Question 1

1. C is the slack parameter. It controls the weights or the emphasis given to slack variables. When we train the model by taking C=1 and C=100, we see that taking large values of C will reduce or eliminate the number of misclassified samples. This is because when we increase the value of C, we reduce the margin to ensure least number of misclassified samples. Smaller C makes the margin from the decision boundary considerably larger.

Linear kernel was used for different values of variable C. The accuracies for the different cases were as follows:

|  |  |  |
| --- | --- | --- |
| Index | Value of C | Accuracy of the model |
| 1 | 1 | 0.9 |
| 2 | 2 | 1 |
| 3 | 10 | 1 |
| 4 | 100 | 1 |

The decision boundaries plotted for C=1 and C=100 are as given below:

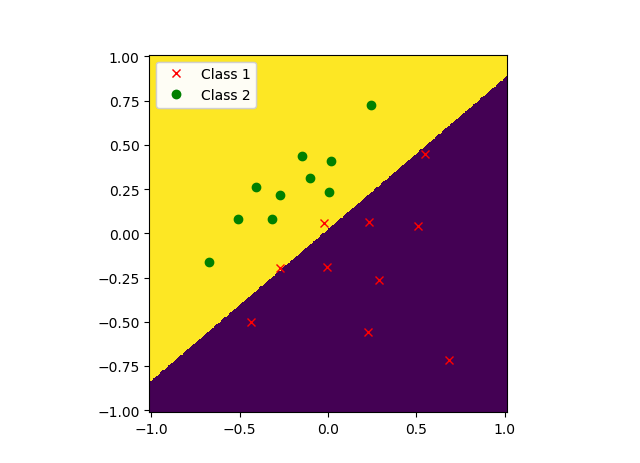


Figure 1 : C = 1

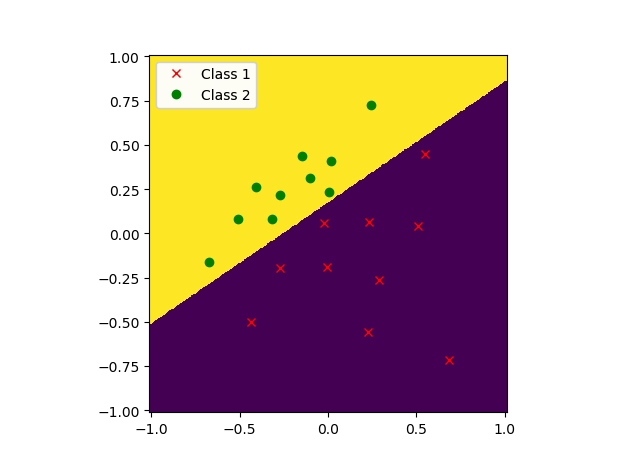
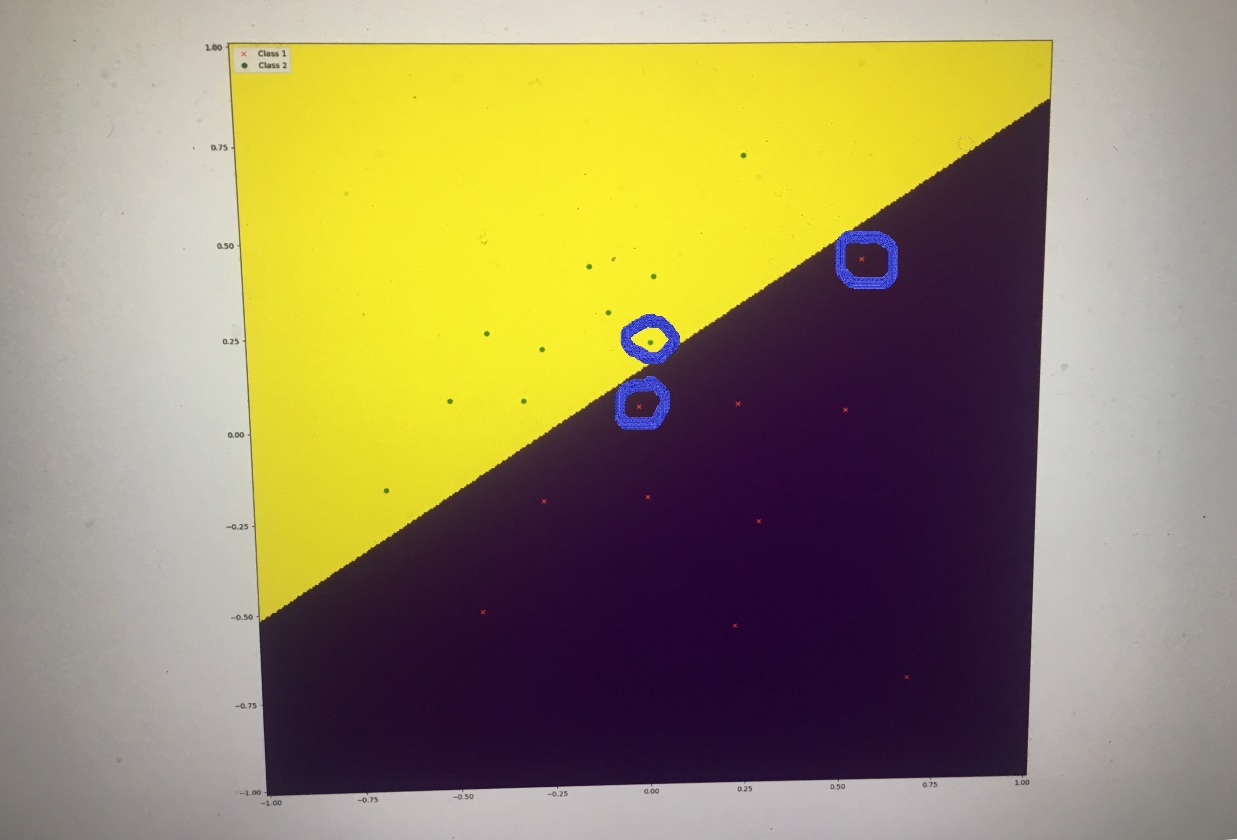


Figure 2 : C =100

From the figures, it is seen that the margin is larger for smaller value of C leading to some misclassified points. When we increase the value of C, we reduce the margin to accommodate the slack variables, thus reducing misclassified points.

1. For C=100, the support vectors are marked by the 3 blue circles on the image.



Numerically, the coordinates of these support vectors are:

1. [ 0.0065 , 0.2339 ]
2. [-0.0239 , 0.0604 ]
3. [ 0.5458 , 0.4503 ]

The weight vectors are:

[ W1 W2 ] = [ -7.11966384 , 10.40264821 ]

W0 = -1.7984

The equation of the decision boundary is :

G(x) => W0 + W1\*X1 + W2\*X2 = 0

=> -1.7984 – 7.11966384 \* X1 + 10.40264821 \* X2 = 0;

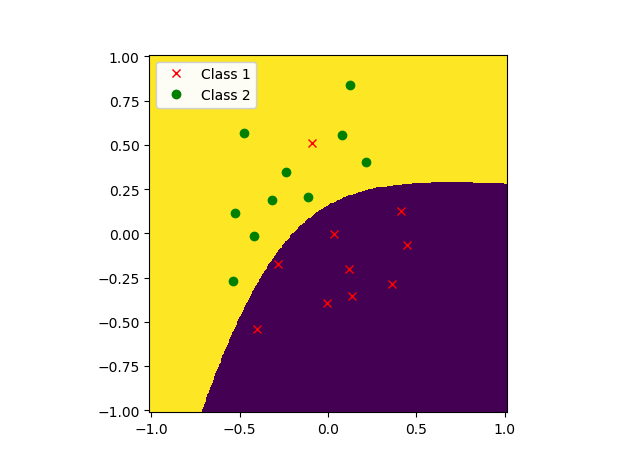
1. Substituting the support vector points in the equation of G(x) => W0 + W1\*X1 + W2\*X2 = 0, we can calculate the value of G(x) for them. Using this formula, I got :

|  |  |
| --- | --- |
| Point | Value of G(X) |
| P1 | 0.5891 |
| P2 | -1 |
| P3 | -1 |

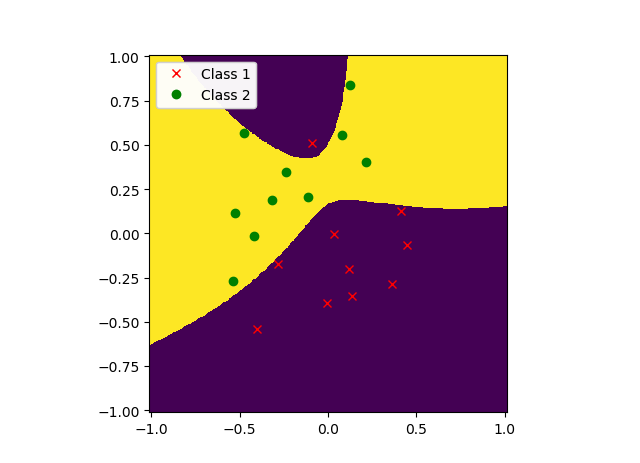
Points P2 and P3 lie on the -1 margin but P1 doesn’t lie on the +1 margin. This might be because support vector P1 corresponds to a misclassified sample.

1. For rbf kernel,

C = 50, Accuracy = 0.9499

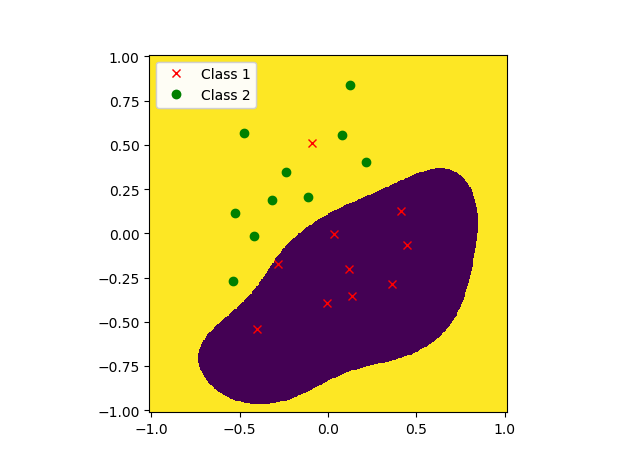


C = 5000, Accuracy = 1.0

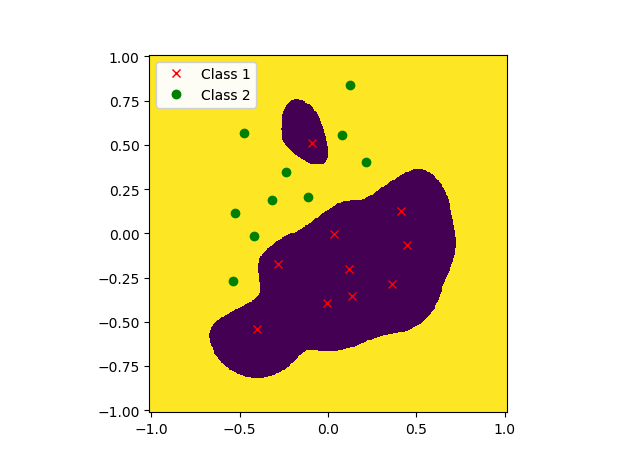


For smaller values of C, there is more tolerance towards non linearly separable case. This allows some more room for misclassification. When there’s a large value of C, we reduce the tolerance and don’t allow misclassification of datapoints by reducing the margin with the datapoints.

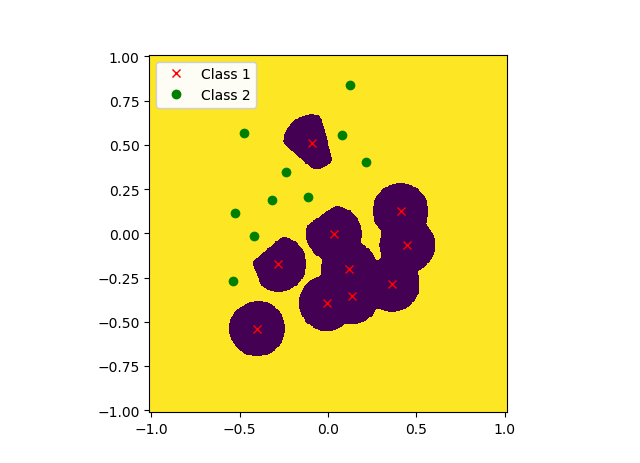
1. G = 10, Accuracy = 0.95



G = 50,Accuracy = 1



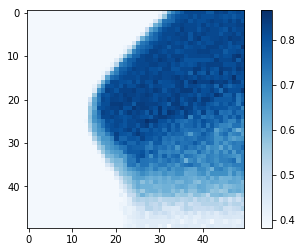
G = 500, Accuracy = 1



For a small value of G, the decision boundary is smooth and the model is more tolerant towards misclassified samples. The margin is a bit big too and the default value of C=1.0. On increasing the value of G, the boundary becomes more complex, but it correctly classifies all the data. With increase in complexity of decision boundary, there is a chance of overfitting. This is seen when G=500. This can lead to poor results on a different test set.

Question 2

1. Average Cross Validation Accuracy = 0. 81045751634
2. Figure:



The best pair is C = 2.02358964773 and gamma = 1.15139539933

The mean Cross-Validation Accuracy for the best pair = 0.866013071895

The Standard deviation for the best pair = 0.0557666117848

1. The values of the 20 chosen pairs of gammas and C are:

* [ 4.94171336e-01 4.71486636e+00]
* [ 1.67683294e-02 2.44205309e+02]
* [ 1.20679264e-01 5.96362332e+01]
* [ 2.81176870e-01 1.93069773e+01]
* [ 2.81176870e-01 8.28642773e+00]
* [ 3.72759372e-01 3.39322177e+01]
* [ 1.20679264e-01 4.49843267e+01]
* [ 3.72759372e-01 4.49843267e+01]
* [ 1.15139540e+00 2.68269580e+00]
* [ 1.59985872e-01 2.55954792e+01]
* [ 6.86648845e-02 5.96362332e+01]
* [ 2.12095089e-01 1.09854114e+01]
* [ 1.59985872e-01 1.09854114e+01]
* [ 1.59985872e-01 3.39322177e+01]
* [ 6.86648845e-02 2.55954792e+01]
* [ 2.12095089e-01 3.39322177e+01]
* [ 3.72759372e-01 1.45634848e+01]
* [ 6.55128557e-01 3.55648031e+00]
* [ 3.90693994e-02 7.54312006e+02]
* [ 2.12095089e-01 4.49843267e+01]

The best pair is gamma = 0.281176869797 and C = 8.28642772855

The mean Cross-Validation Accuracy for the best pair = 0.866666666667

The Standard deviation for the best pair = 0.0902670933848

The best chosen pair is the one where the accuracy is the maximum. A run on T=20 becomes more reproducible as the set of best pair of gamma and C repeats itself after long runs due to convergence of the pair at a repetitive value.

1. Test accuracy = 0.7865068529225844

The estimate is within approximately 1 standard deviation of the mean cross-validation accuracy from c-(II).

