ML Lab Week 10 SVM Lab Instructions

Name	Namritha Diya Lobo
SRN	PES2UG23CS362
Section	F
Date	12/10/2025

Moons Dataset Questions (2 questions):

1. Inferences about the Linear Kernel's performance.
The linear SVM draws a straight line to separate the 2 classes. The decision boundary clearly misclassifies several points near the curved edges of the 2 half-moon shapes.
The linear Kernel struggles with non-linear patterns, like the Moons dataset, which has an inherent crescent (non linear) structure. It is too simple for this dataset. Would work better for a linearly separable dataset.

2. Comparison between RBF and Polynomial kernel decision boundaries.
On comparing RBF and polynomial kernels, its observed that the RBF kernel captures the shape of the data better and more naturally. The data which has an inherent crescent structure is classified more accurately with the RBF kernel.

Banknote Dataset Questions (2 questions):

1. Which kernel was most effective for this dataset?
Linear Kernel works very well on this data(since its linearly separable). RBF also performs quite well , but slightly overfits

2. Why might the Polynomial kernel have underperformed here?
Polynomial kernels performs worse than linear of RBF here. It creates a curved higher-order decision boundary, which is not needed for linearly separable data.
Thus the Polynomial kernel underperforms because the Banknote dataset is most linearly separable , so the higher order curved boundary is unnecessary and can lead to overfitting.

Hard vs. Soft Margin Questions (4 questions):
1. Which margin (soft or hard) is wider?
The soft margin SVM(C=0.1) produces a wider margin because it tolerates some misclassifications, prioritizing a larger separation between classes/.

2. Why does the soft margin model allow "mistakes"?

The SVM allows these “mistakes” to achieve a wider margin and better generalization. Its primary goal is not to perfectly classify every training point, but to find a decision boundary that generalizes well to unseen data.

3. Which model is more likely to be overfitting and why?

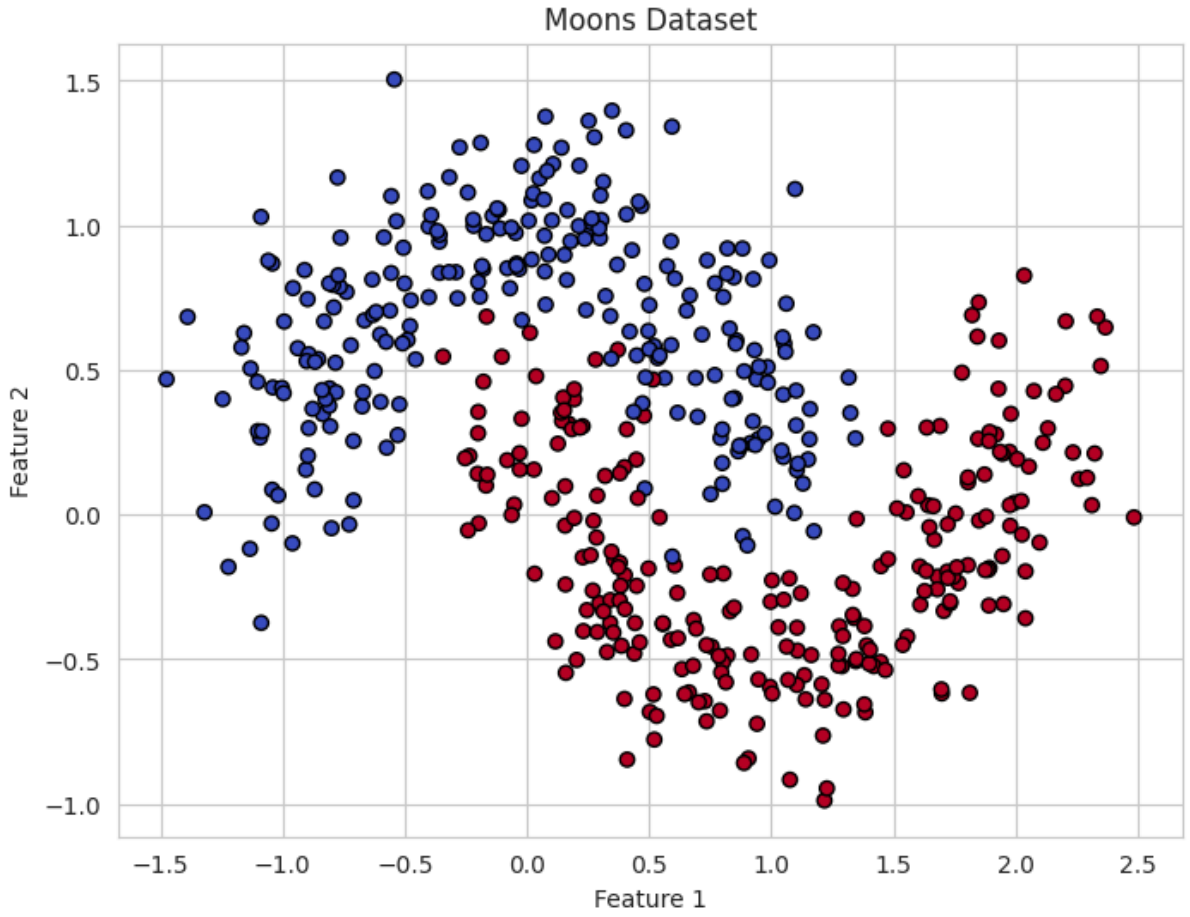
The Hard Margin SVM(C=100) is more likely to overfit, because it tries to classify every training point correctly, including outliers, which reduces generalization.

4. Which model would you trust more for new data and why?

I would trust the Soft margin SVM(c=0.1) more for classifying new data. In practice, starting with a low C value is generally better in noisy environments, as it balances margin width and tolerance to misclassification.

SCREENSHOTS

1. Moon dataset



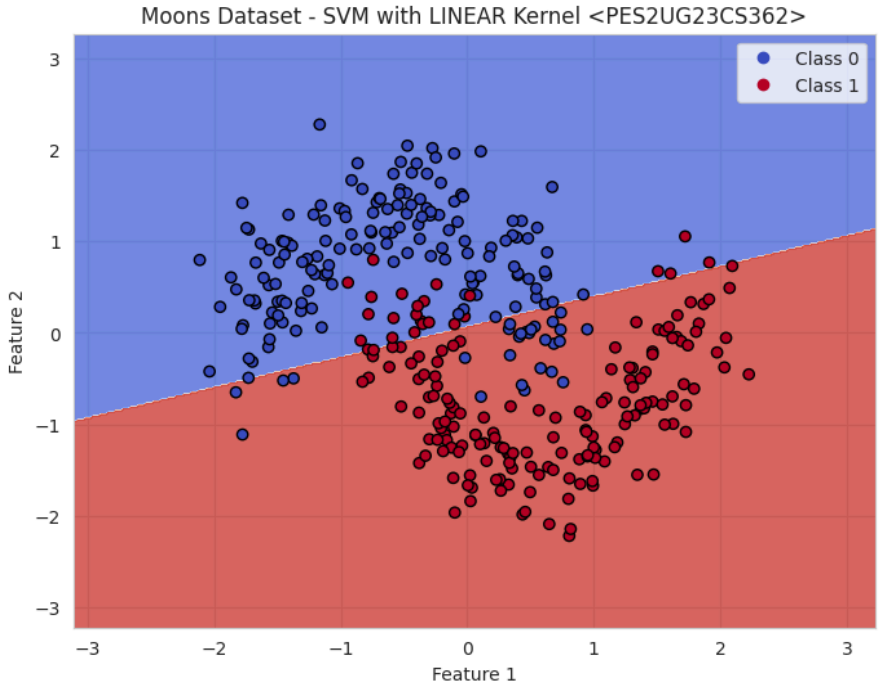
Classification report

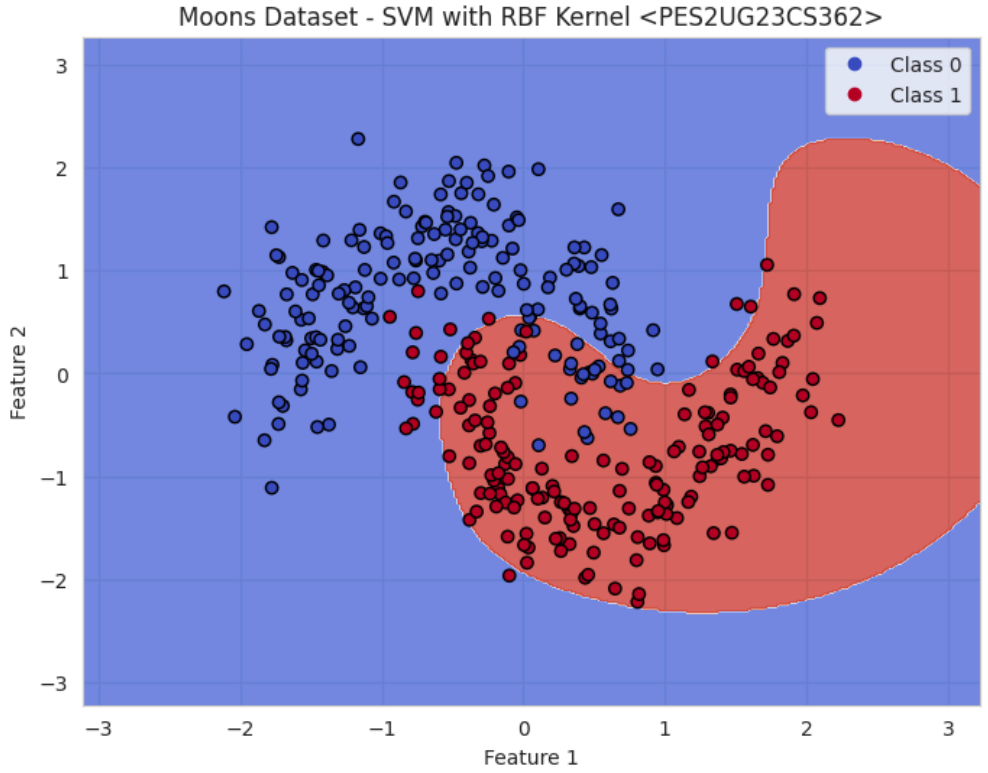
```
SVM with LINEAR Kernel <PES2UG23CS362>
              precision    recall  f1-score   support

           0       0.85      0.89      0.87        75
           1       0.89      0.84      0.86        75

    accuracy                           0.87       150
   macro avg       0.87      0.87      0.87       150
weighted avg       0.87      0.87      0.87       150
```

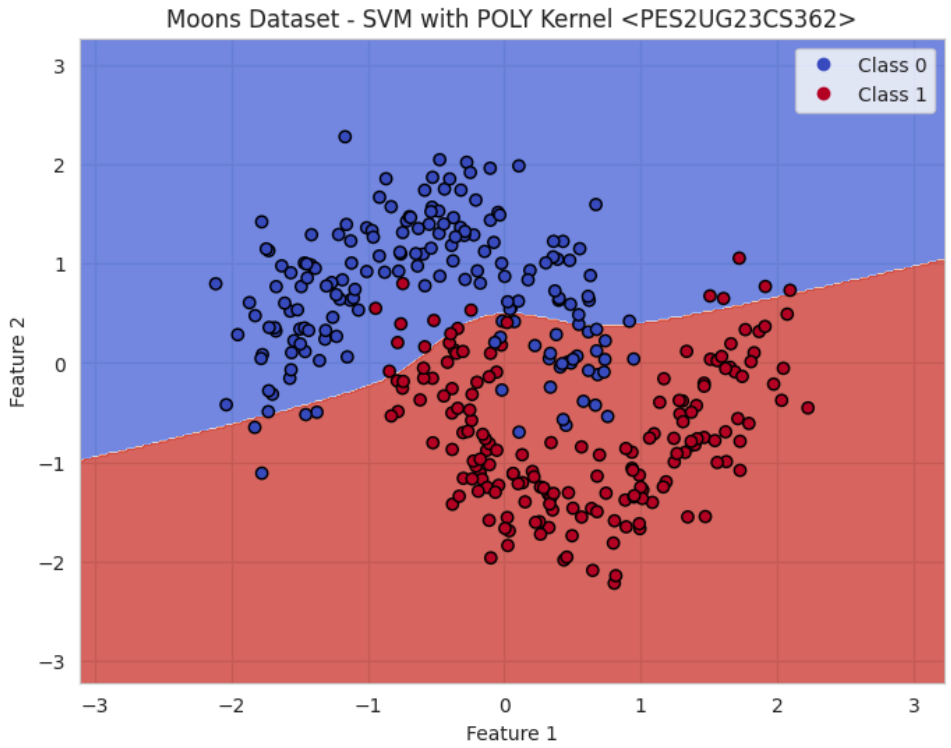
```
SVM with RBF Kernel <PES2UG23CS362>
              precision    recall  f1-score   support

           0       0.96      1.00      0.98        75
           1       1.00      0.96      0.98        75

    accuracy                           0.98       150
   macro avg       0.98      0.98      0.98       150
weighted avg       0.98      0.98      0.98       150
```

```
SVM with POLY Kernel <PES2UG23CS362>
              precision    recall  f1-score   support

           0       0.93      0.88      0.90        75
           1       0.89      0.93      0.91        75

    accuracy                           0.91       150
   macro avg       0.91      0.91      0.91       150
weighted avg       0.91      0.91      0.91       150
```

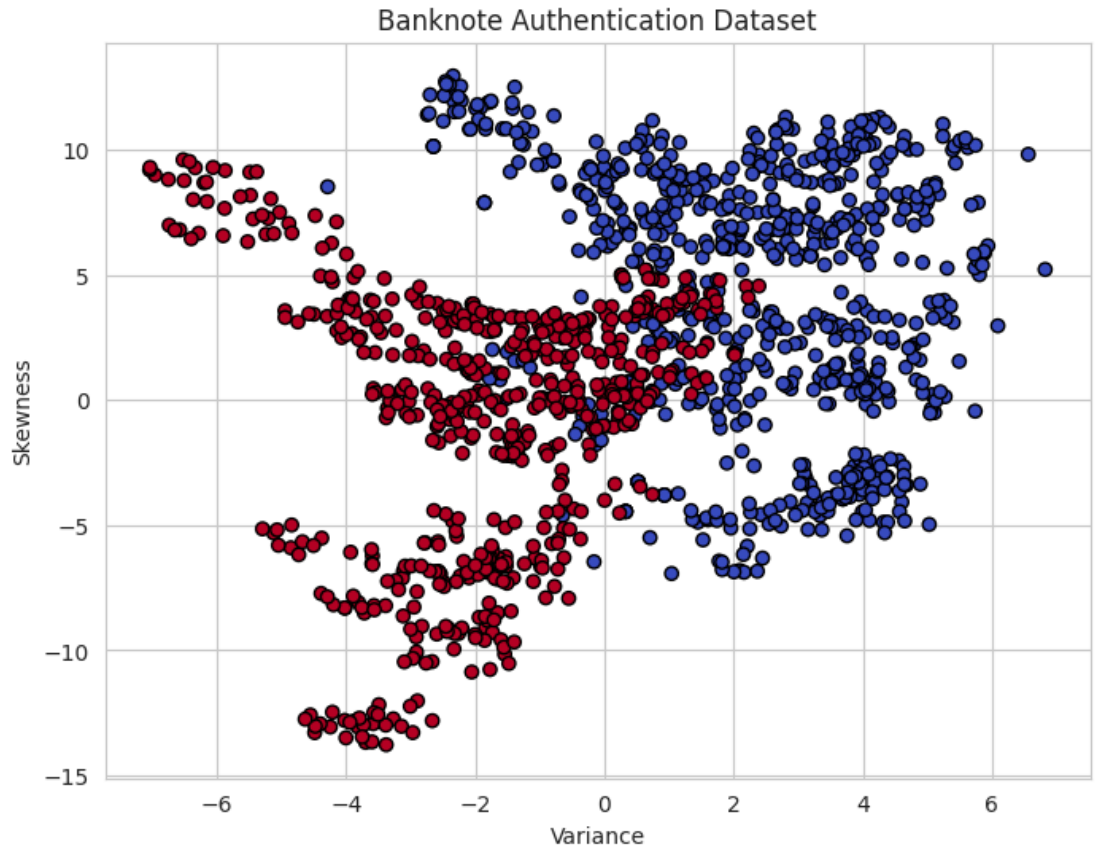
Decision boundary Visualization







2. Banknote Authentication Dataset



Classification Report

SVM with LINEAR Kernel <PES2UG23CS362>

	precision	recall	f1-score	support
Forged	0.90	0.88	0.89	229
Genuine	0.86	0.88	0.87	183
accuracy			0.88	412
macro avg	0.88	0.88	0.88	412
weighted avg	0.88	0.88	0.88	412

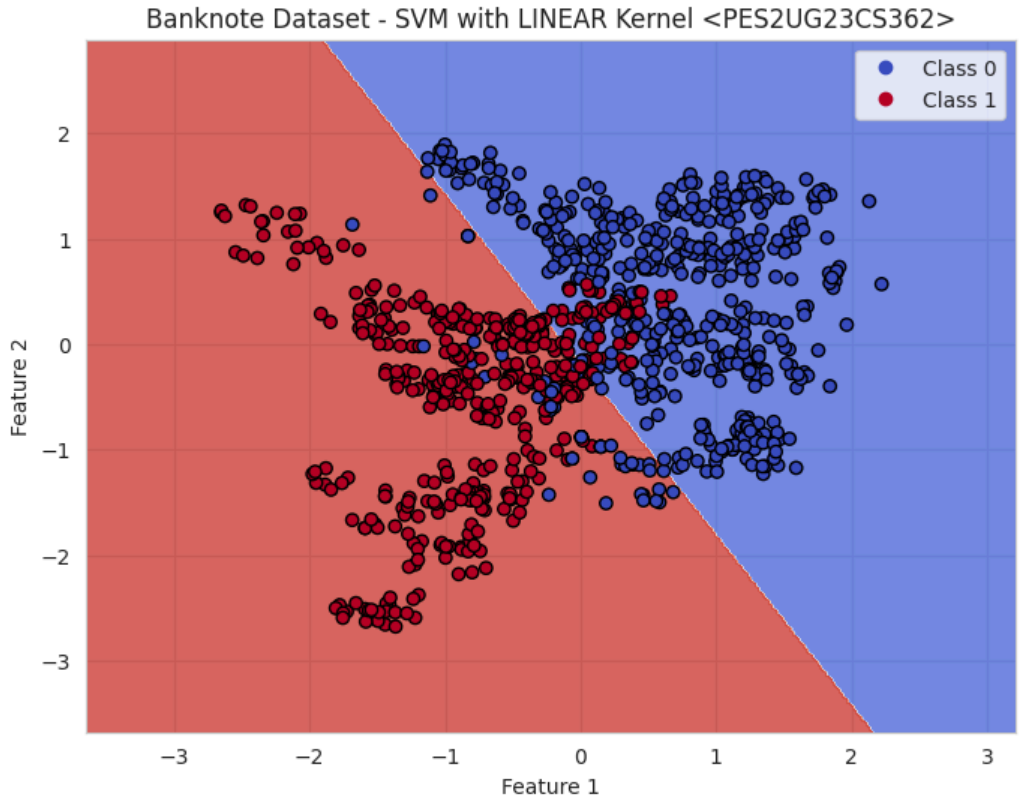
SVM with RBF Kernel <PES2UG23CS362>

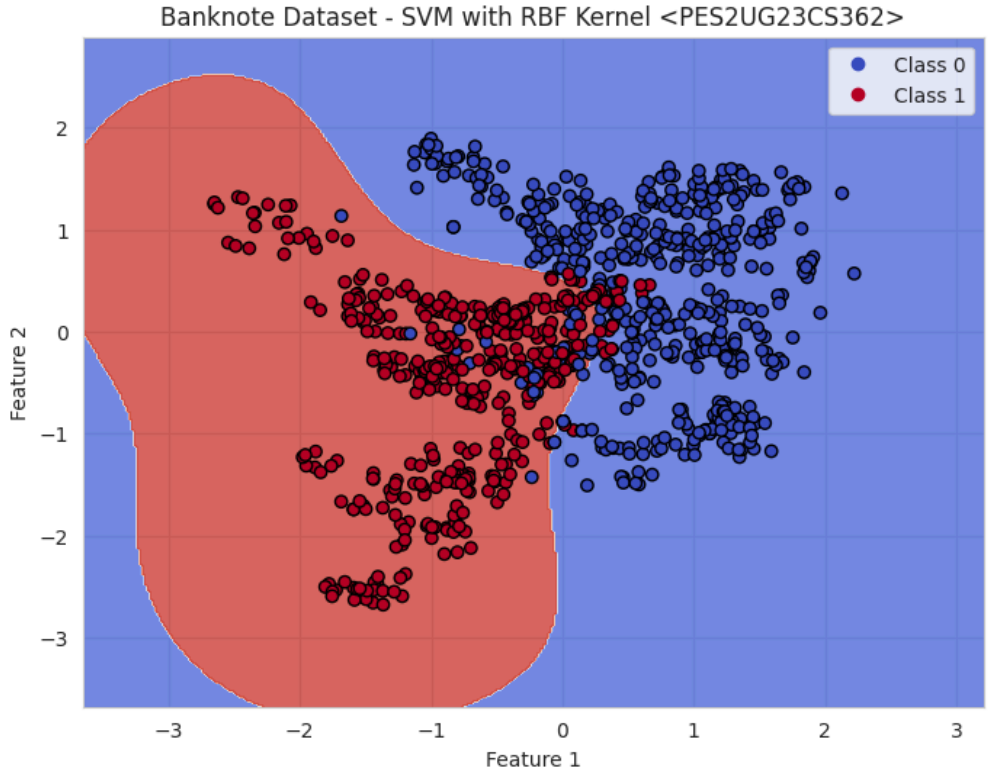
	precision	recall	f1-score	support
Forged	0.96	0.91	0.94	229
Genuine	0.90	0.96	0.93	183
accuracy			0.93	412
macro avg	0.93	0.93	0.93	412
weighted avg	0.93	0.93	0.93	412

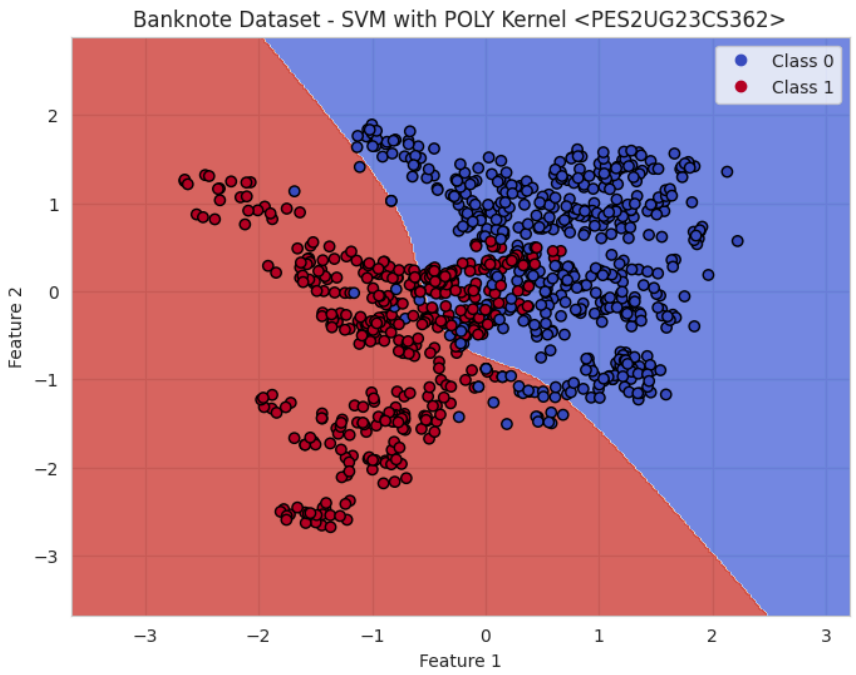
SVM with POLY Kernel <PES2UG23CS362>

	precision	recall	f1-score	support
Forged	0.82	0.91	0.87	229
Genuine	0.87	0.75	0.81	183
accuracy			0.84	412
macro avg	0.85	0.83	0.84	412
weighted avg	0.85	0.84	0.84	412

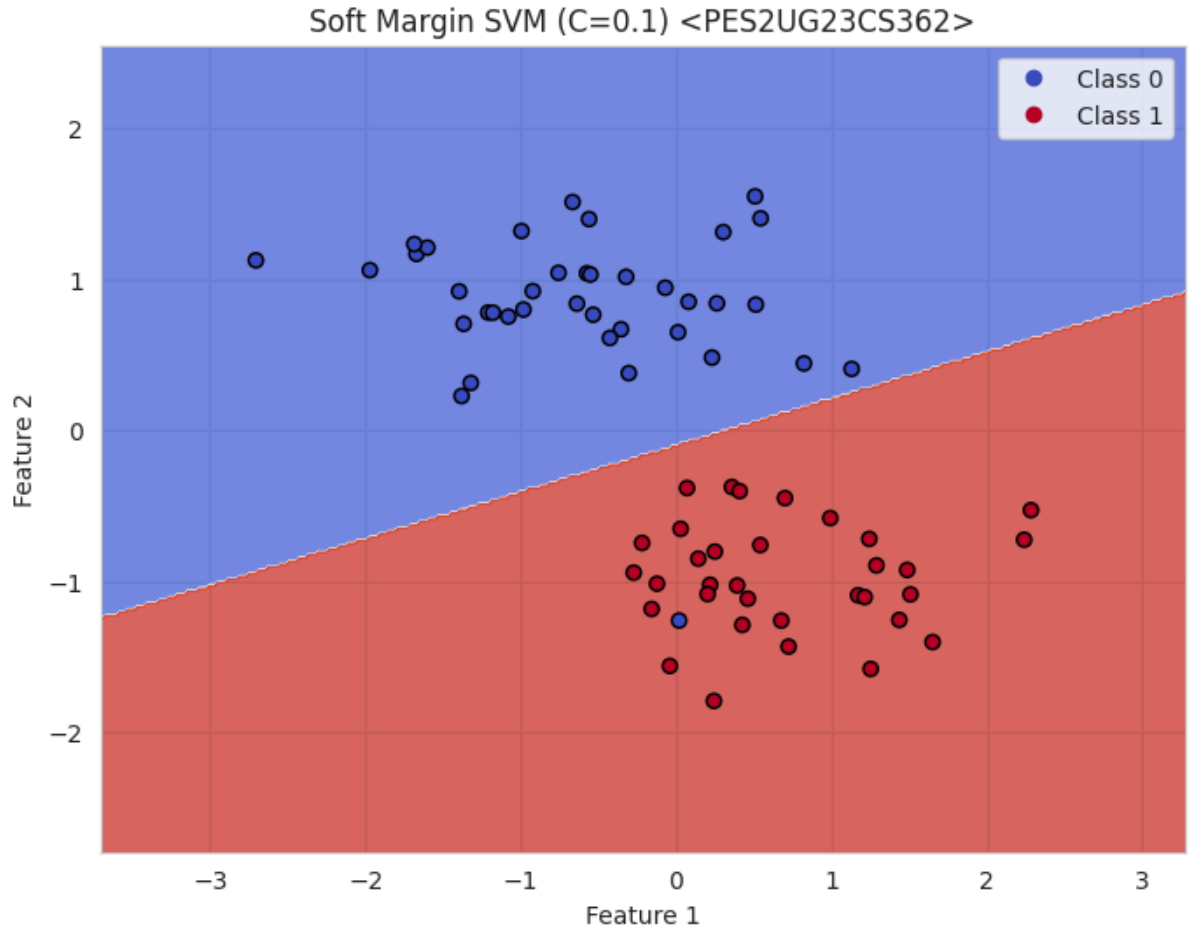
Decision Boundary Visualization

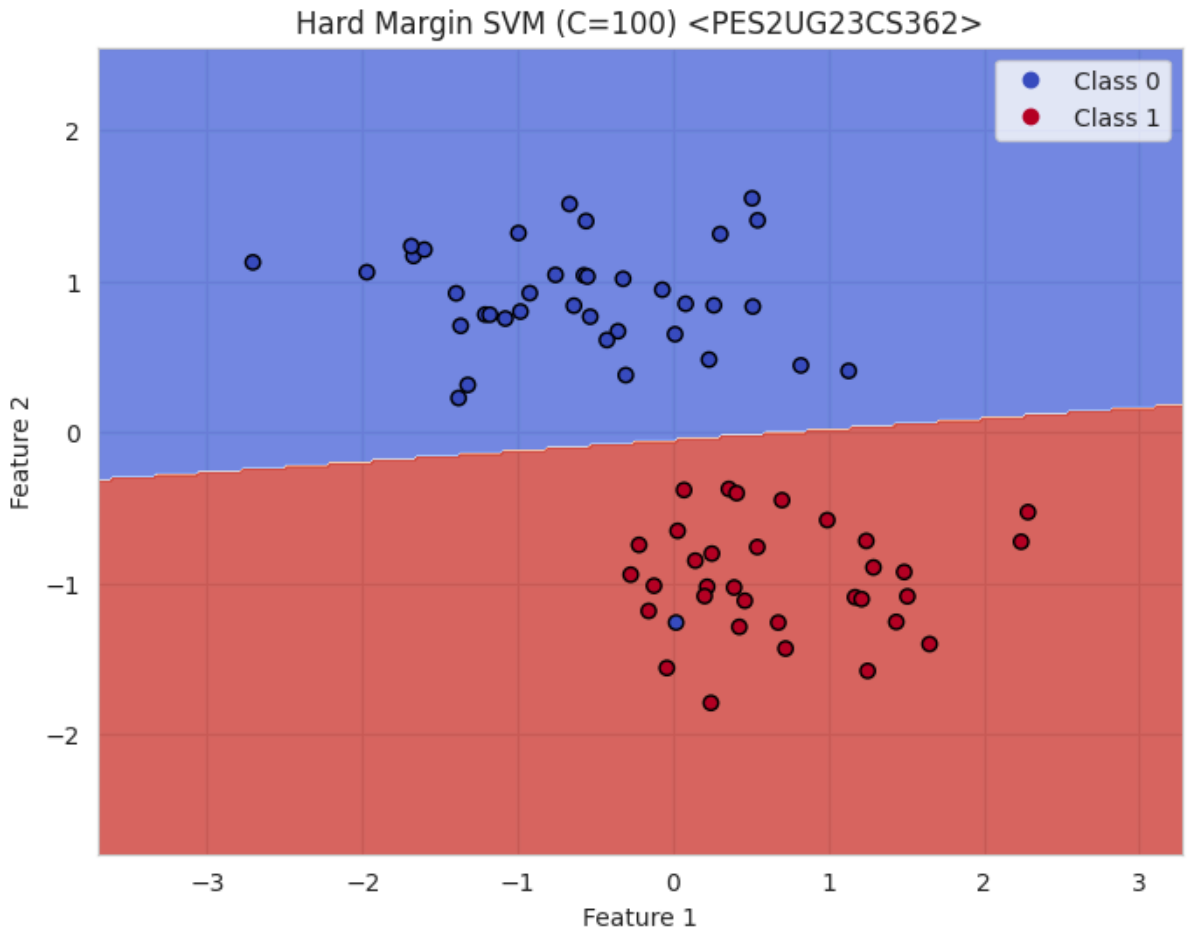






3. Margin Analysis




Hard Margin SVM (C=100) <PES2UG23CS362>
Class 0
Class 1
Feature 2
Feature 1
2
1
0
−1
−2
−3
−2
−1
0
1
2
3