Akshay Manjunath

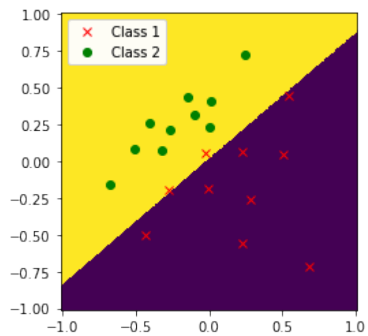
Homework 8

Problem 1

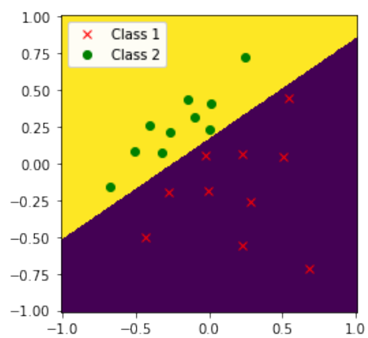
Part(a)

C is the slack parameter. It controls the weights or the emphasis given to slack variables. When we train the model by taking different values of C, 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.

Accuracy when C = 1 is 90%



Accuracy when C = 100 is 100%



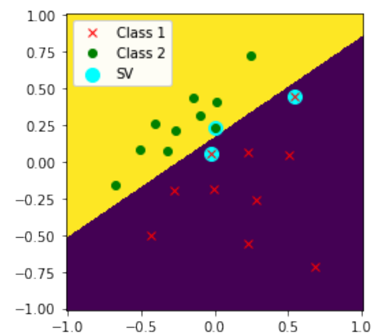
From the figures, it is seen that for C = 1 the penalty for those samples misclassified is small leading to some misclassified points. When we increase the value of C to 100, the penalty is higher and we reduce the margin to accommodate the slack variables, thus reducing misclassified points.

Part (b)

The support vectors are:
1. [-0.023855 , 0.06042]
2. [0.54579 , 0.45029]
3. [0.0064864 , 0.23394]

The weight vectors are:
W0 = -1.7983676
[W1 W2] = [-7.11966384 , 10.40264821]

The equation of the decision boundary is :
G(x) = W0 + W1*X1 + W2*X2 = 0
= -1.7983676 – 7.11966384 * X1 + 10.40264821 * X2 = 0



Part (c)

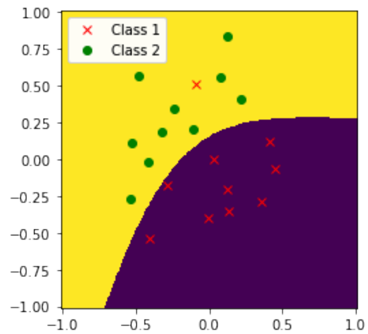
Substituting the support vector points in the equation of G(x) = W0 + W1*X1 + W2*X2 = 0, we can calculate the value of G(x) for them. Using this formula, I got :

Support Vector	G(x)
[0.0064864 , 0.23394]	0.5891
[-0.023855 , 0.06042]	-1
[0.54579 , 0.45029]	-1

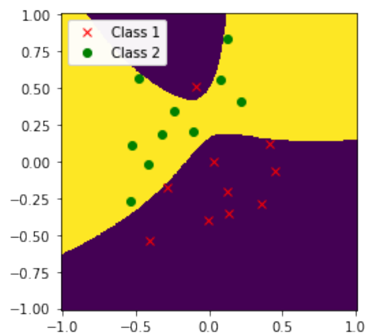
Points 2 and 3 lie on the -1 margin but point 1 doesn't lie on the +1 margin. This might be because support vector P1 corresponds to a misclassified sample.

Part (d)

For RBF Kernel, Accuracy when C = 50 is 95.0



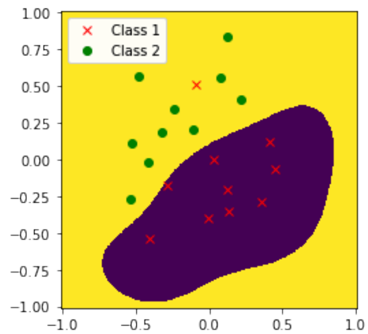
Accuracy when C = 5000 is 100.0



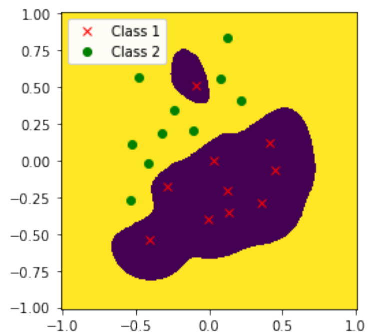
With Gaussian Kernel The boundary is nonlinear. For smaller values of C, there is more tolerance towards non linearly separable case. This allows some more room for misclassification. When there's a large value of C, we reduce the tolerance and don't allow misclassification of datapoints by reducing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

Part (e)

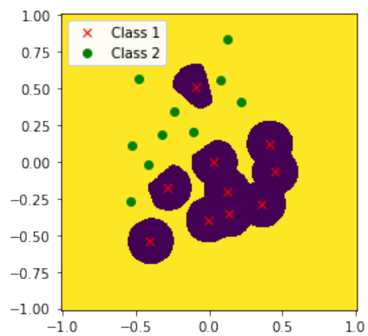
Accuracy when gamma = 10 is 95.0



Accuracy when gamma = 50 is 100.0



Accuracy when gamma = 500 is 100.0



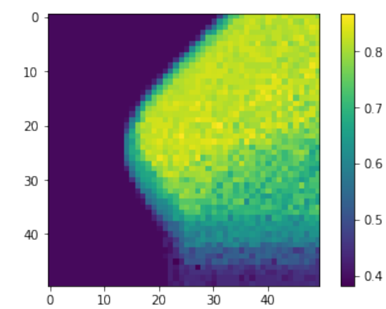
For a small value of gamma, the decision boundary is smooth and the model is more tolerant towards misclassified samples. On increasing the value of gamma it will try more to avoid misclassifying training data and the boundary becomes more complex, but it correctly classifies all the data. This is because a higher gamma gives a narrow Gaussian width. 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.

Problem 2

Part(a)

The average cross validation accuracy = 81.96078431372548

Part(b)



The best pair is C = 0.868511373751352 and gamma = 0.49417133613238334
The mean Cross-Validation Accuracy for the best pair = 0.8660130718954248
The Standard deviation for the best pair = 0.0433052778391567

Part(c)

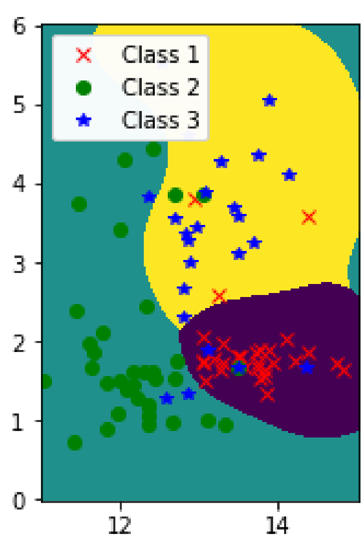
The 20 chosen pairs =
[[6.55128557e-01 1.52641797e+00]
[3.72759372e-01 4.71486636e+00]
[2.12095089e-01 4.49843267e+01]
[6.55128557e-01 1.09854114e+01]
[1.59985872e-01 3.23745754e+02]
[4.94171336e-01 1.52641797e+00]
[2.12095089e-01 1.09854114e+01]
[2.81176870e-01 1.45634848e+01]
[3.72759372e-01 2.02358965e+00]
[2.12095089e-01 3.39322177e+01]
[2.94705170e-02 2.44205309e+02]
[3.72759372e-01 3.55648031e+00]
[1.59985872e-01 1.04811313e+02]

[2.12095089e-01 1.93069773e+01]
[2.12095089e-01 1.52641797e+00]
[6.55128557e-01 1.09854114e+01]
[1.59985872e-01 2.55954792e+01]
[2.12095089e-01 1.38949549e+02]
[4.09491506e-03 1.84206997e+02]
[1.20679264e-01 1.04811313e+02]]

The best pair is C = 1.5264179671752334 and gamma = 0.49417133613238334
The mean Cross-Validation Accuracy for the best pair = 0.877124183006536
The Standard deviation for the best pair = 0.04053341629729058

The results are more reproducible and more well defined now. 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.

Part(d)



Test accuracy = 0.797752808988764
The mean Cross-Validation Accuracy = 0.877124183006536

The estimate is not within approximately 1 standard deviation of the mean cross-validation accuracy. This is because the validation accuracy is not a good predictor of the test accuracy and vice versa.