dollmann

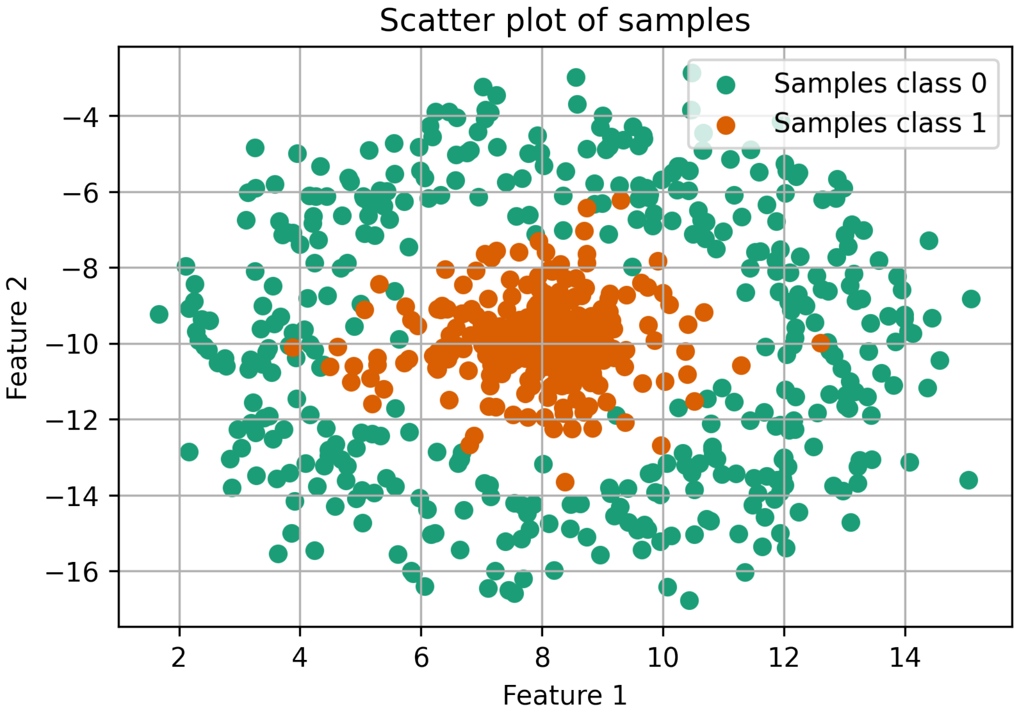
May 3, 2021

1 Foundations of Data Analysis

1.1 Supervised Learning Lab

1.1.1 Marc Martin Dollmann 11928878

2 2 Understanding the Dataset



The data is not linearly separable as it is given. But as sample class 1 forms a disk and sample class 0 forms a ring around sample class 1, they can be separated by a circle. Hence, mapping the data points to their euclidean distance from their central point, which is roughly at $(8, -10)$, would make them linearly separable.

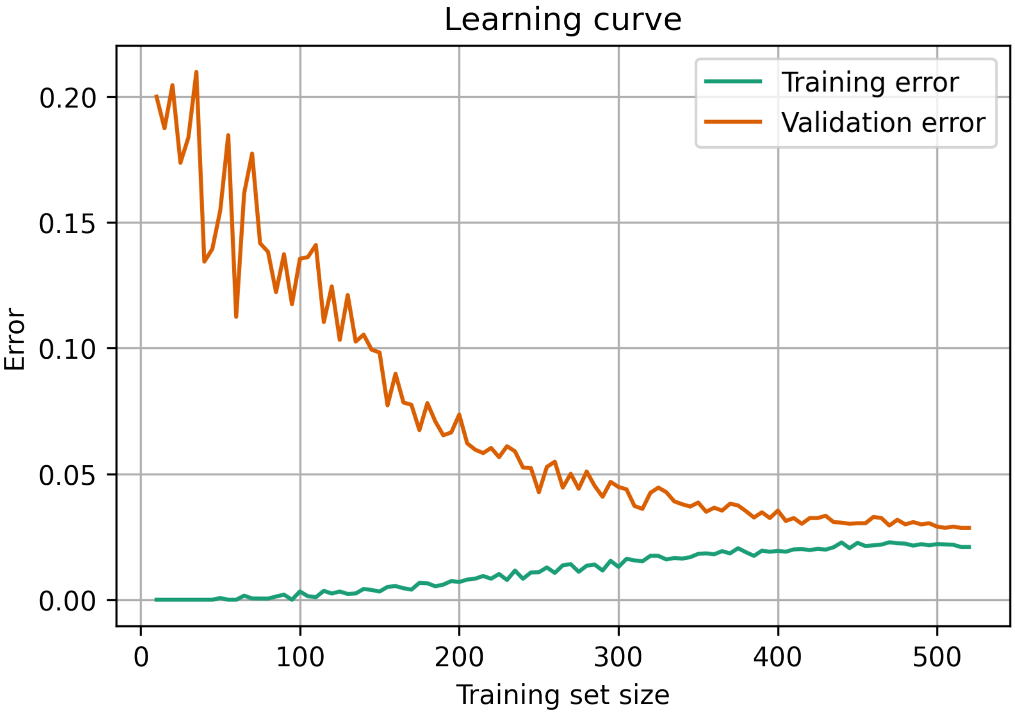
3 3 Setting up a classifier

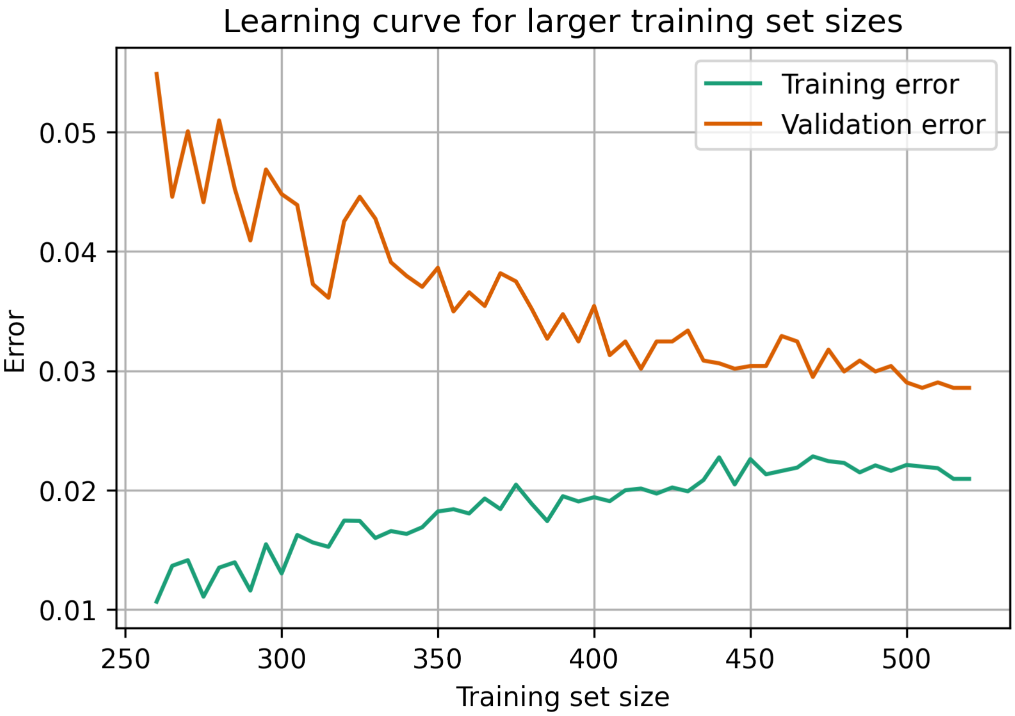
3.1 Splitting the data set

```
Score of training data: 0.979047619047619
Score of testing data:  0.9714285714285714
```

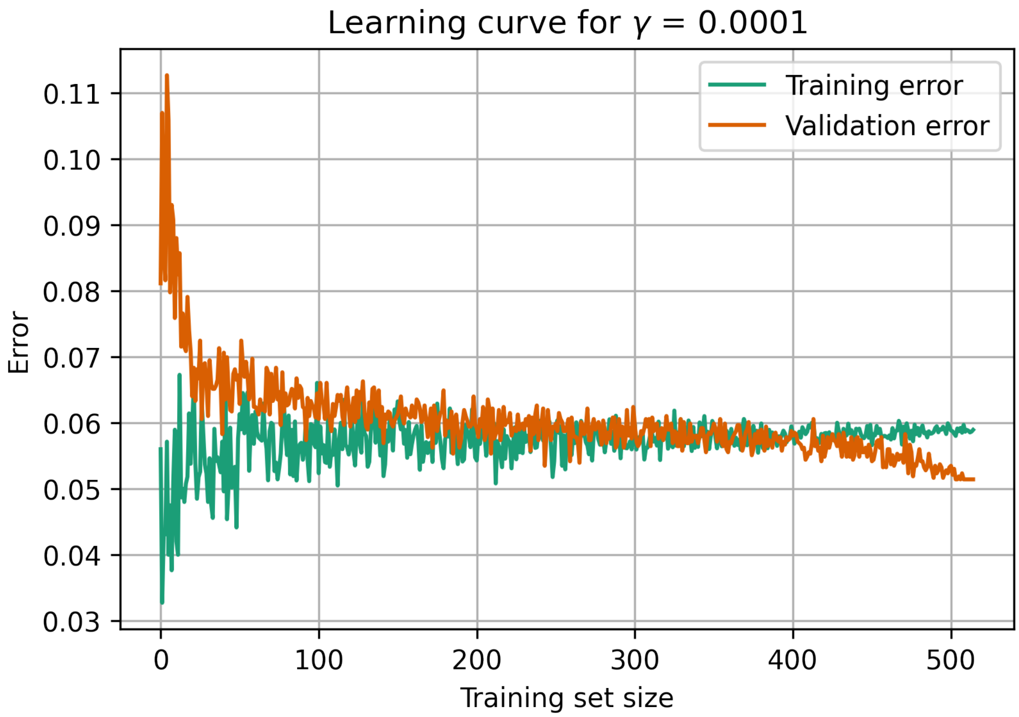
When plotting the data, one could see that there are some outliers. Some data points of one sample class are well into the area of the other. This explains why the classifiers isn't perfect. The difference in scores could come from the random allocation of outliers. The outliers in the training data will skew the results towards that, giving a worse result for the testing data where there will be different outliers. If the outliers are evenly distributed, the results should be nearly identical. But with such a low number of outliers (maybe 20 or 30), it might be the case that the most significant outliers of class 0 will get allocated to the training data and the most significant ones of class 1 to the testing data.

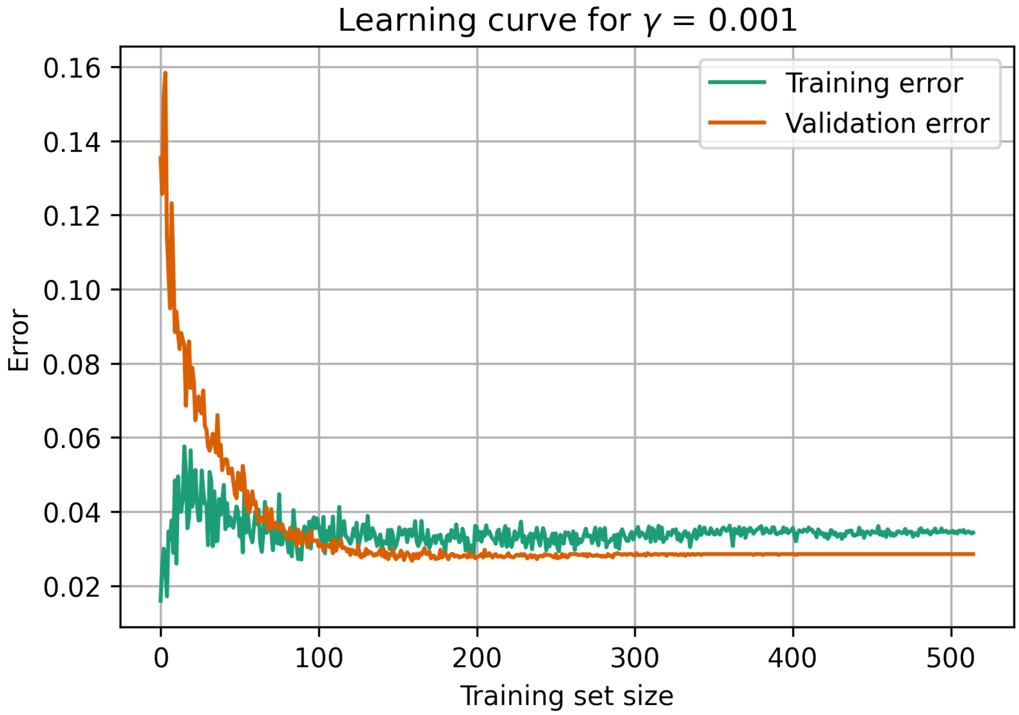
4 4 Learning Curves

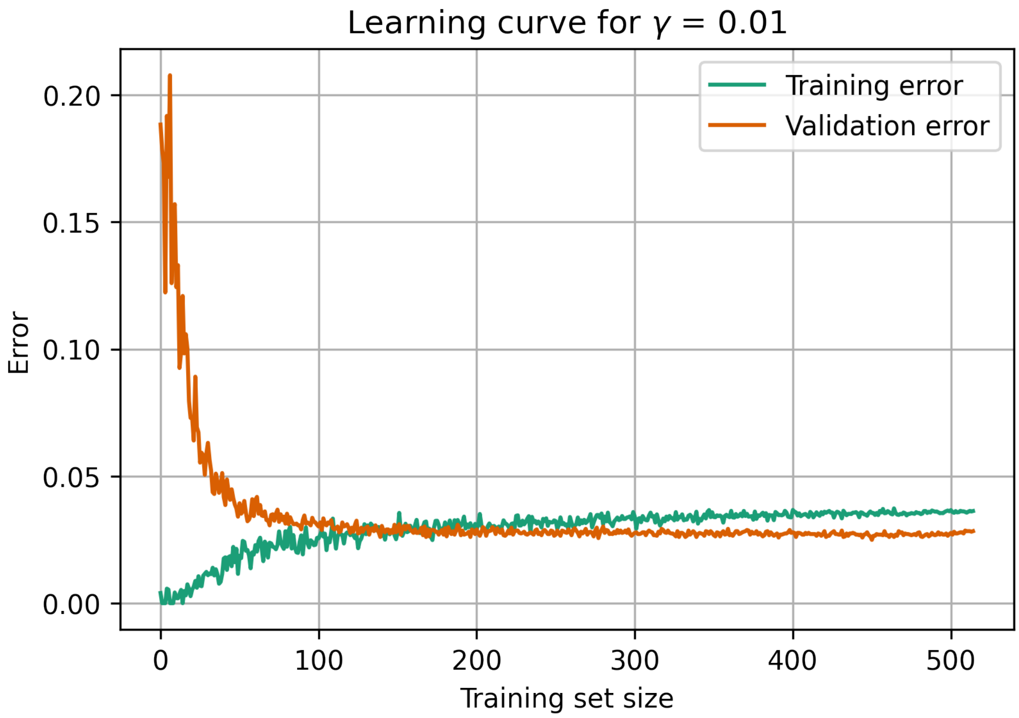


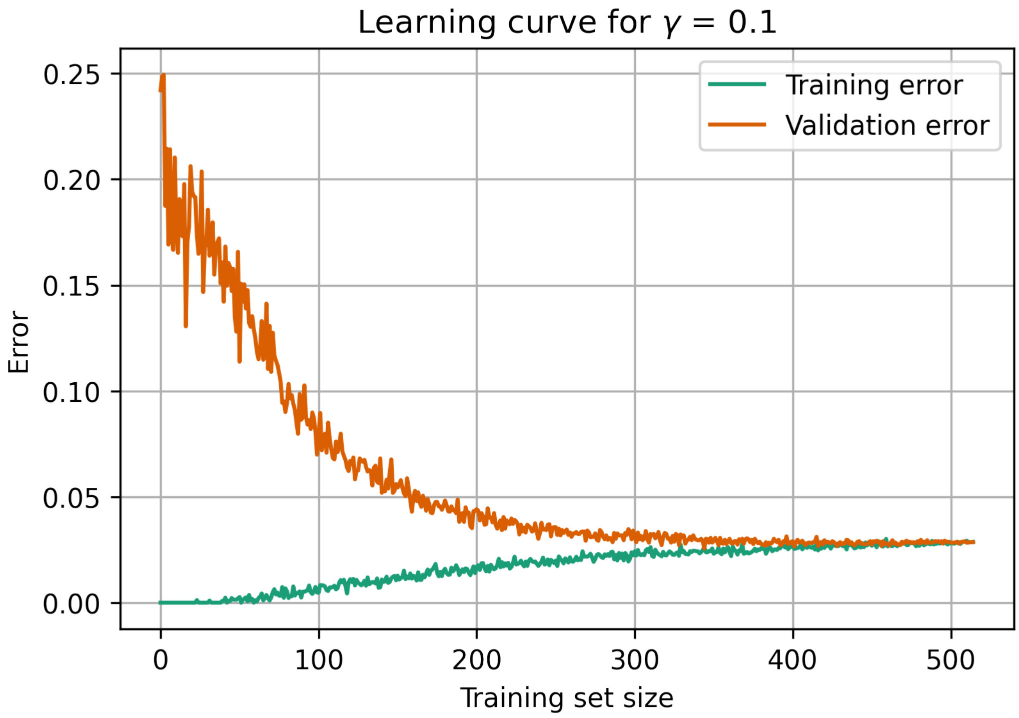


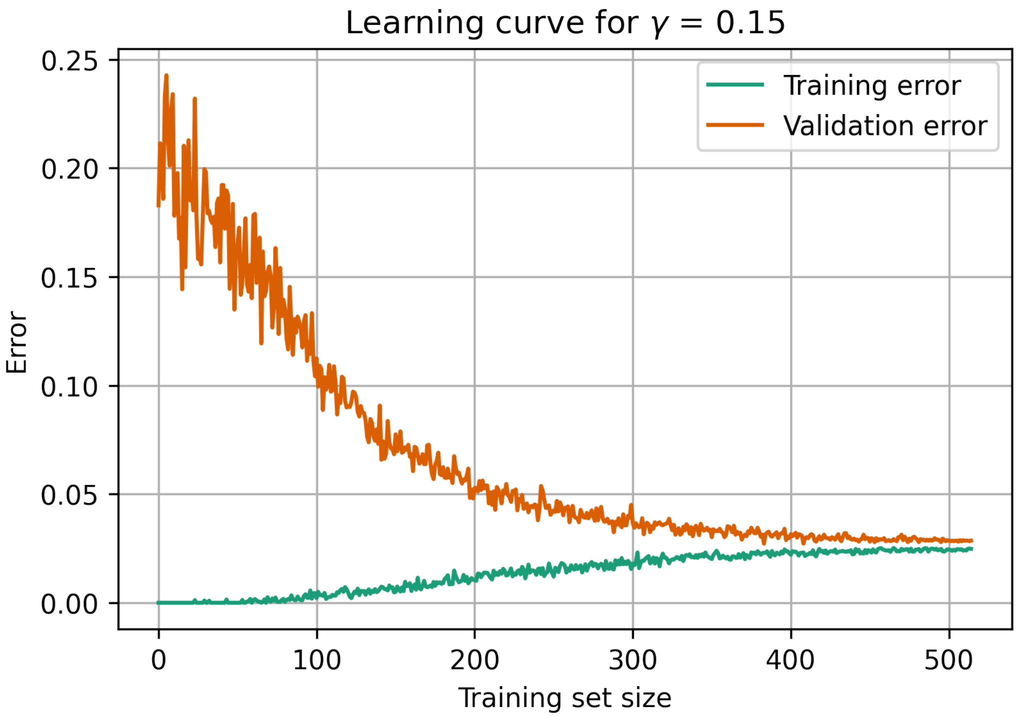
The training error rises linearly from 0 to a little over 2%. This is expected as increasing the training set size will increase the number of ouliers included in it. Increasing the training set size will also cause the quality of the classifier to increase. This is reflected in the decreasing validation error. It starts at over 20% and after reaching a training ste size of 250 stop falling linearly and starts approaching the training error asymptotically. With a maximal training set size there is only a percentage point of difference between the training and validaiton error. To furhter decrease the validation error, one would have to collect more samples to train the classifier on, and to decrease the training error one would need to consider a larger or tweak the existing hypothesis class. The latter will be done next, when the learning curves are plotted for different values of gamma.

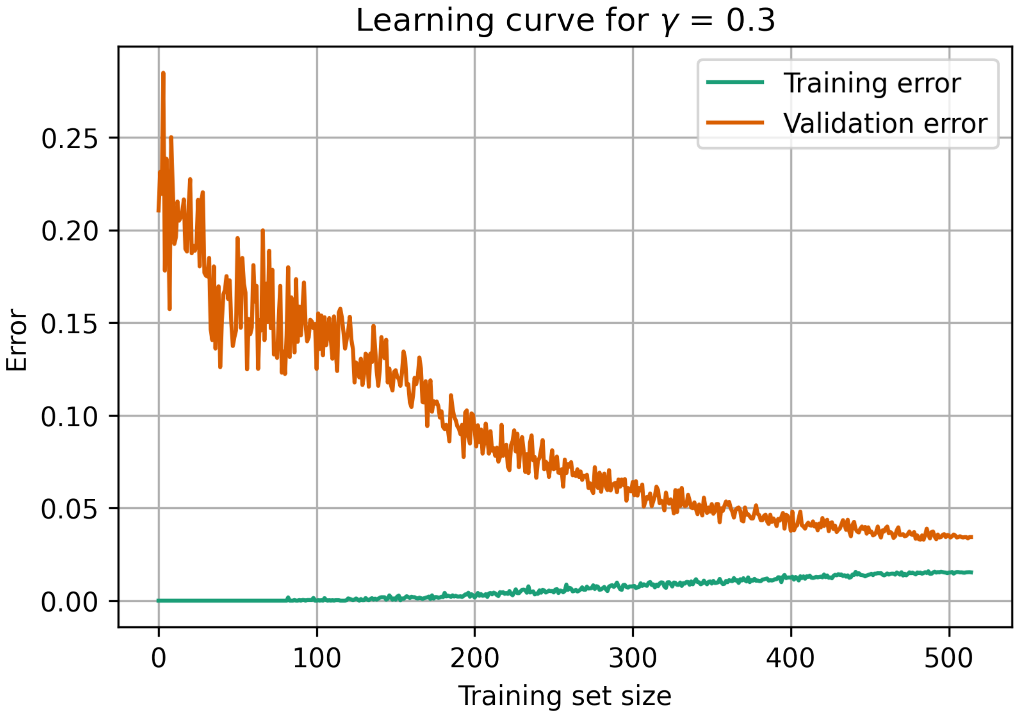

Learning curve for γ = 0.0001
Training error
Validation error
Error
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
0.11
0
100
200
300
400
500
Training set size


Learning curve for γ = 0.001
Training error
Validation error
Error
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
0
100
200
300
400
500
Training set size


Learning curve for γ = 0.01
Training error
Validation error
Error
0.00
0.05
0.10
0.15
0.20
0
100
200
300
400
500
Training set size


Learning curve for γ = 0.1
Training error
Validation error
Error
0.00
0.05
0.10
0.15
0.20
0.25
0
100
200
300
400
500
Training set size


Learning curve for γ = 0.15
Training error
Validation error
Error
0.25
0.20
0.15
0.10
0.05
0.00
0
100
200
300
400
500
Training set size

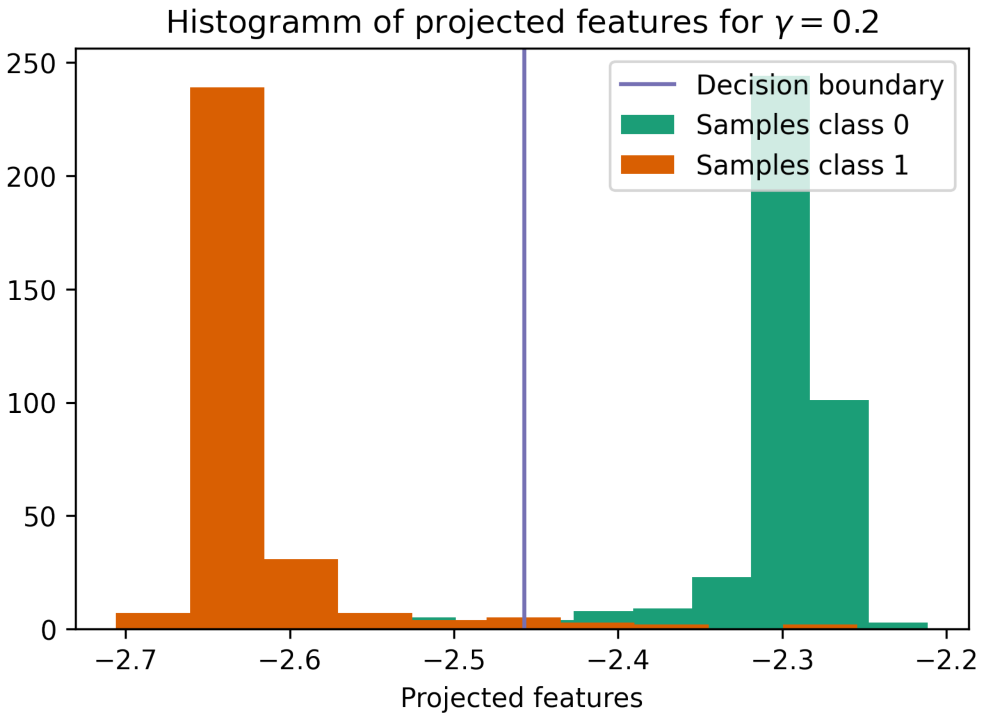

Learning curve for γ = 0.3
Training error
Validation error
Error
0.25
0.20
0.15
0.10
0.05
0.00
0
100
200
300
400
500
Training set size

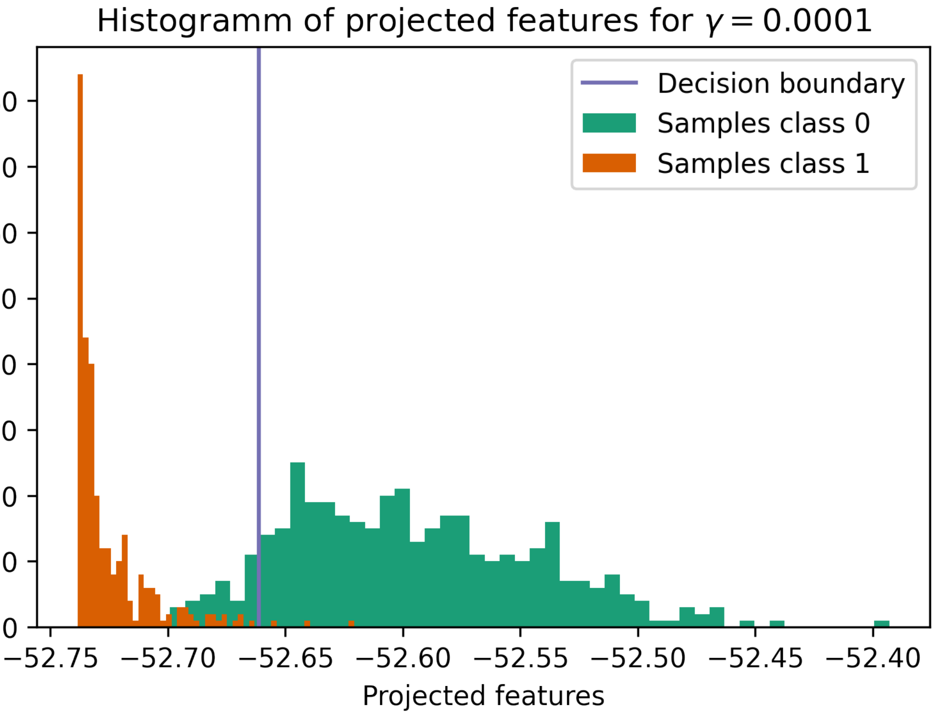
Looking at the different learning curves shows how the γ-factor is relevant to make sure the model does not over- or underfit the training data. Going forward, γ will be set to 0.1 as for this value the training and validation error converge to around 3%. For $\gamma = 0.0001$ the model underfits the data as the training error remains high. For $\gamma = 0.001$ and $\gamma = 0.01$, the model overfits less, but the training error still remain higher than for $\gamma = 0.1$. In both cases increasing the training set size doesn't lead to a reduction of validation error, meaning the hypothesis class is exhausted and one should change it. For $\gamma = 0.1$ and for $\gamma = 0.3$ the model overfits the data, leading to a very low training error but a much higher validation error because it focuses too much on the outlier in the training data.

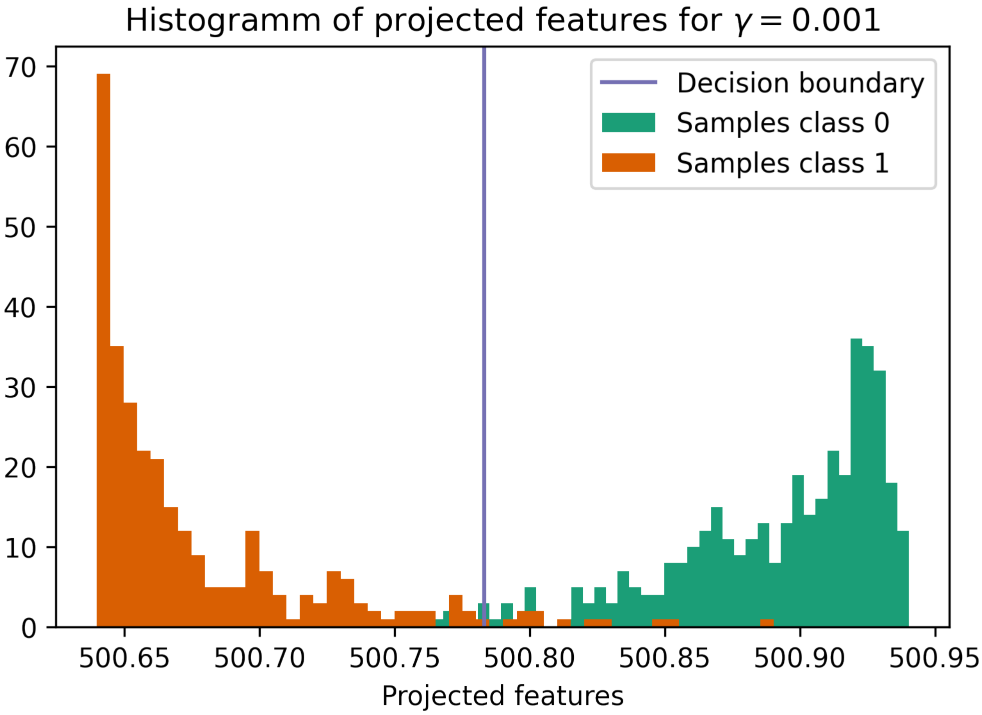
```
Score of training data: 0.9752380952380952
Score of testing data:  0.9714285714285714
Predictions:  [False False False]
```

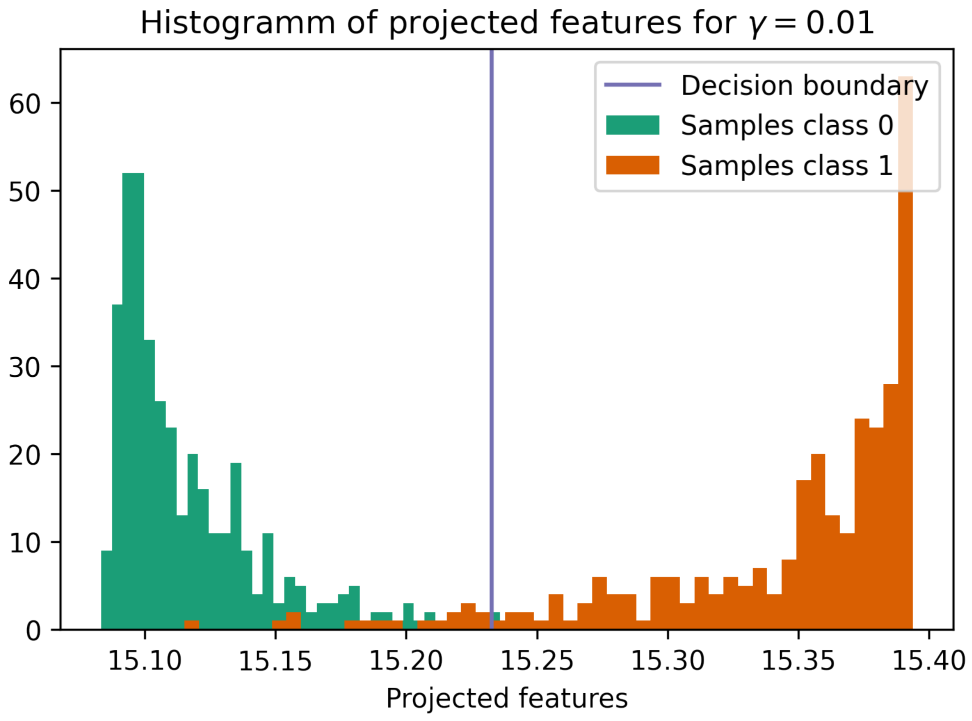
The classifier predicts the label 0 for all three samples.

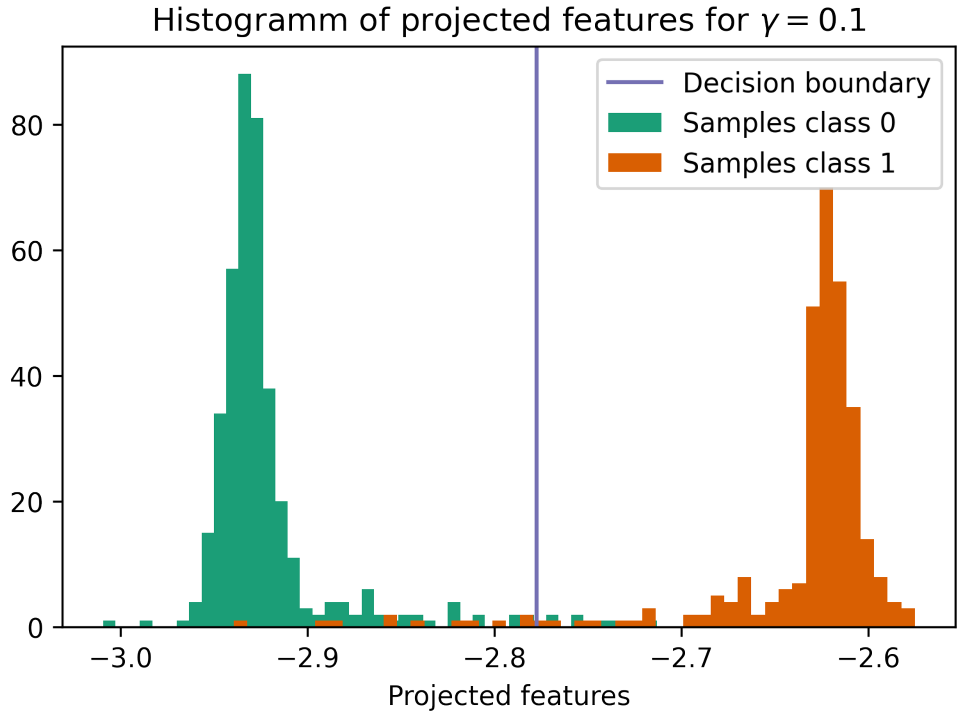
5 5 In the projected space

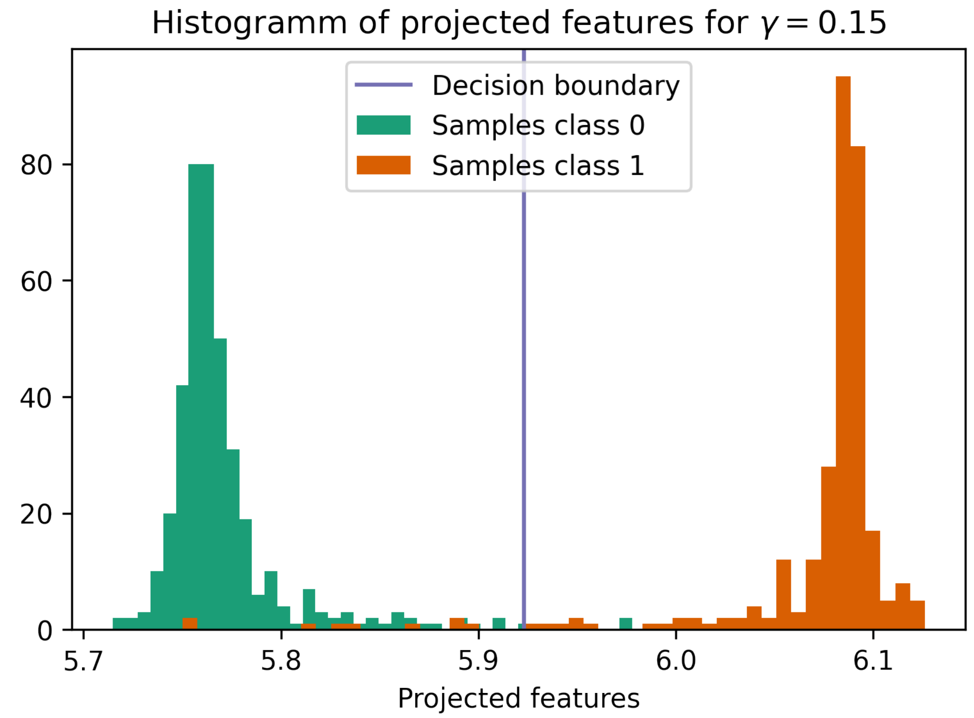


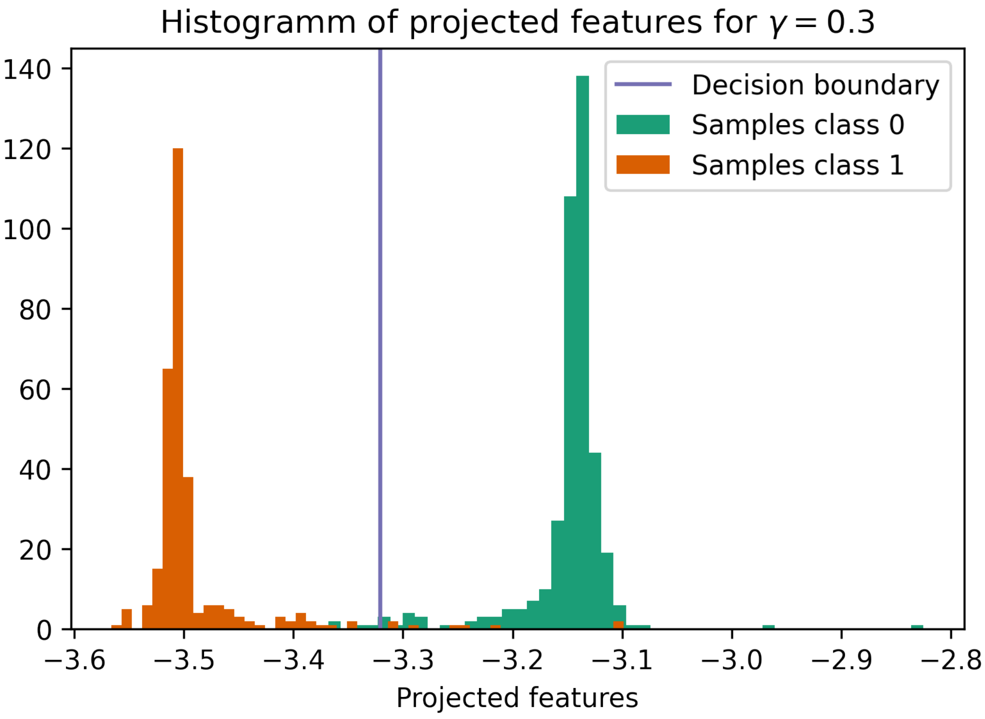

Histogramm of projected features for $\gamma = 0.0001$
Decision boundary
Samples class 0
Samples class 1
80
70
60
50
40
30
20
10
0
−52.75
−52.70
−52.65
−52.60
−52.55
−52.50
−52.45
−52.40
Projected features


Histogramm of projected features for $\gamma = 0.001$
Decision boundary
Samples class 0
Samples class 1
70
60
50
40
30
20
10
0
500.65
500.70
500.75
500.80
500.85
500.90
500.95
Projected features


Histogramm of projected features for $\gamma = 0.01$
Decision boundary
Samples class 0
Samples class 1
0
10
20
30
40
50
60
15.10
15.15
15.20
15.25
15.30
15.35
15.40
Projected features


Histogramm of projected features for $\gamma = 0.1$
Decision boundary
Samples class 0
Samples class 1
0
20
40
60
80
−3.0
−2.9
−2.8
−2.7
−2.6
Projected features

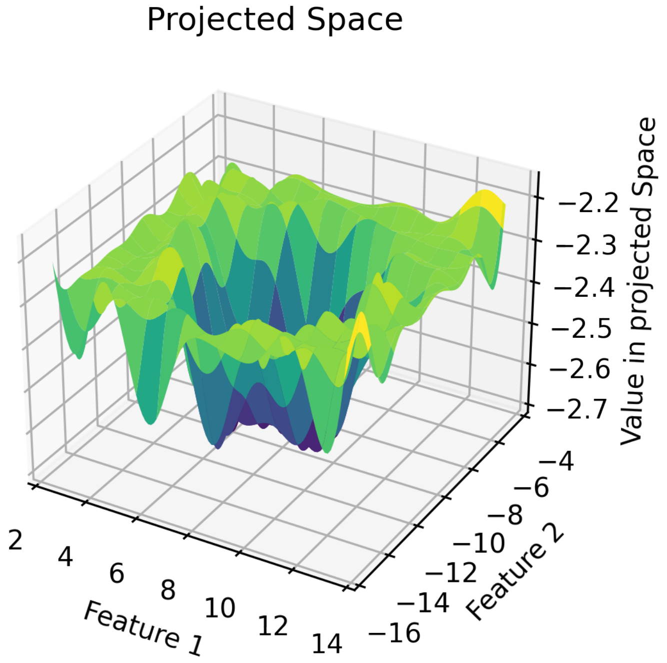

Histogramm of projected features for $\gamma = 0.15$
Decision boundary
Samples class 0
Samples class 1
0
20
40
60
80
5.7
5.8
5.9
6.0
6.1
Projected features


Histogramm of projected features for $\gamma = 0.3$
Decision boundary
Samples class 0
Samples class 1
0
20
40
60
80
100
120
140
−3.6
−3.5
−3.4
−3.3
−3.2
−3.1
−3.0
−2.9
−2.8
Projected features

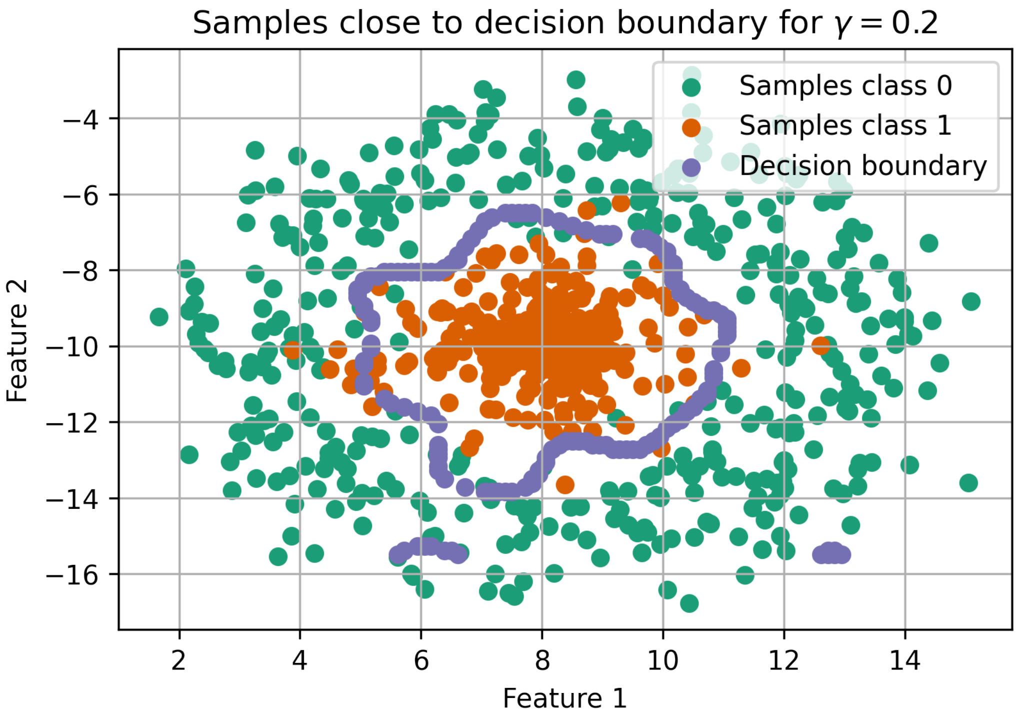
Increasing γ pushes the outlier further away from the main cluster of the data. Hence the historgrams seem to get more concentrated. As we are considering a RBF kernel, increasing γ means decreasing the similarity between two points. Hence, the classifier becomes more relaxed and may project features to a wider range of values.

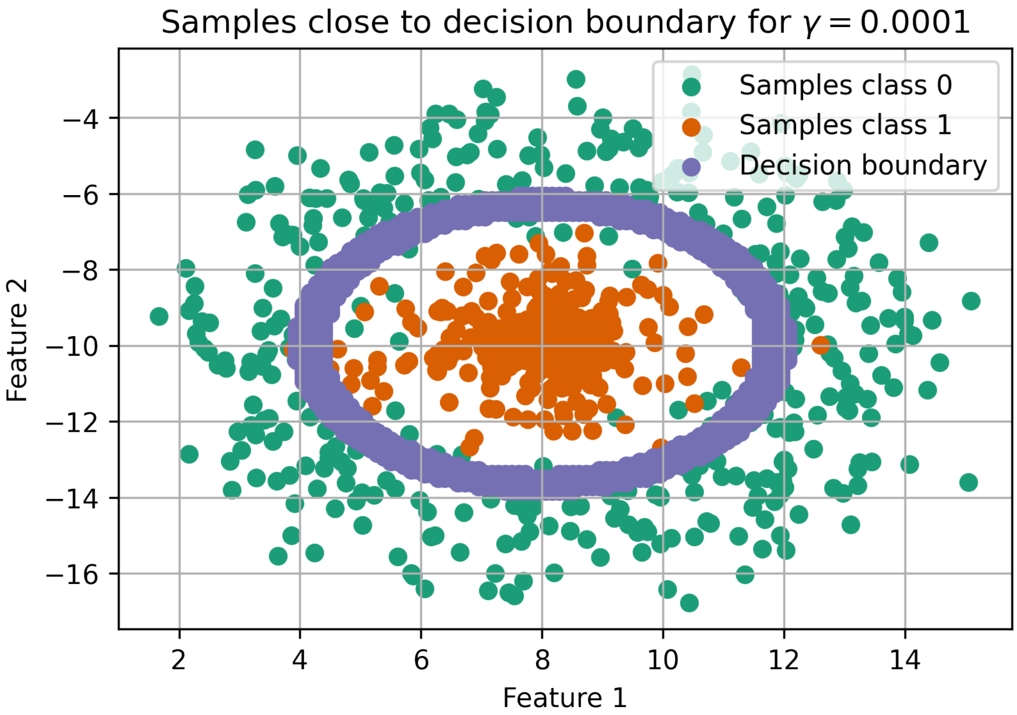
```
Separation ratios
for gamma=0.0001 is 5.221928546594546
for gamma=0.001 is 12.309132729531475
for gamma=0.01 is 12.751630700049066
for gamma=0.1 is 19.822998431516638
for gamma=0.15 is 20.92507266667788
for gamma=0.3 is 22.699612629592647
```

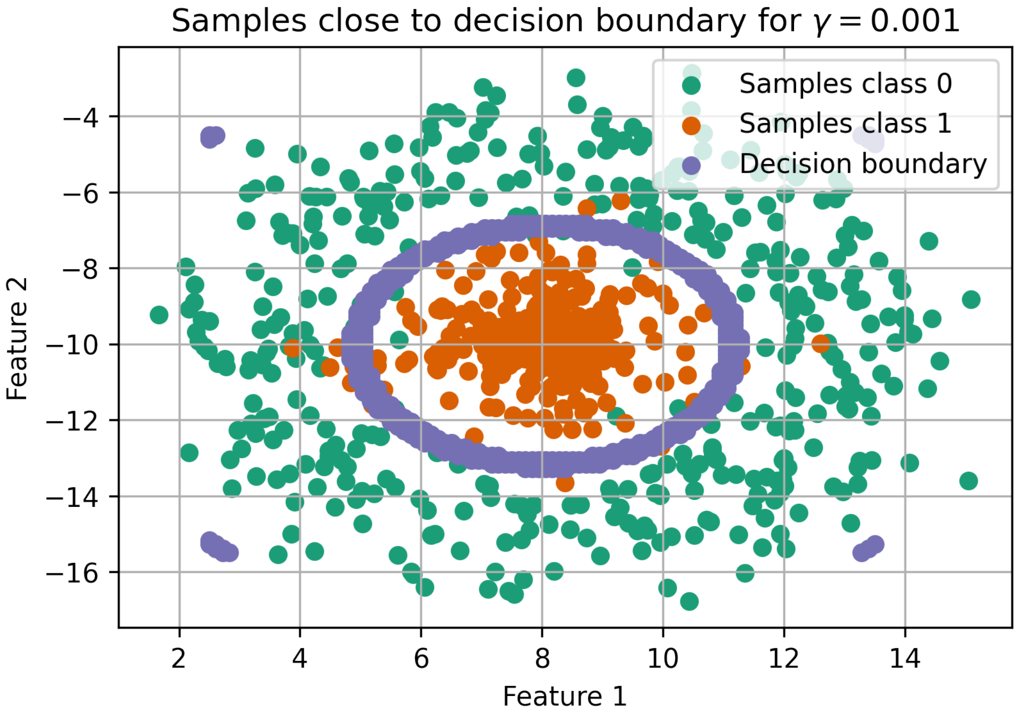
6 6 The decision boundary

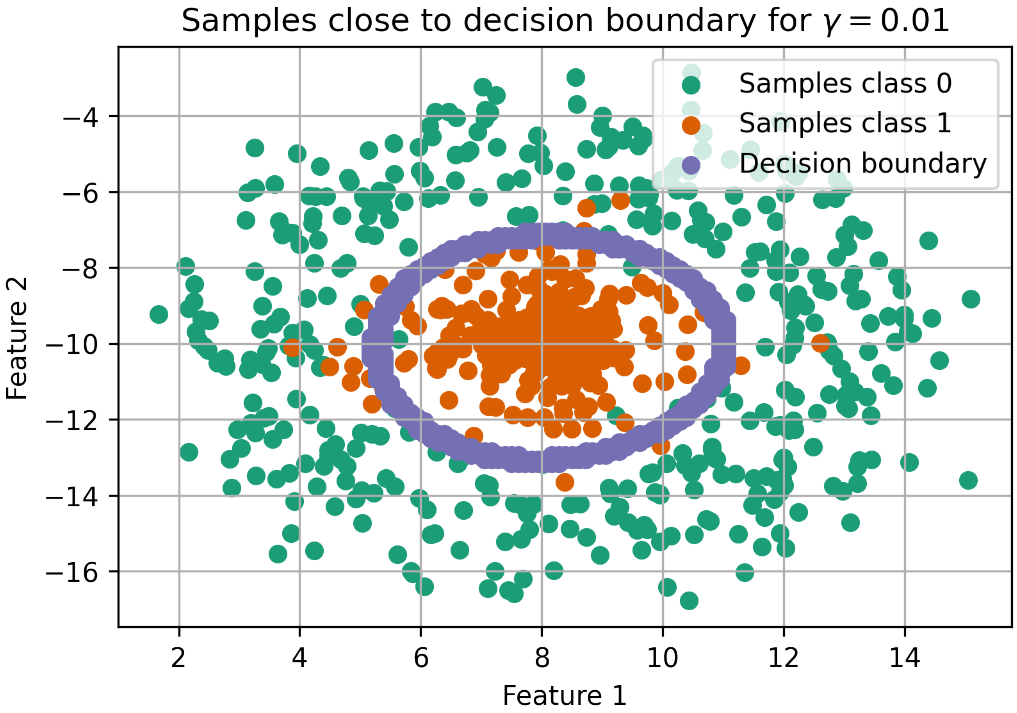


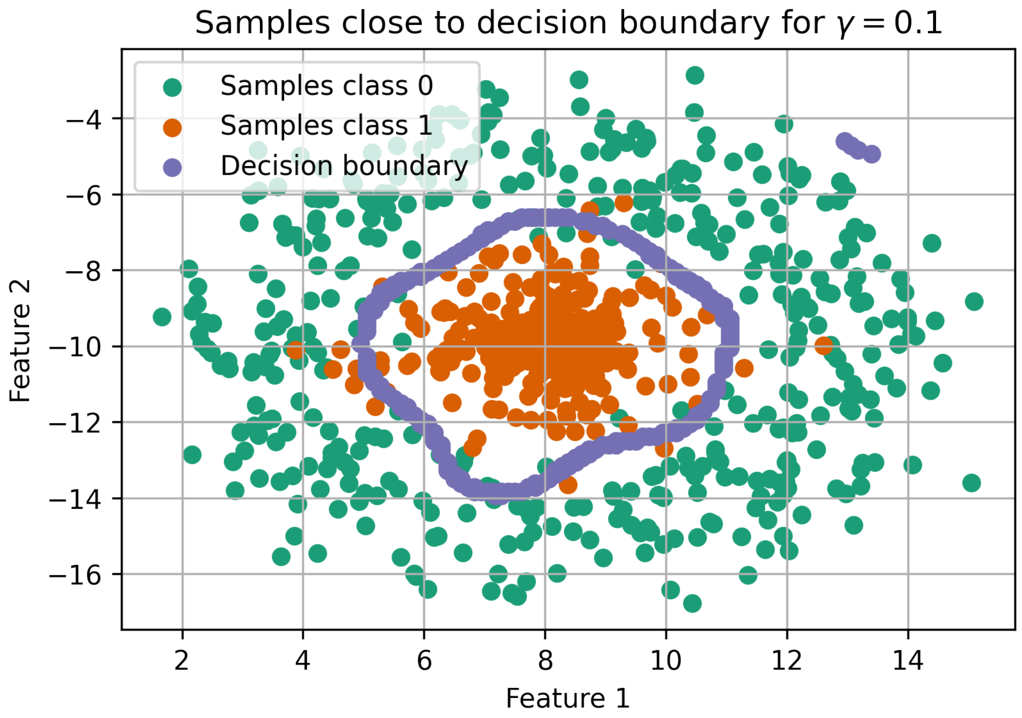
The plot shows how the model chooses some lower value for class 0 and a higher values for class 1. It reflects the roundness of the border between the two classes.

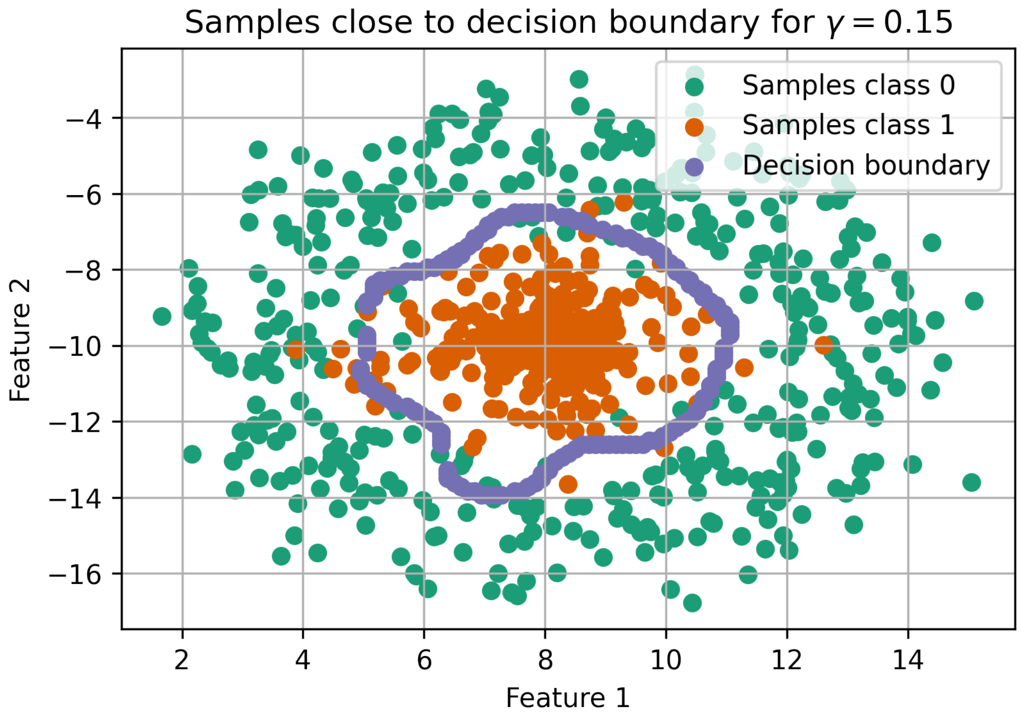

Samples close to decision boundary for $\gamma = 0.2$
Samples class 0
Samples class 1
Decision boundary
Feature 2
−4
−6
−8
−10
−12
−14
−16
2
4
6
8
10
12
14
Feature 1

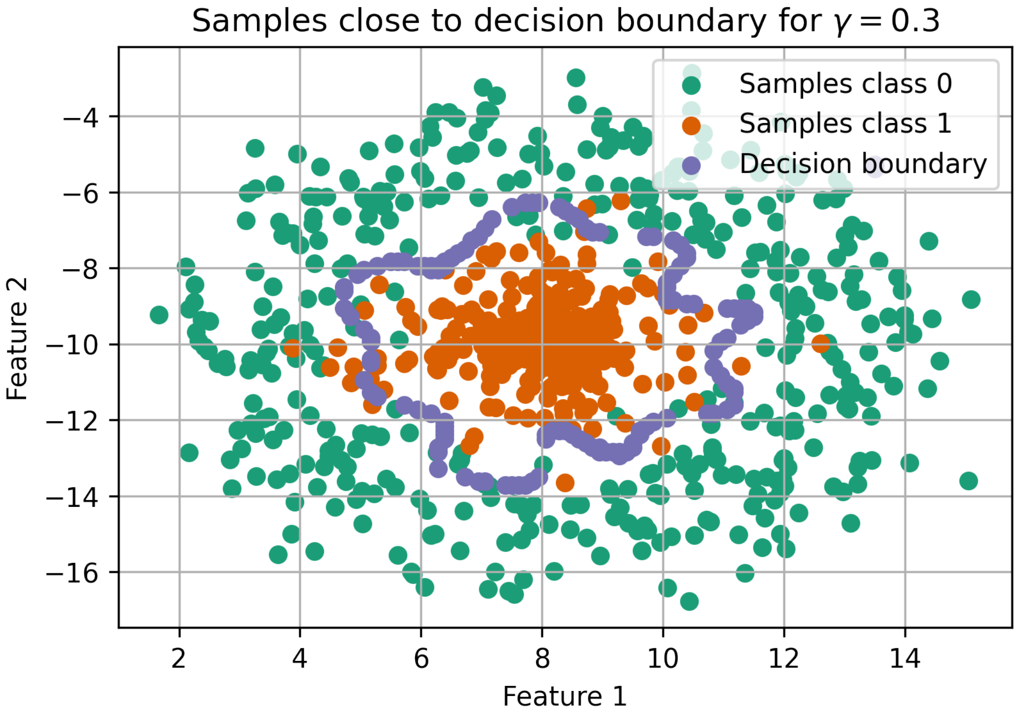

Samples close to decision boundary for $\gamma = 0.0001$
Samples class 0
Samples class 1
Decision boundary
Feature 2
−4
−6
−8
−10
−12
−14
−16
2
4
6
8
10
12
14
Feature 1


Samples close to decision boundary for $\gamma = 0.001$
Samples class 0
Samples class 1
Decision boundary
Feature 2
−4
−6
−8
−10
−12
−14
−16
Feature 1
2
4
6
8
10
12
14


Samples close to decision boundary for $\gamma = 0.01$
Samples class 0
Samples class 1
Decision boundary
Feature 2
−4
−6
−8
−10
−12
−14
−16
Feature 1
2
4
6
8
10
12
14


Samples close to decision boundary for $\gamma = 0.1$
Samples class 0
Samples class 1
Decision boundary
Feature 1
Feature 2
2
4
6
8
10
12
14
−4
−6
−8
−10
−12
−14
−16


Samples close to decision boundary for $\gamma = 0.15$
Samples class 0
Samples class 1
Decision boundary
Feature 1
Feature 2
2
4
6
8
10
12
14
−4
−6
−8
−10
−12
−14
−16



Comparing the decision boundaries for the different choices of γ, one sees again how increasing γ means decreasing the sensitivity of the similarity measure. For the smaller values of γ, the decision boundary is close to elliptical and very thick, which dosn't consider the outliers in our data and refelcts the smaller range of values assigned to the projected features. But for the largest value of γ, the decision boundary is too complex for the data set, showing signs of overfitting.