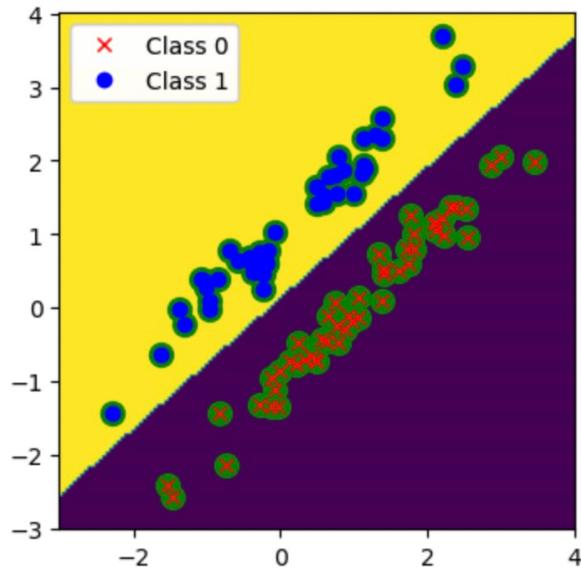


1. SVM using sklearn

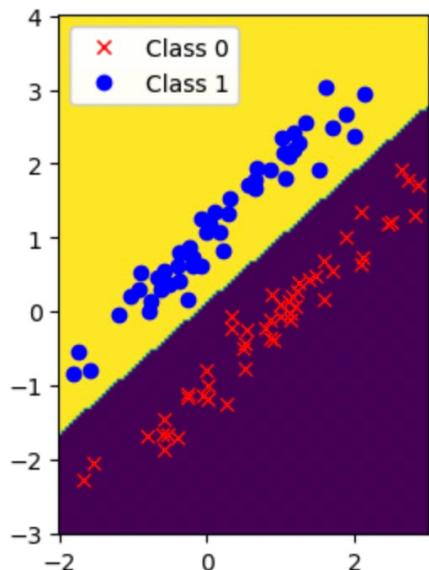
(a) When $C = 0.01$:

```
train acc = 1.0, test acc = 1.0
weight vector w = [[-0.47617753  0.53480959]]
offset w0 = [-0.06369157]
```

Train dataset1:



Test dataset1:



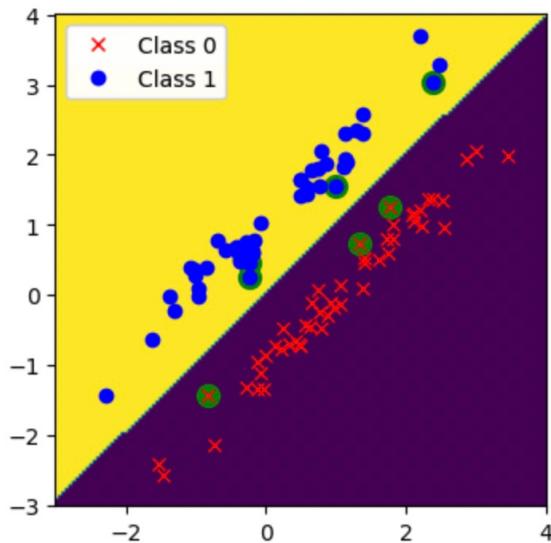
When $C = 1$:

```

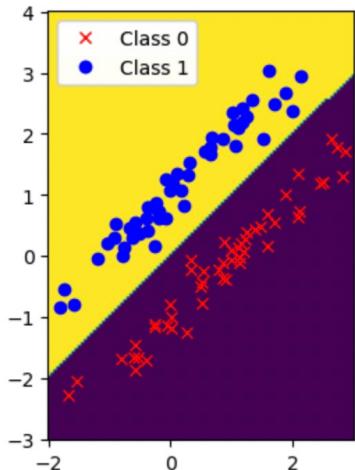
train acc = 1.0, test acc = 1.0
weight vector w = [[-1.5291679    1.54596349]]
offset w0 = [-0.04317366]

```

Train dataset1:



Test dataset1:



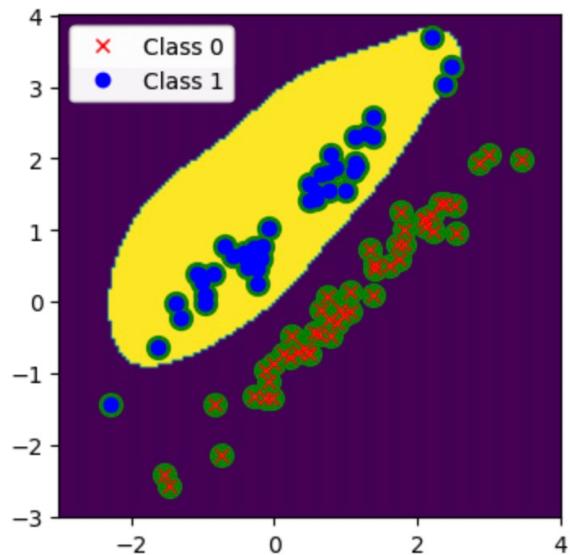
Discuss your results and explain the performance and the differences for the different values of C.

Obviously, the dataset1 can be classified 100% correct because the data is linearly separable. By plotting the support vectors, we can find that, when $C = 0.01$, every data point is considered as a support vector while there are only a few data points are considered as support vectors when $C = 1$. Because the value of C stands for a regularization parameter in the SVM. The 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.

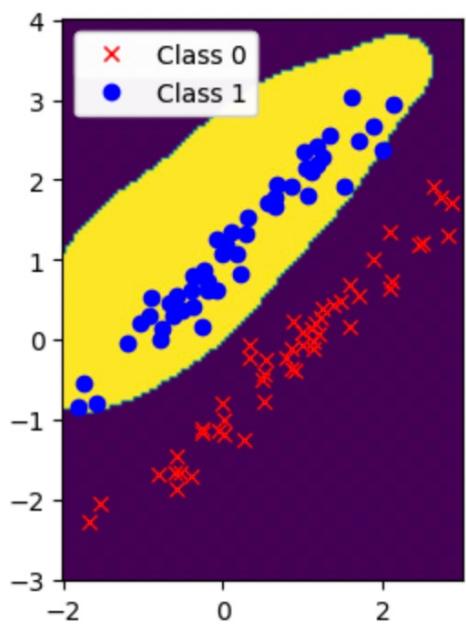
(b) $C = 0.01, \gamma = 1$

train acc = 0.99, test acc = 0.99

Train dataset1:



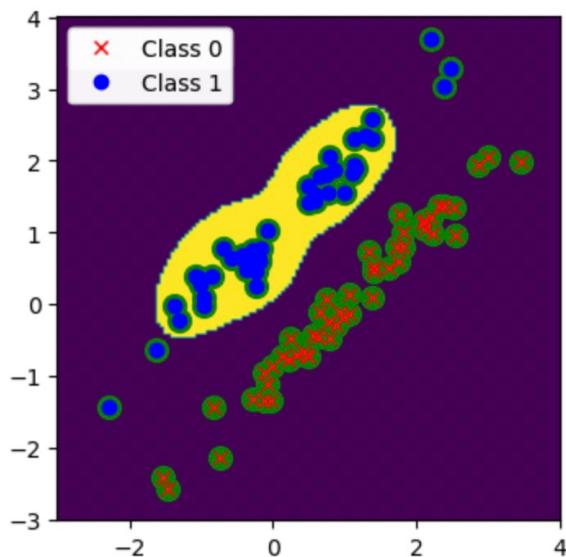
Test dataset1:



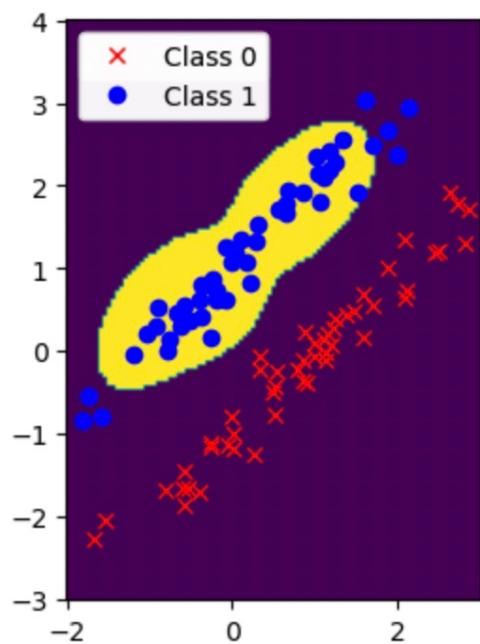
$C = 0.01, \gamma = 3$

train acc = 0.95, test acc = 0.92

Train dataset1:



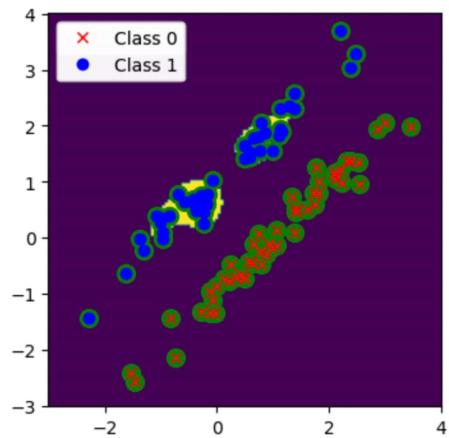
Test dataset1:



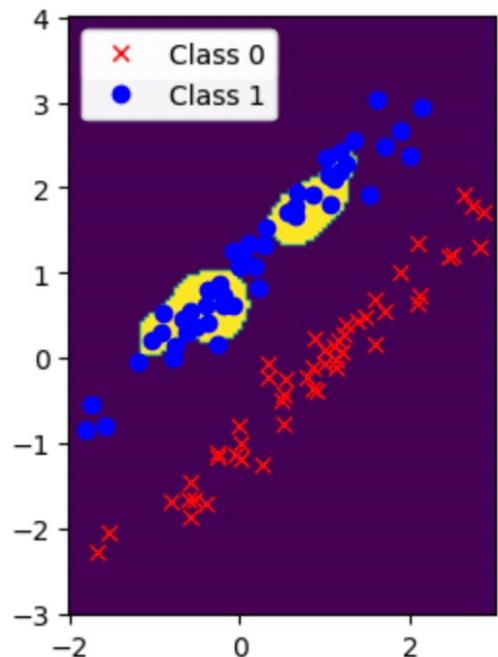
$C = 0.01, \gamma = 10$

train acc = 0.88, test acc = 0.77

Train dataset1:



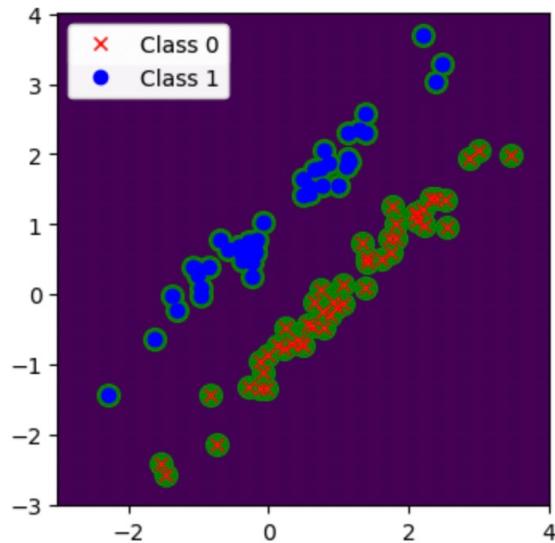
Test dataset1:



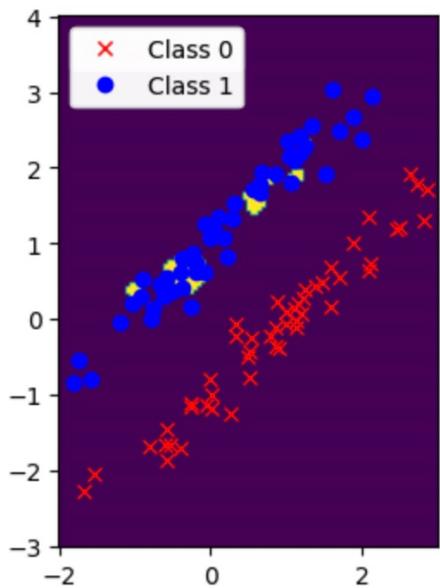
$C = 0.01, \gamma = 50$

train acc = 0.79, test acc = 0.6

Train dataset1:



Test dataset1:



Explain the linearity or nonlinearity of the decision boundary and explain the difference in decision regions for the various values of γ . State where (if anywhere) you observe underfitting or overfitting.

The decision boundary is non-linear due to the RBF kernel.

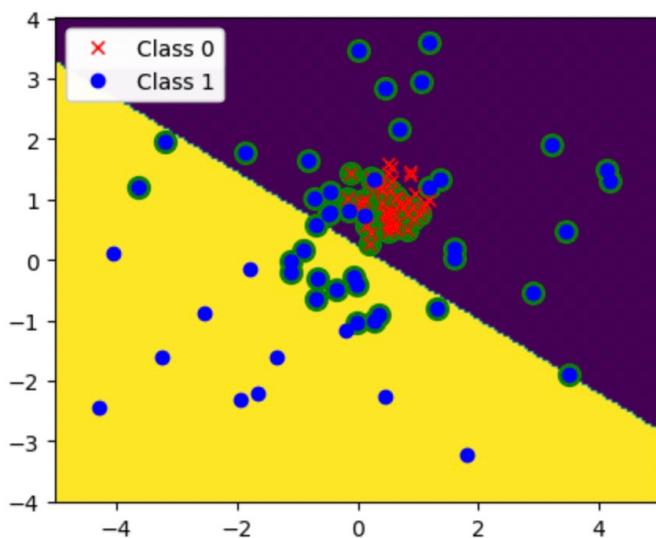
The gamma parameters can be seen as the inverse of the radius of influence of samples selected by the model as support vectors. Therefore, the larger the gamma is, the radius of the area of influence of the support vectors only includes the support vector itself. As the gamma getting larger, the decision boundary is getting closer to the support vectors.

When gamma equals and larger then 10, there exist underfitting according to the low accuracy of both train dataset and test dataset.

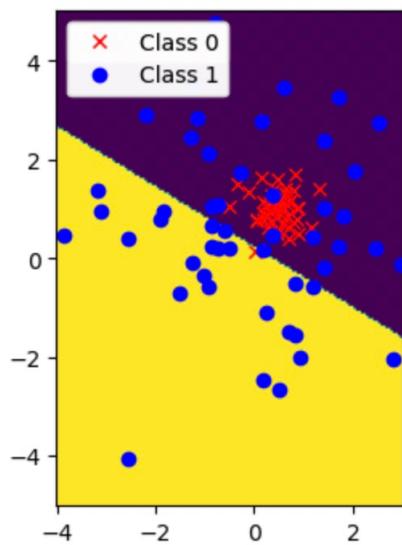
(c) $C = 1$

```
train acc = 0.76, test acc = 0.74
weight vector w =  [[-0.48069371 -0.78654222]]
offset w0 = [0.1697672]
```

Train dataset3:



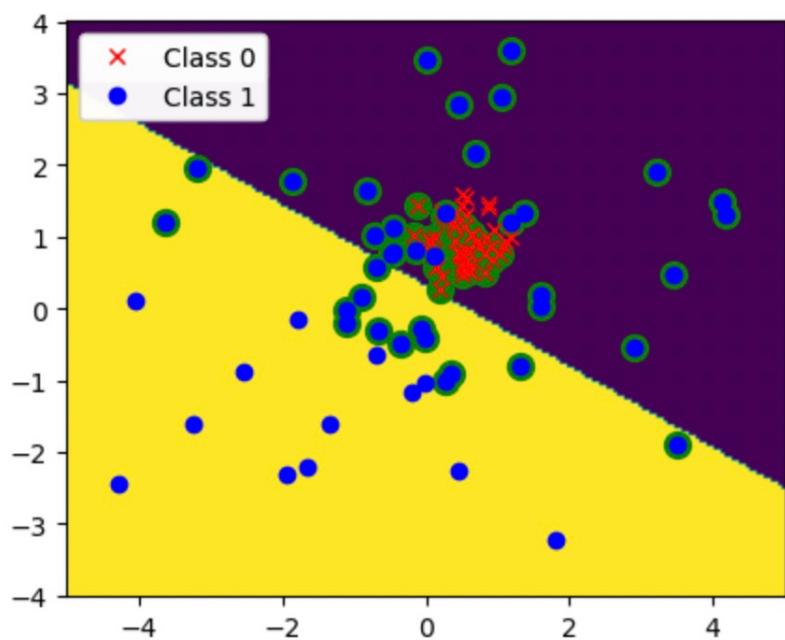
Test dataset3:



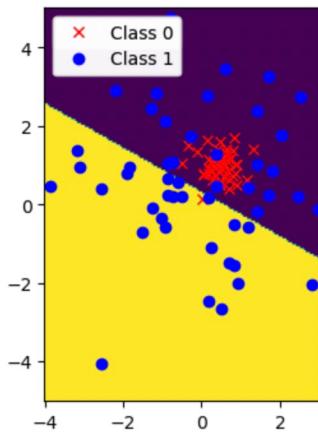
$C = 100$

```
train acc = 0.77, test acc = 0.75
weight vector w = [[-0.49088354 -0.86901288]]
offset w0 = [0.26395834]
```

Train dataset3:



Test dataset3:

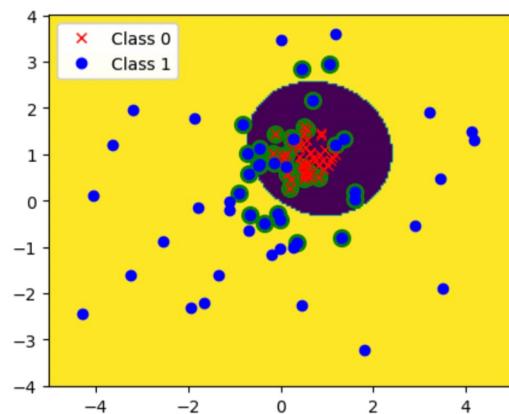


The distribution of dataset3 is not linear. Therefore, the linear kernel cannot classify the data 100% correctly and the performance of both values of C is almost the same.

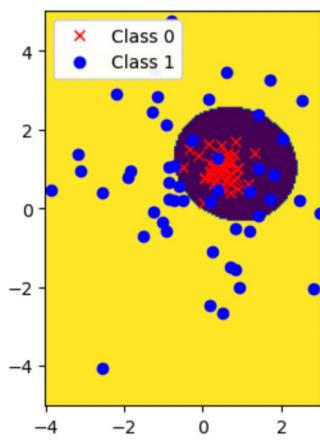
(d) $C = 1, \gamma = 0.1$

train acc = 0.89, test acc = 0.89

Train dataset3:



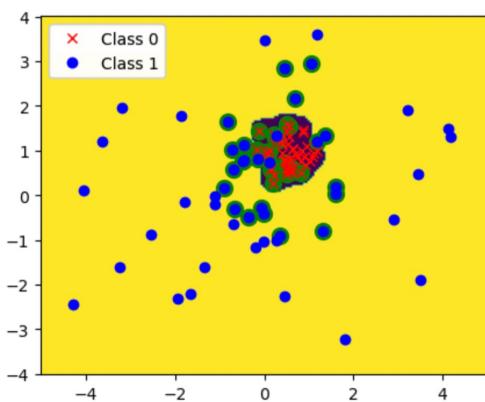
Test dataset3:



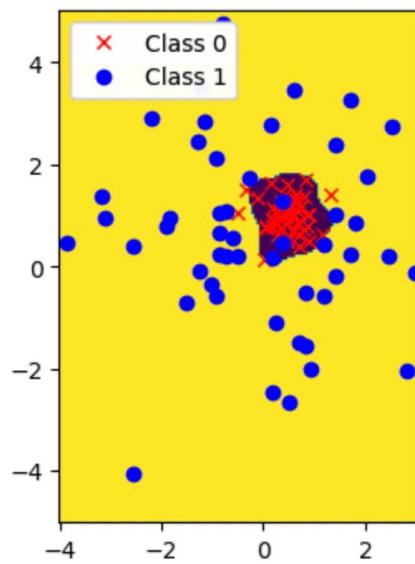
$C = 1, \gamma = 10$

train acc = 0.98, test acc = 0.94

Train dataset3:

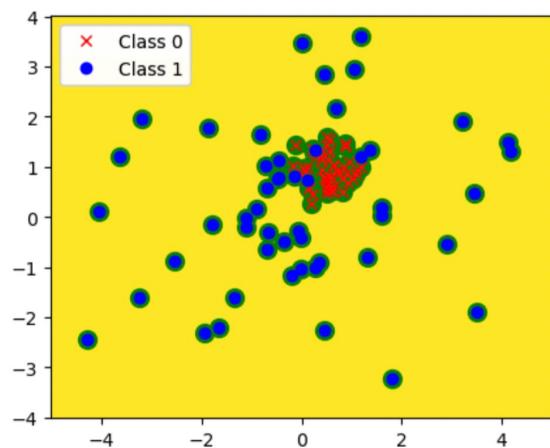


Test dataset3:

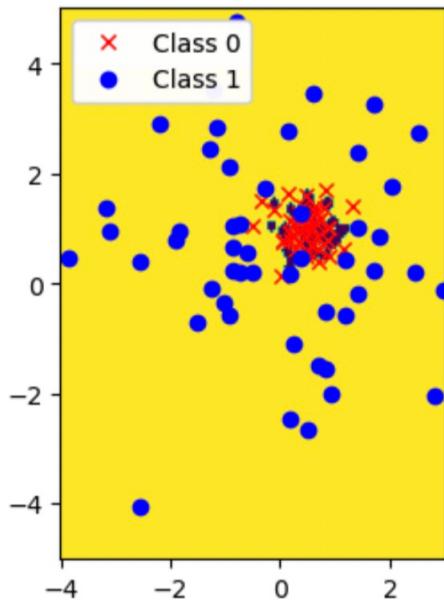


$C = 1, \gamma = 200$
train acc = 0.98, test acc = 0.77

Train dataset3:



Test dataset3:



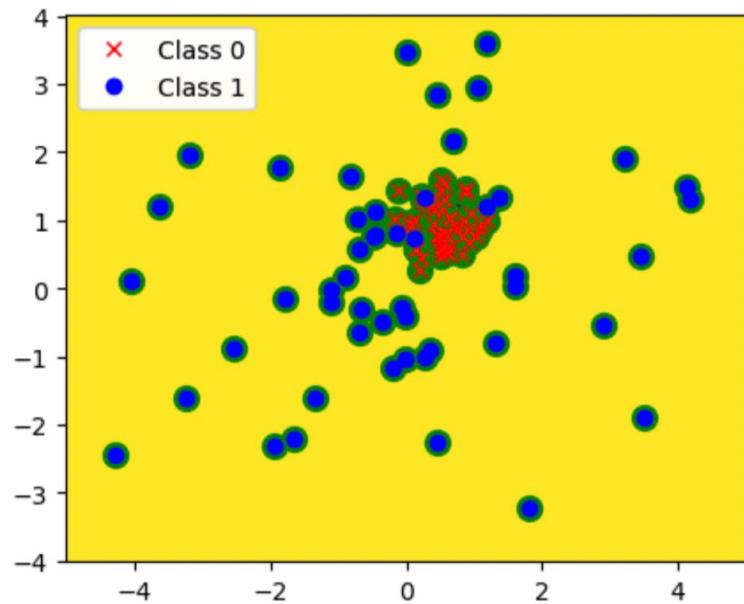
Explain the difference in decision regions for the different values of γ . Do you observe any overfitting or underfitting for any of the given values of γ ?

The larger the value of gamma is, the closer the decision boundary are to the support vectors. When the gamma is too large (e.g. 200), there existing overfitting according to the high accuracy of train dataset and low accuracy of test dataset.

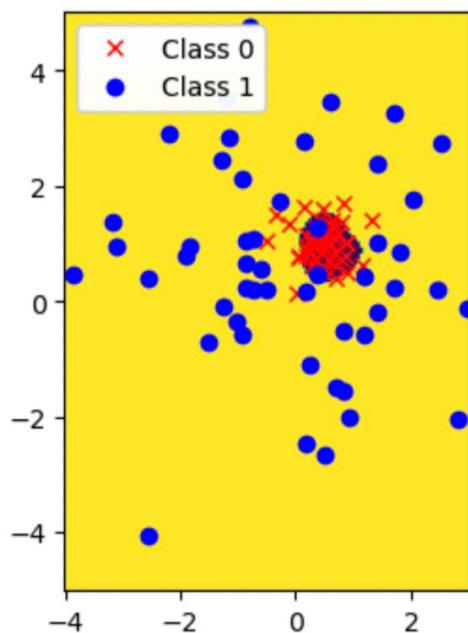
(e) $C = 0.01, \gamma = 10$

train acc = 0.88, test acc = 0.86

Train dataset3:



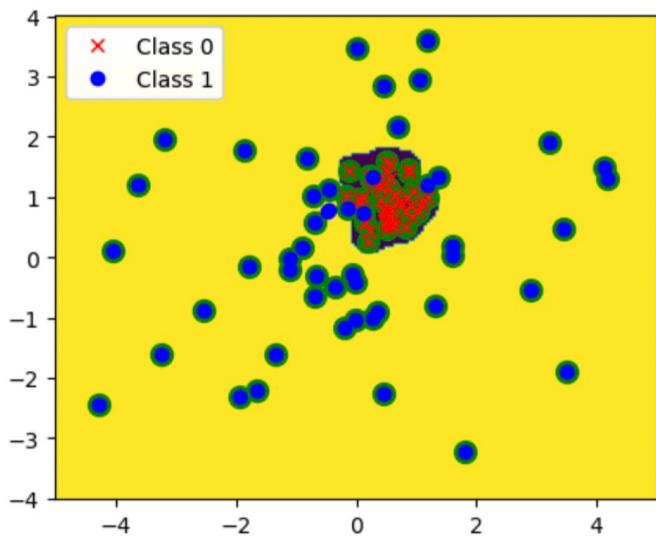
Test dataset3:



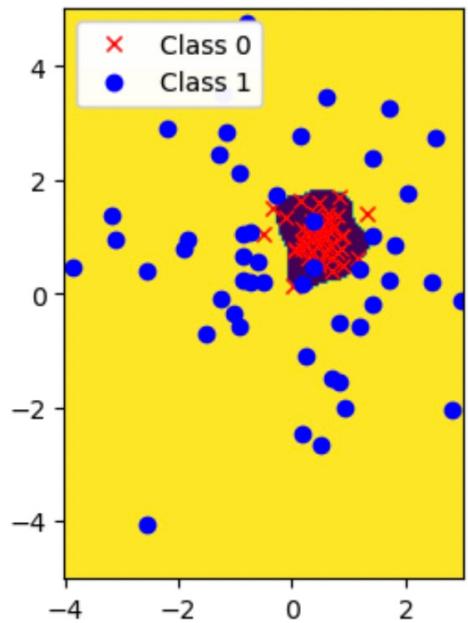
$$C = 1, \gamma = 10$$

train acc = 0.98, test acc = 0.94

Train dataset3:

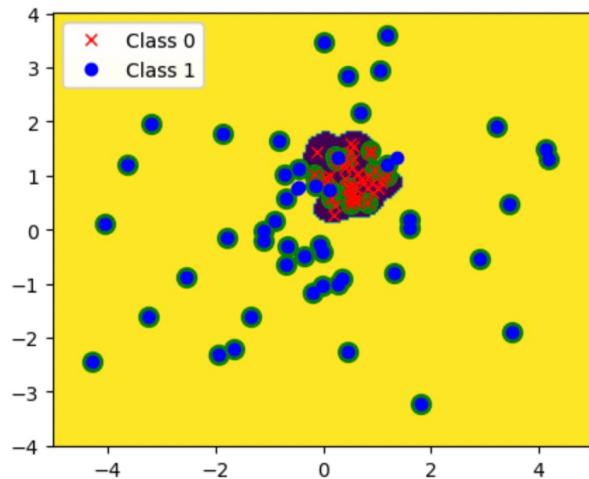


Test dataset3:

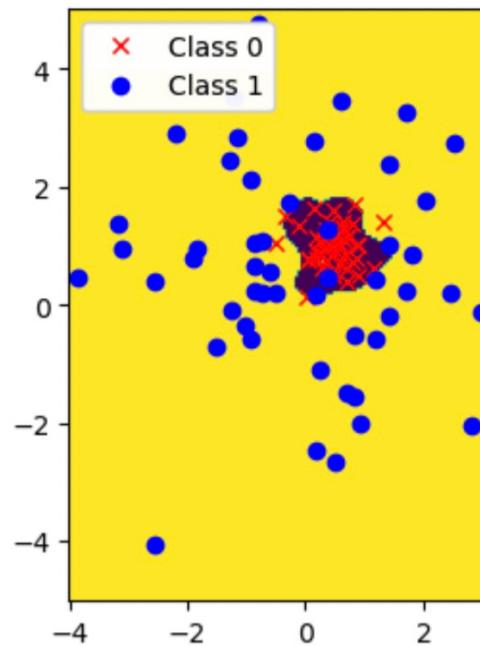


$C = 100, \gamma = 10$
train acc = 0.98, test acc = 0.95

Train dataset3:



Test dataset3:



Explain your observations in the different decision boundaries and the support vectors for the different values of C.

While the increasing of the value of C, the margin between the datapoints from class 0 and the decision boundary become larger. For a small value of C, every datapoints in class 0 are selected as support vectors. For a large value of C, only a few datapoints near the boundary are selected.