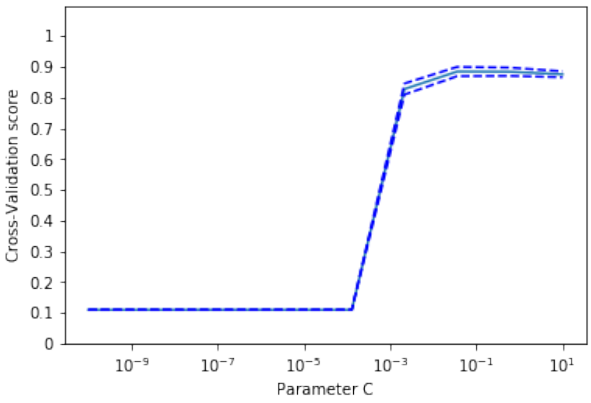
Confusion Matrix

	precision	recall	f1-score	support
0	0.93	0.95	0.94	84
1	0.99	1.00	0.99	85
2	0.82	0.90	0.86	69
3	0.86	0.79	0.82	84
4	0.90	0.89	0.89	71
5	0.88	0.88	0.88	69
6	0.95	0.91	0.93	66
7	0.90	0.96	0.93	68
8	0.89	0.88	0.88	65
9	0.91	0.89	0.90	89
accuracy			0.90	750
macro avg	0.90	0.90	0.90	750
weighted avg	0.90	0.90	0.90	750



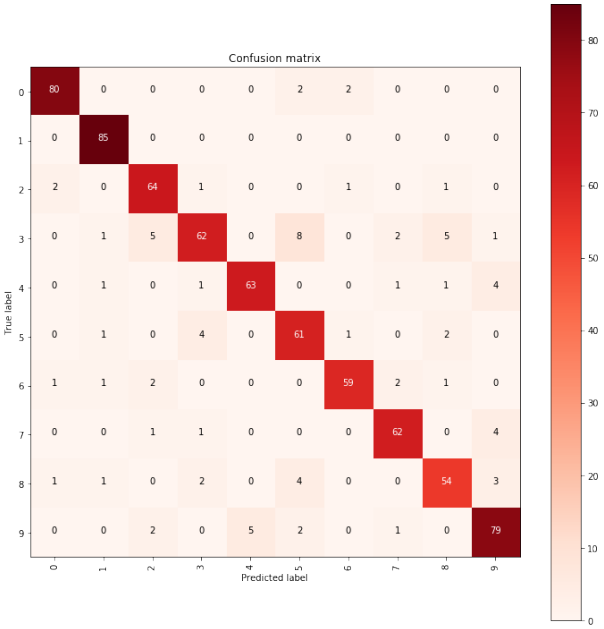
Linear

Effect of C



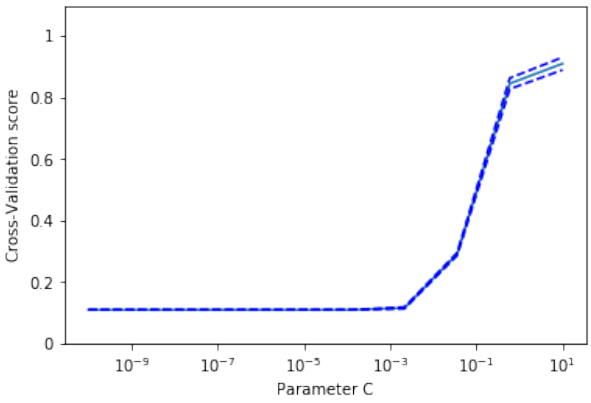
Confusion Matrix

	precision	recall	f1-score	support
0	0.95	0.95	0.95	84
1	0.94	1.00	0.97	85
2	0.86	0.93	0.90	69
3	0.87	0.74	0.80	84
4	0.93	0.89	0.91	71
5	0.79	0.88	0.84	69
6	0.94	0.89	0.91	66
7	0.91	0.91	0.91	68
8	0.84	0.83	0.84	65
9	0.87	0.89	0.88	89
accuracy			0.89	750
macro avg	0.89	0.89	0.89	750
weighted avg	0.89	0.89	0.89	750



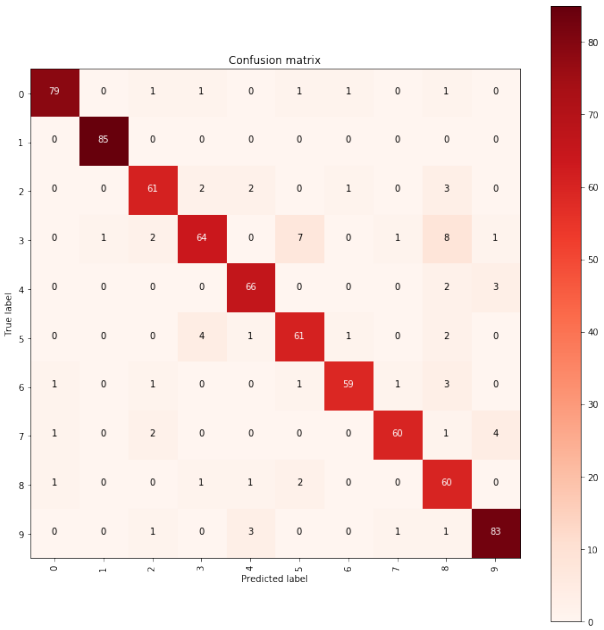
Polynomial Kernel

Effect of C



Confusion Matrix

	precision	recall	f1-score	support
0	0.96	0.94	0.95	84
1	0.99	1.00	0.99	85
2	0.90	0.88	0.89	69
3	0.89	0.76	0.82	84
4	0.90	0.93	0.92	71
5	0.85	0.88	0.87	69
6	0.95	0.89	0.92	66
7	0.95	0.88	0.92	68
8	0.74	0.92	0.82	65
9	0.91	0.93	0.92	89
accuracy			0.90	750
macro avg	0.90	0.90	0.90	750
weighted avg	0.91	0.90	0.90	750



S.No.	Kernel	C	Score/Accuracy
1.	RBF	25	0.904
3.	Linear	1	0.892
4.	Polynomial(deg$deg = 2$)	2	0.9026666666666666
4.	Polynomial(deg$deg = 3$)	10	0.904
5.	Polynomial(deg$deg = 5$)	500	0.8666666666666667

Observation: Adding more features leads to increase in accuracy in multi-class classification.This might tell us having relevant features which might not seems useful at first sight may play vital role in achieving good accuracy on our models

In comparison to Binary classification, Multi-class classification required higher value of regularization parameter C, that seems obvious because in multi-class classification there is more chances of overfitting.

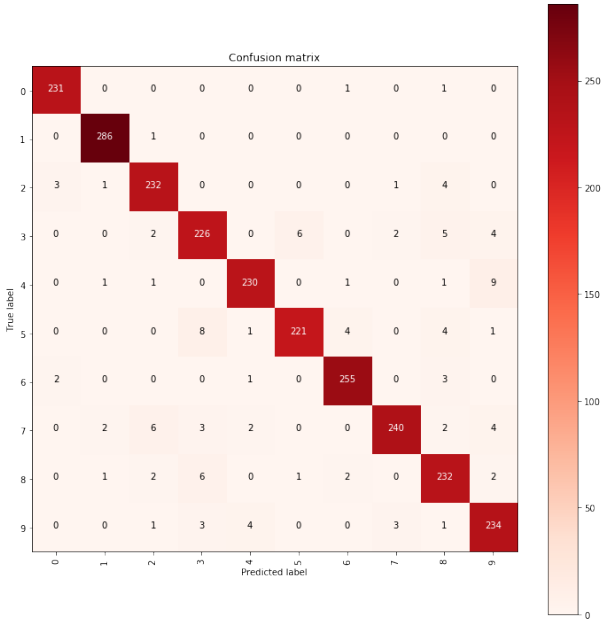
4 Conclusion

- In Binary classification with 0 and 1 as class labels, best accuracy achieved is 98% with 10 features and 100% with all features.
- In Binary classification with 8 and 9 as class labels, best accuracy achieved is 98% with 10 features and 99% with all features.
- In Multi-class classification, best accuracy achieved is 88% with RBF kernel with 10 features only and 90% with poly(deg=3) as well as rbf (tie) in case of considering all features.
- Important point to note that is in binary classification almost all kernels are able to achieve similar accuracy (obviously with different regularization settings), but in multi-class classification higher degree polynomial(5) doesn't able to achieve similar accuracy even if very high regularization.

5 (Part 2)

On the line of above experimentation, We choose our model for kaggle part is either rfb or poly with degree 3.

Confusion Matrix



	precision	recall	f1-score	support
0	0.98	0.99	0.99	233
1	0.98	1.00	0.99	287
2	0.95	0.96	0.95	241
3	0.92	0.92	0.92	245
4	0.97	0.95	0.96	243
5	0.97	0.92	0.95	239
6	0.97	0.98	0.97	261
7	0.98	0.93	0.95	259
8	0.92	0.94	0.93	246
9	0.92	0.95	0.94	246
accuracy			0.95	2500
macro avg	0.95	0.95	0.95	2500
weighted avg	0.96	0.95	0.95	2500

Accuracy achieved in both the cases is 95.5%