Lab Report

ML Lab-Week 10

SVM Lab

Name: Megha Dhanya

SRN: PES2UG23CS336

Section: F

Course: UE23CS352A – Machine Learning

Date: 12/10/2025

Analysis Questions

A. Moons Dataset

1. Inferences about the Linear Kernel's performance

The linear kernel achieved moderate accuracy (around 87%) but failed to capture the curved structure of the Moons dataset. The decision boundary was a straight line, which resulted in several misclassifications near the overlapping regions. This indicates that a linear SVM is not suitable for data with non-linear patterns.

2. Comparison between RBF and Polynomial kernel decision boundaries

The RBF kernel produced a smoother and more flexible boundary that followed the curved moon shapes accurately. The polynomial kernel also curved but appeared more rigid and slightly less precise. Overall, the RBF kernel fit the dataset's structure more naturally and achieved higher accuracy.

B. Banknote Authentication Dataset

1. Which kernel was most effective for this dataset?

The RBF kernel was the most effective, achieving around 93% accuracy. It handled the slight non-linear separation between forged and genuine banknotes better than other kernels. The linear kernel also performed well since the dataset is almost linearly separable.

2. Why might the Polynomial kernel have underperformed here?

The Banknote dataset is mostly linear, so a polynomial kernel adds unnecessary complexity. The polynomial decision boundary becomes overly curved, which can reduce accuracy when the data doesn't require such flexibility.

C. Hard vs Soft Margin (Linear SVM with C = 0.1 and C = 100)

1. Which margin (soft or hard) is wider?

The soft margin model (C = 0.1) produces a wider margin, allowing more distance between the separating hyperplane and the nearest data points.

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

The soft margin SVM allows some misclassified points to achieve a smoother and more general decision boundary. Its goal is to maximize the margin while maintaining reasonable classification accuracy, rather than perfectly fitting every point.

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

The hard margin model (C = 100) is more likely to overfit since it tries to classify every

training point correctly, including outliers. This creates a very narrow margin that may not generalize well to new data.

4. Which model would be trusted more for new data and why?
The soft margin model is more reliable for unseen data because it is less sensitive to noise and small variations. In real-world scenarios where data is noisy, a lower C value (soft margin) is generally preferred to ensure better generalization.

Screenshots Summary

Training Results (6 Screenshots)

1. Classification Report – SVM with LINEAR Kernel (Moons)

```
SVM with LINEAR Kernel <PES2UG23CS336>
              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

----------------------------------------
```

2. Classification Report – SVM with RBF Kernel (Moons)

```
SVM with RBF Kernel <PES2UG23CS336>
              precision    recall  f1-score   support

           0       0.95      1.00      0.97        75
           1       1.00      0.95      0.97        75

    accuracy                           0.97       150
   macro avg       0.97      0.97      0.97       150
weighted avg       0.97      0.97      0.97       150

----------------------------------------
```

3. Classification Report – SVM with POLY Kernel (Moons)

```
SVM with POLY Kernel <PES2UG23CS336>
              precision    recall  f1-score   support

           0       0.85      0.95      0.89        75
           1       0.94      0.83      0.88        75

    accuracy                           0.89       150
   macro avg       0.89      0.89      0.89       150
weighted avg       0.89      0.89      0.89       150

----------------------------------------
```

4. Classification Report – SVM with LINEAR Kernel (Banknote)

```
SVM with LINEAR Kernel (PES2UG23CS336)
              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

----------------------------------------
```

5. Classification Report – SVM with RBF Kernel (Banknote)

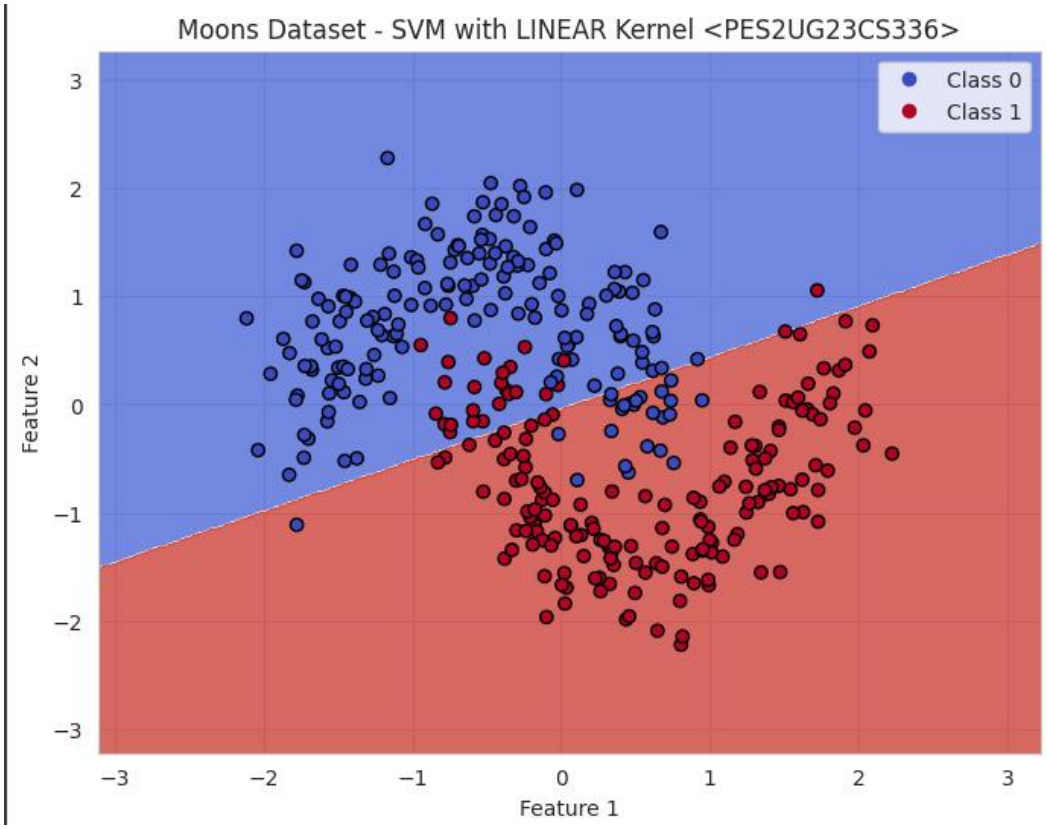
```
SVM with RBF Kernel (PES2UG23CS336)
              precision    recall  f1-score   support

      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

----------------------------------------
```

6. Classification Report – SVM with POLY Kernel (Banknote)

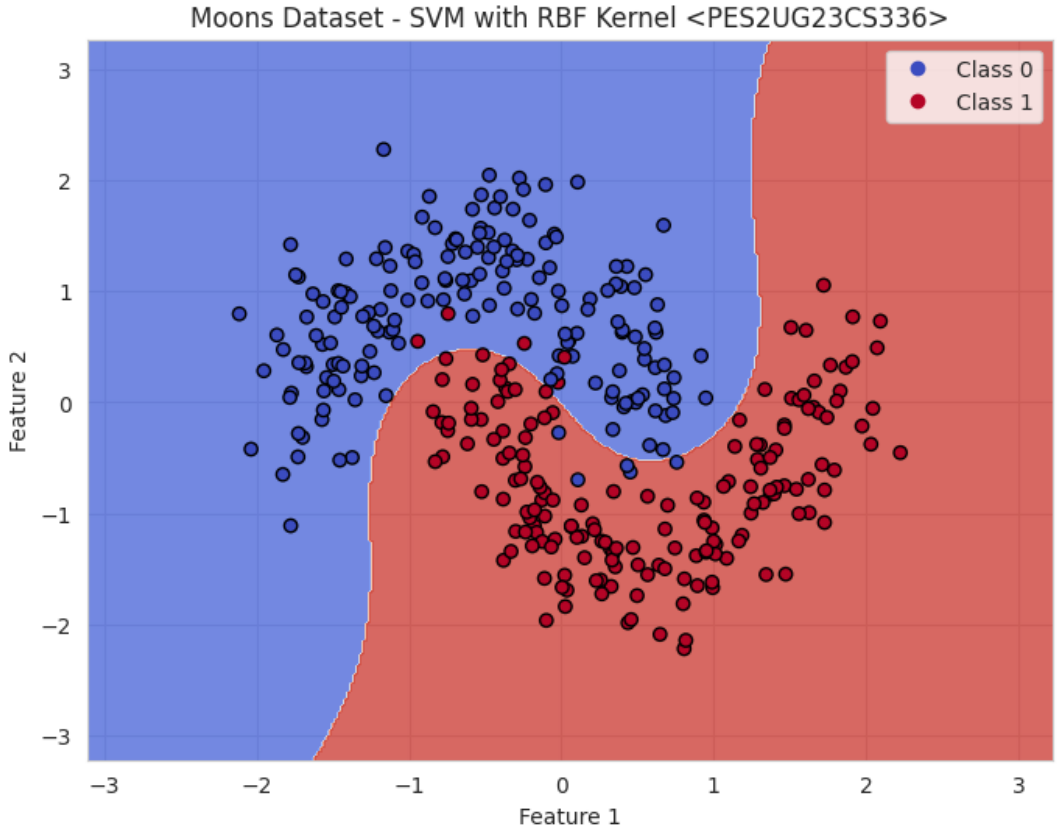
```
SVM with POLY Kernel (PES2UG23CS336)
              precision    recall  f1-score   support

      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 Visualizations (8 Screenshots)

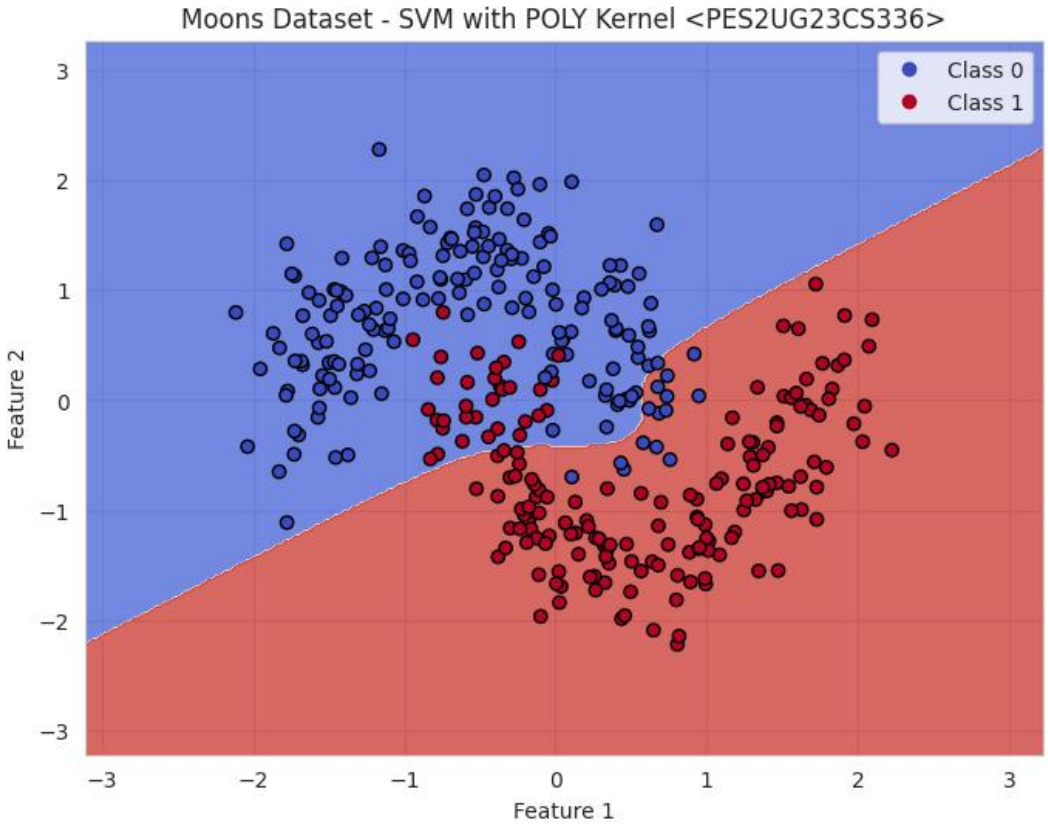
7. Moons Dataset – SVM with LINEAR Kernel



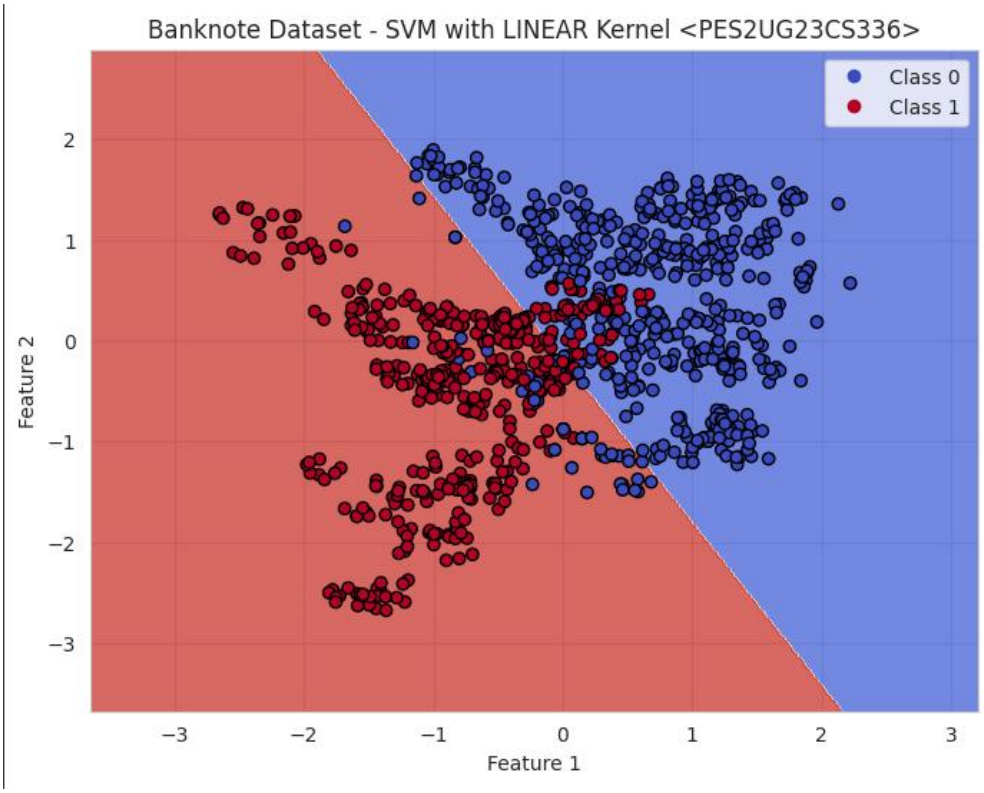
8. Moons Dataset – SVM with RBF Kernel


Moons Dataset - SVM with RBF Kernel <PES2UG23CS336>
Class 0
Class 1
Feature 1
Feature 2
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

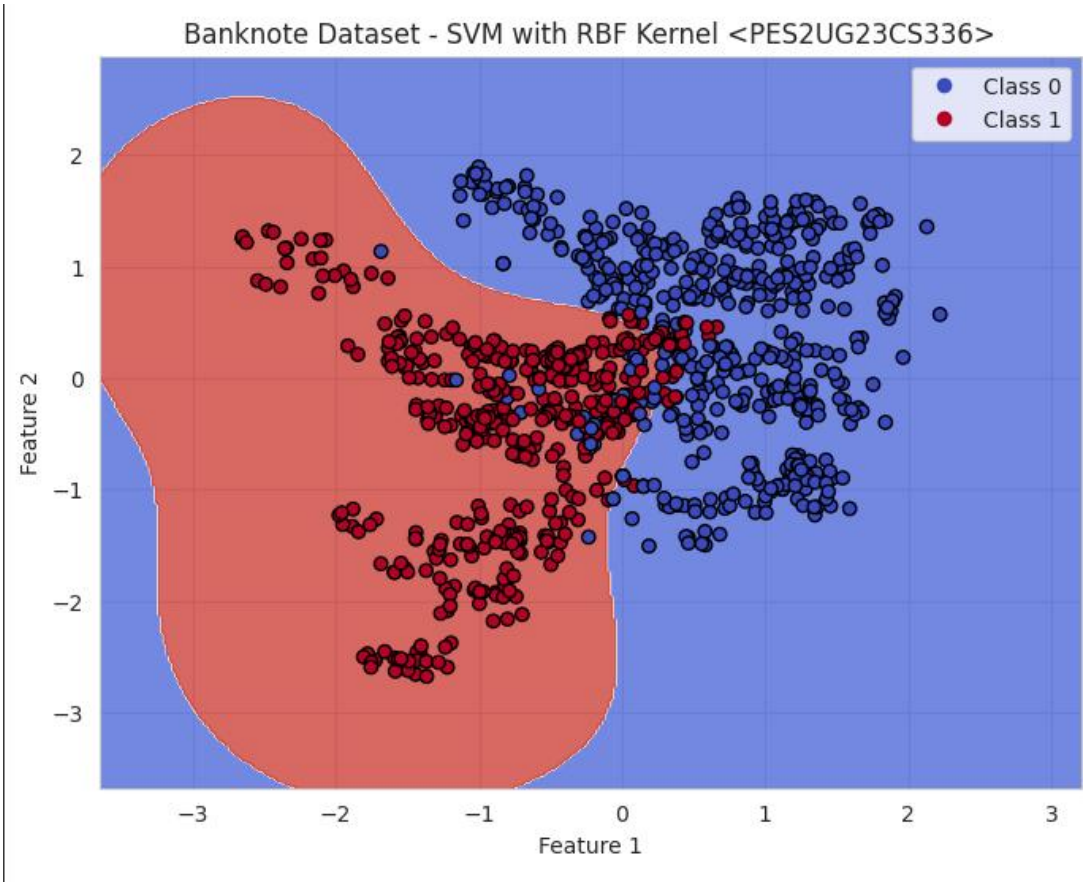
9. Moons Dataset – SVM with POLY Kernel



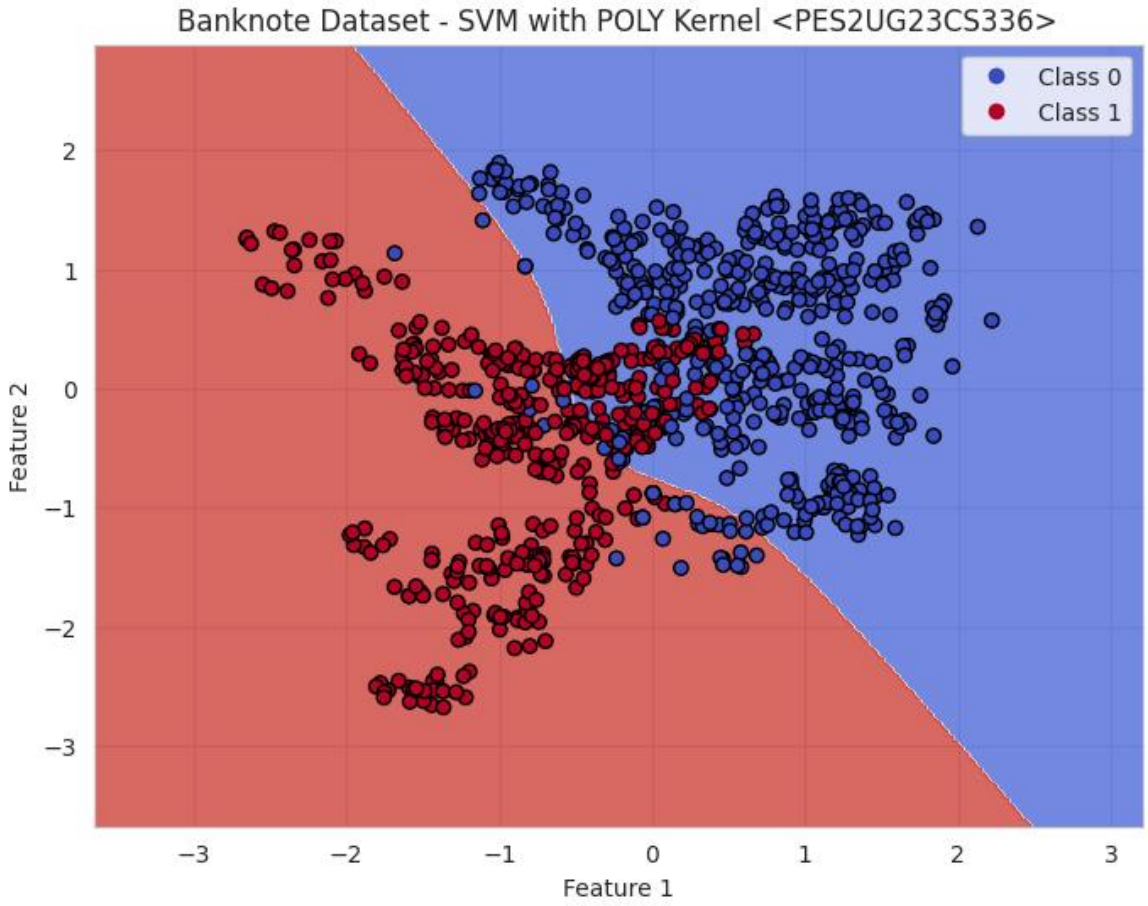
10. Banknote Dataset – SVM with LINEAR Kernel



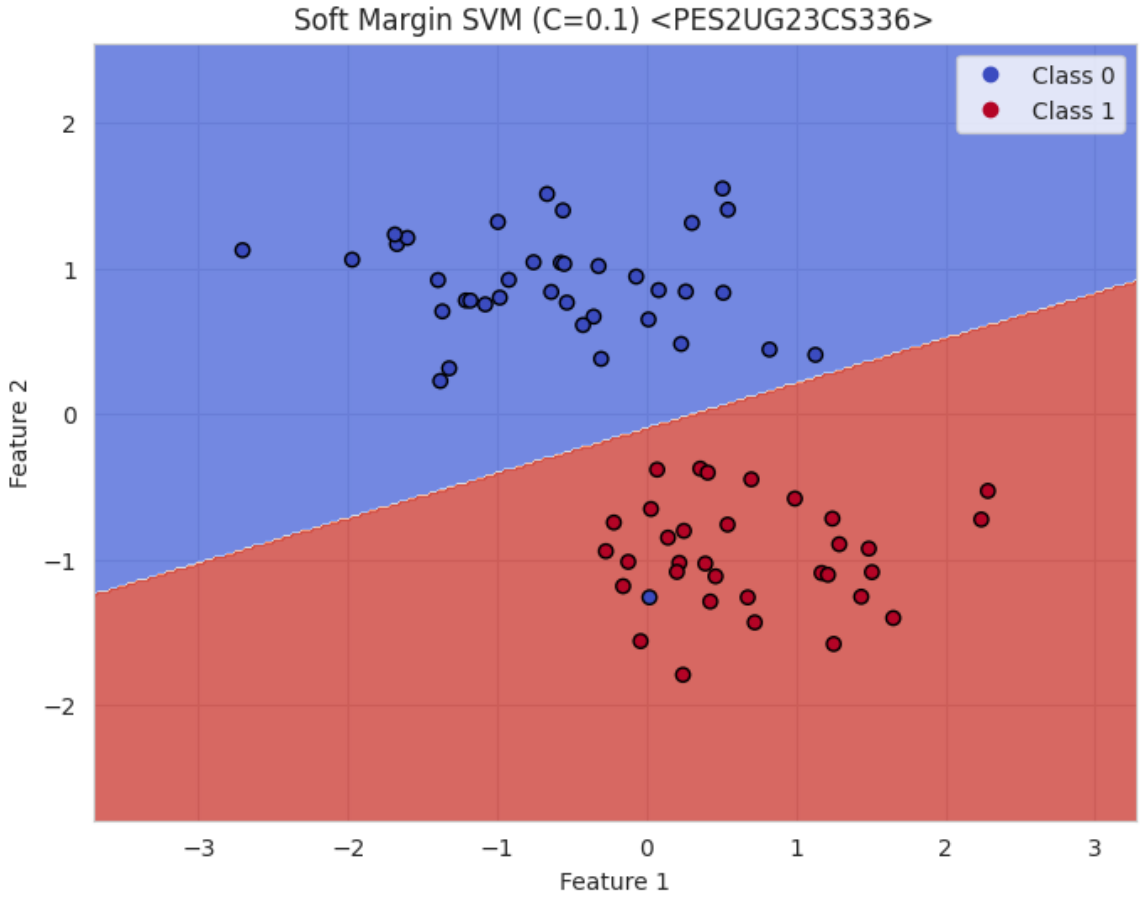
11. Banknote Dataset – SVM with RBF Kernel



12. Banknote Dataset – SVM with POLY Kernel



13. Soft Margin SVM (C = 0.1)



14. Hard Margin SVM (C = 100)