Machine Learning Lab Week- 10 Report

Name: Rithvik Rajesh Matta

SRN: PES2U23CS485

Semester: 5

Sec on: CS H

Analysis Questions

Moons Dataset

1. Inference about the Linear Kernel's Performance

The linear kernel creates a flat decision boundary, which proves inadequate for datasets exhibiting non-linear patterns like the moons dataset. Given that the two crescent-shaped clusters curve around each other, a straight-line separator fails to correctly classify numerous data points, especially in areas where the classes intermingle. This shortcoming is evident in the evaluation metrics—accuracy, precision, recall, and F1-score all fall considerably below those achieved by non-linear kernels. Essentially, the linear kernel cannot accommodate curved structures, rendering it ineffective for the moons dataset, whereas RBF and polynomial kernels deliver substantially superior results.

2. Comparison Between RBF and Polynomial Kernel Decision Boundaries

The RBF (Radial Basis Function) kernel produces a flexible, non-linear decision boundary that wraps tightly around the crescent-shaped data clusters. It successfully distinguishes between the two classes, even where they overlap, by conforming its boundary to match the local data distribution. The polynomial kernel, however, while introducing some curvature, exhibits less flexibility than RBF. Its boundary appears more uniform and constrained, potentially limiting its ability to follow the intricate curves within the moons' configuration.

Conclusion: The RBF kernel shows greater flexibility and accuracy, creating a boundary that follows the data's shape more organically than the polynomial kernel achieves.

Banknote Dataset

1. Which Kernel Was Most Effective for This Dataset?

In the banknote authentication task, the RBF kernel proved most successful. It achieved better accuracy and stronger classification performance compared to both linear and polynomial alternatives. Its capacity for non-linear transformation enables it to detect nuanced differences in the data, making it particularly effective for distinguishing counterfeit from authentic banknotes.

2. Why Might the Polynomial Kernel Have Underperformed Here?

While the polynomial kernel can handle non-linear relationships, its success is highly sensitive to the degree parameter selection. A degree setting that's too conservative may miss complex patterns (resulting in underfitting), whereas an excessive degree can lead to overfitting on training examples. For this particular dataset, the polynomial kernel probably didn't achieve the necessary flexibility or equilibrium to generalize effectively. The RBF kernel, by comparison, adjusts more fluidly to detailed patterns, accounting for its stronger performance.

Hard vs Soft Margin SVM

1. Which Margin (Soft or Hard) Is Wider?

The soft margin configuration (C = 0.1) creates a substantially wider margin. It permits some classification errors to achieve greater flexibility and a smoother separation, emphasizing generalization over perfect training accuracy.

2. Why Does the Soft Margin Model Allow "Mistakes"?

When C is set to a low value, the SVM emphasizes margin maximization rather than strictly enforcing correct classification of every point. This approach enables the model to accept a few incorrectly classified samples if doing so yields a more straightforward, stable decision boundary. This adaptability enhances the model's effectiveness when handling noisy or previously unseen data.

3. Which Model Is More Likely to Overfit and Why?

The hard margin configuration (C = 100) has a higher tendency toward overfitting. By demanding near-perfect classification with little room for error, it generates a tight margin that conforms to even trivial variations or noise present in the training set, compromising its generalization capacity.

4. Which Model Would You Trust More for New Data and Why?

The soft margin SVM is more dependable when working with new or noisy datasets. Its wider margin and acceptance of minor misclassifications enable better generalization, minimizing overfitting risk. In practical machine learning applications, configurations with lower C values are typically favored for their resilience and consistent performance on real-world data.

SCREENSHOTS MOON DATASET

```
SVM with LINEAR Kernel PES2UG23CS485
              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

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

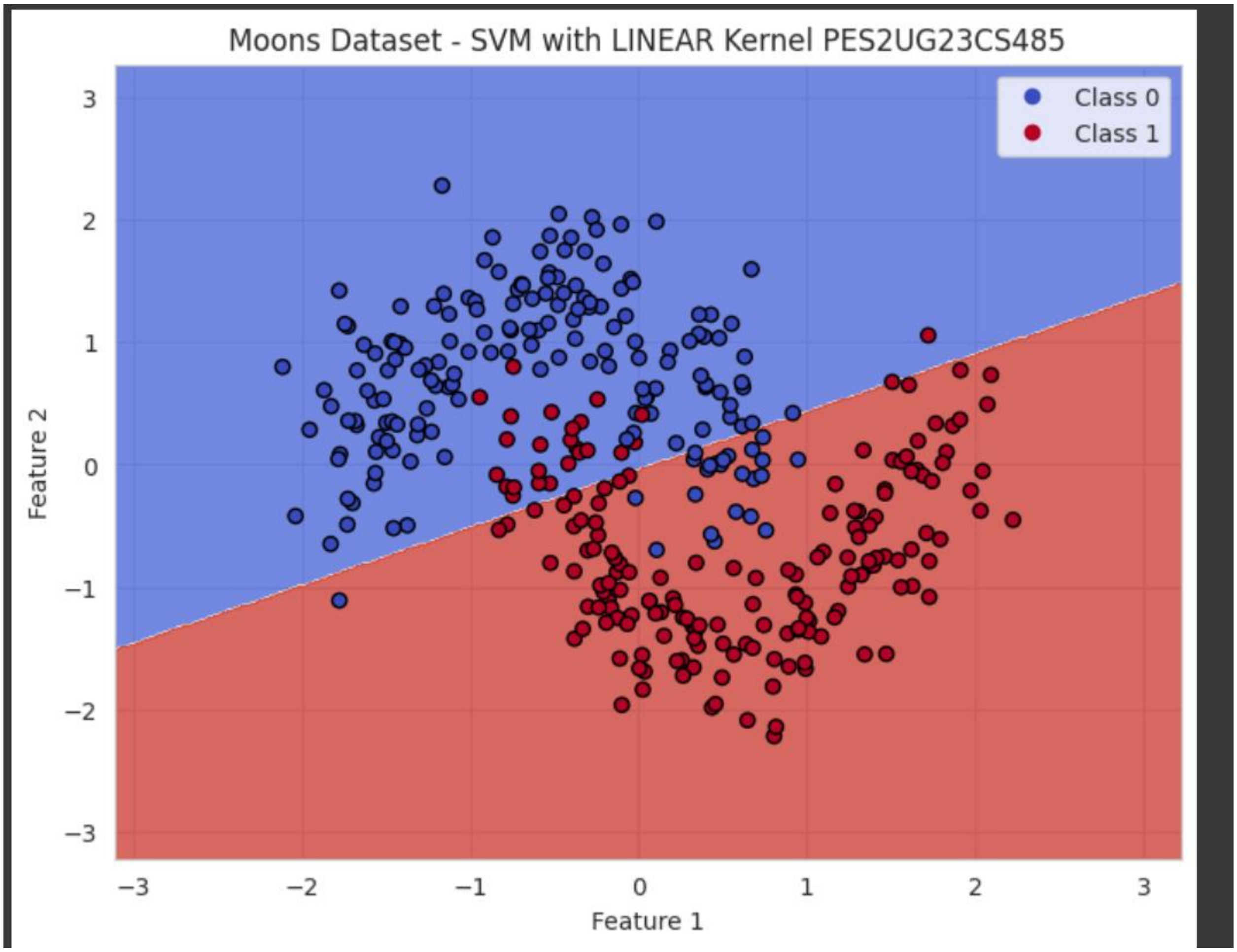
```
SVM with RBF Kernel PES2UG23CS485
              precision    recall  f1-score   support

           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

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

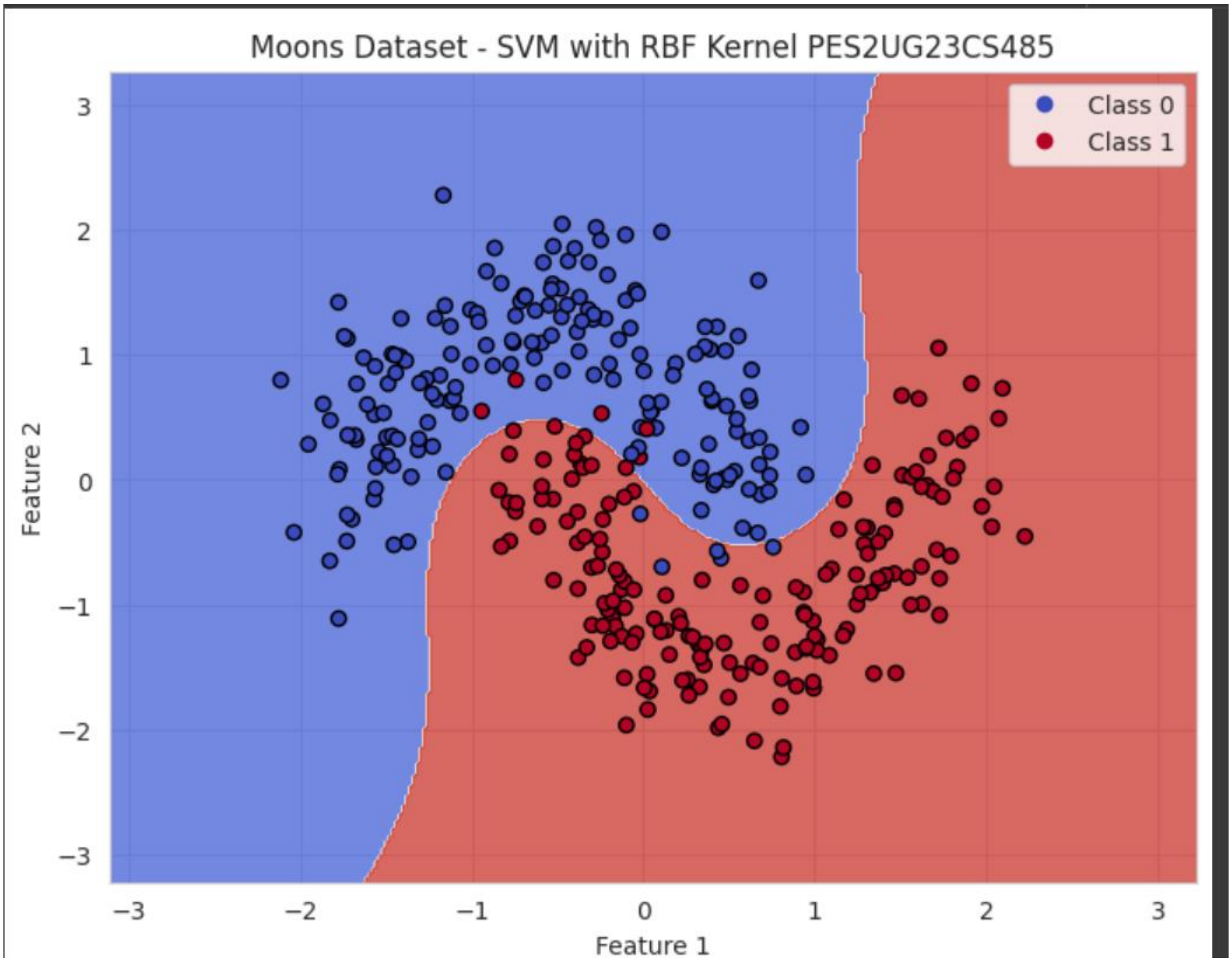
```
SVM with POLY Kernel PES2UG23CS485
              precision    recall  f1-score   support

           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

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

BANKNOTE

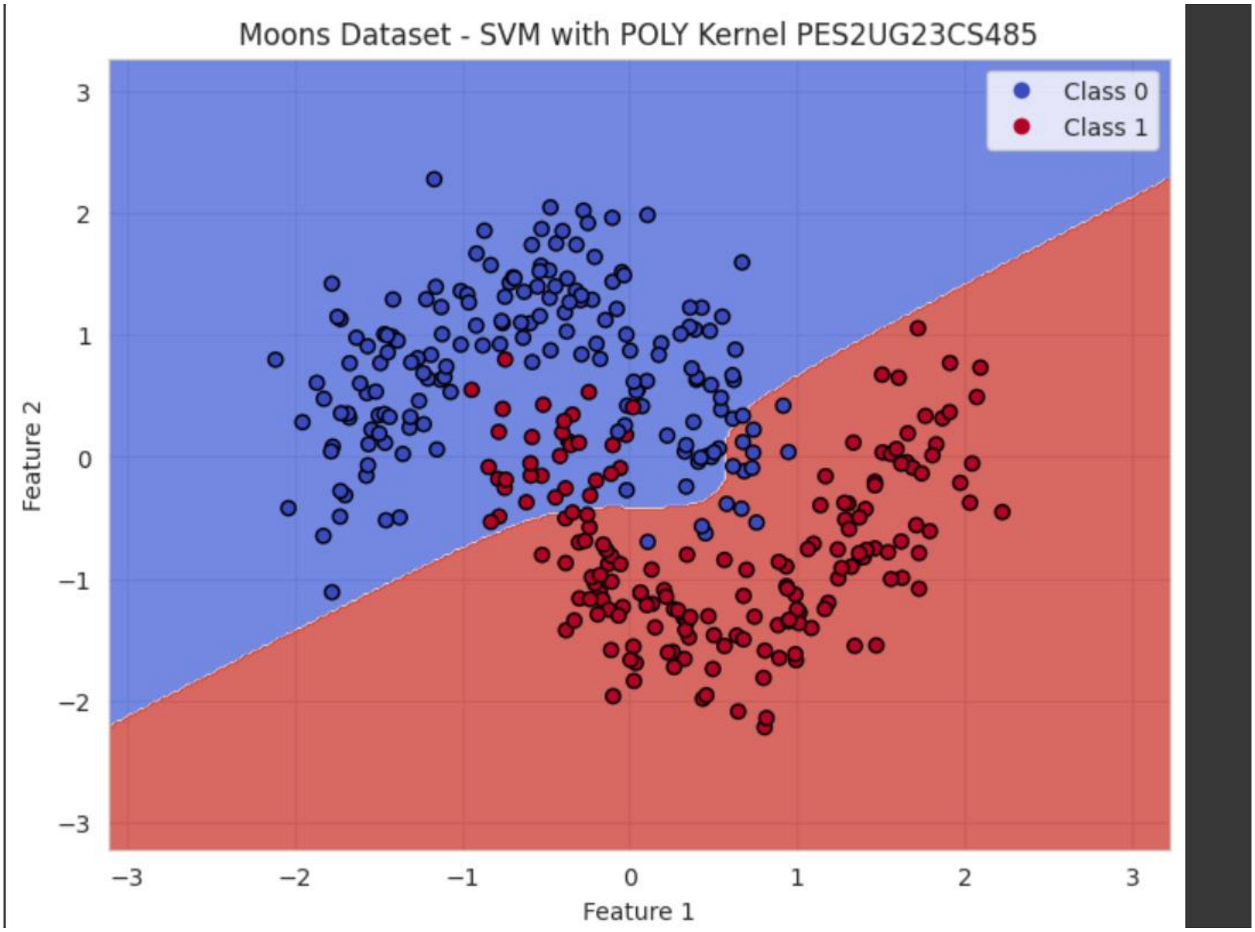
```
SVM with LINEAR Kernel PES2UG23CS485
              precision    recall  f1-score   support

      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 PES2UG23CS485
              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 PES2UG23CS485
              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 MOON DATASET

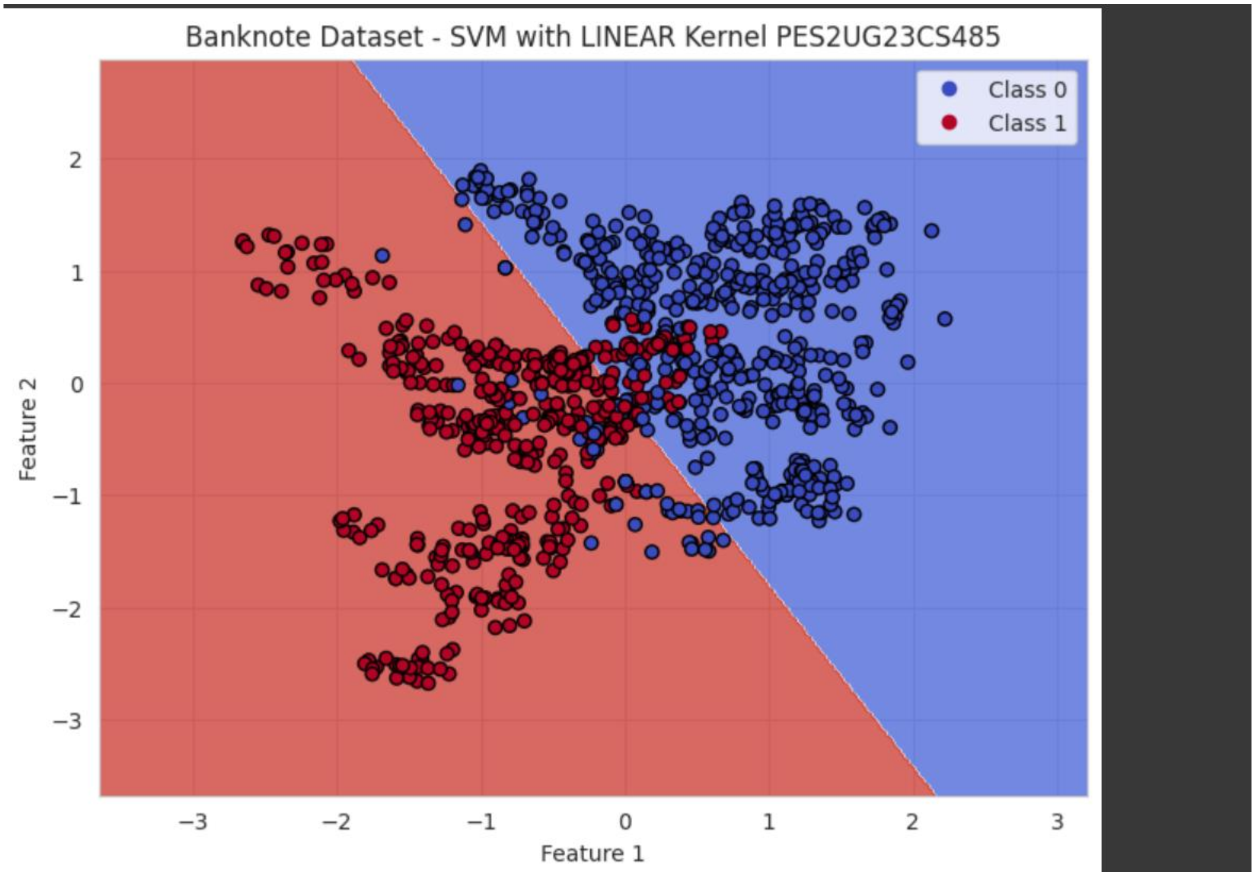

Moons Dataset - SVM with LINEAR Kernel PES2UG23CS485
Class 0
Class 1
Feature 2
Feature 1
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

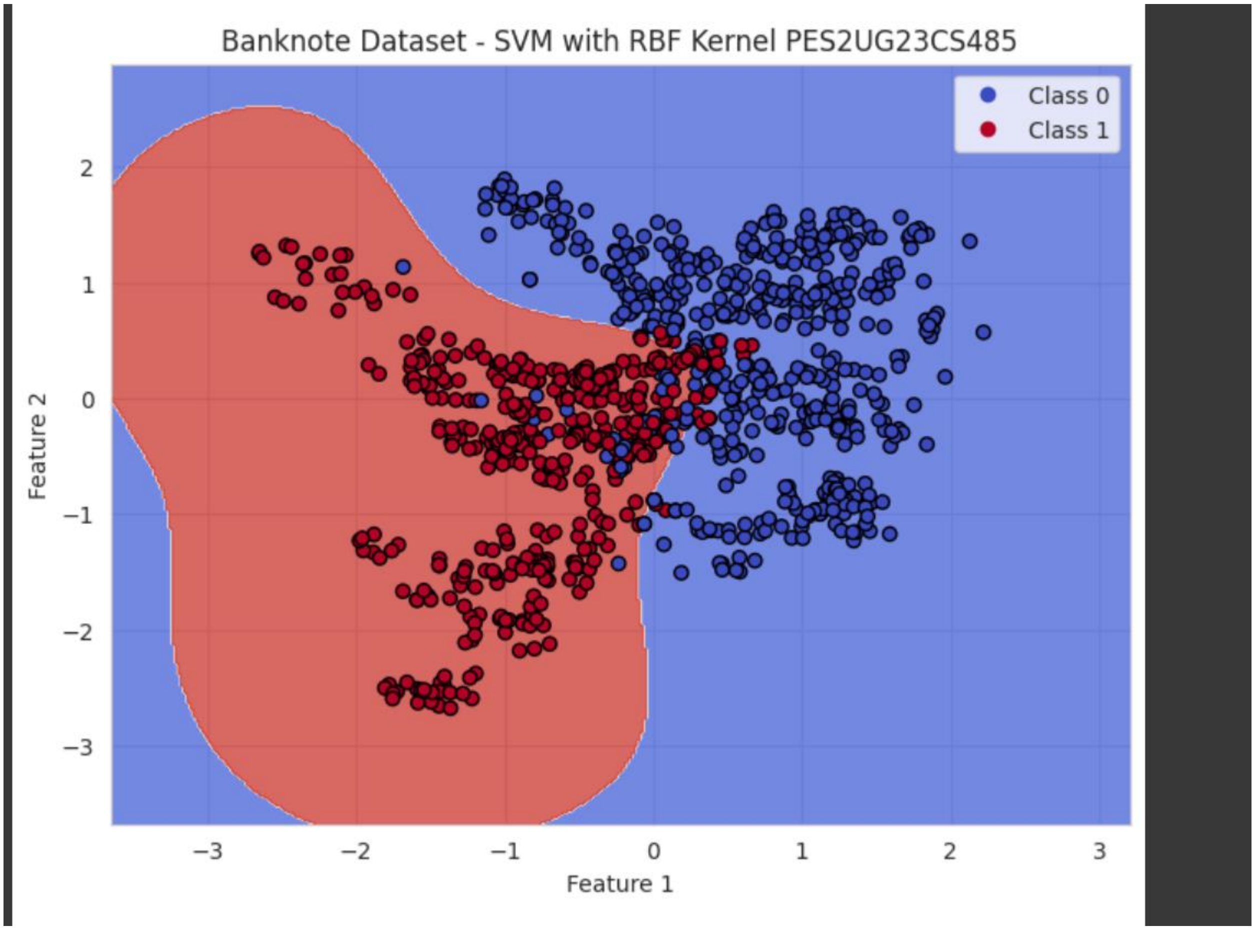


Moons Dataset - SVM with RBF Kernel PES2UG23CS485
Class 0
Class 1
Feature 1
Feature 2
−3
−2
−1
0
1
2
3

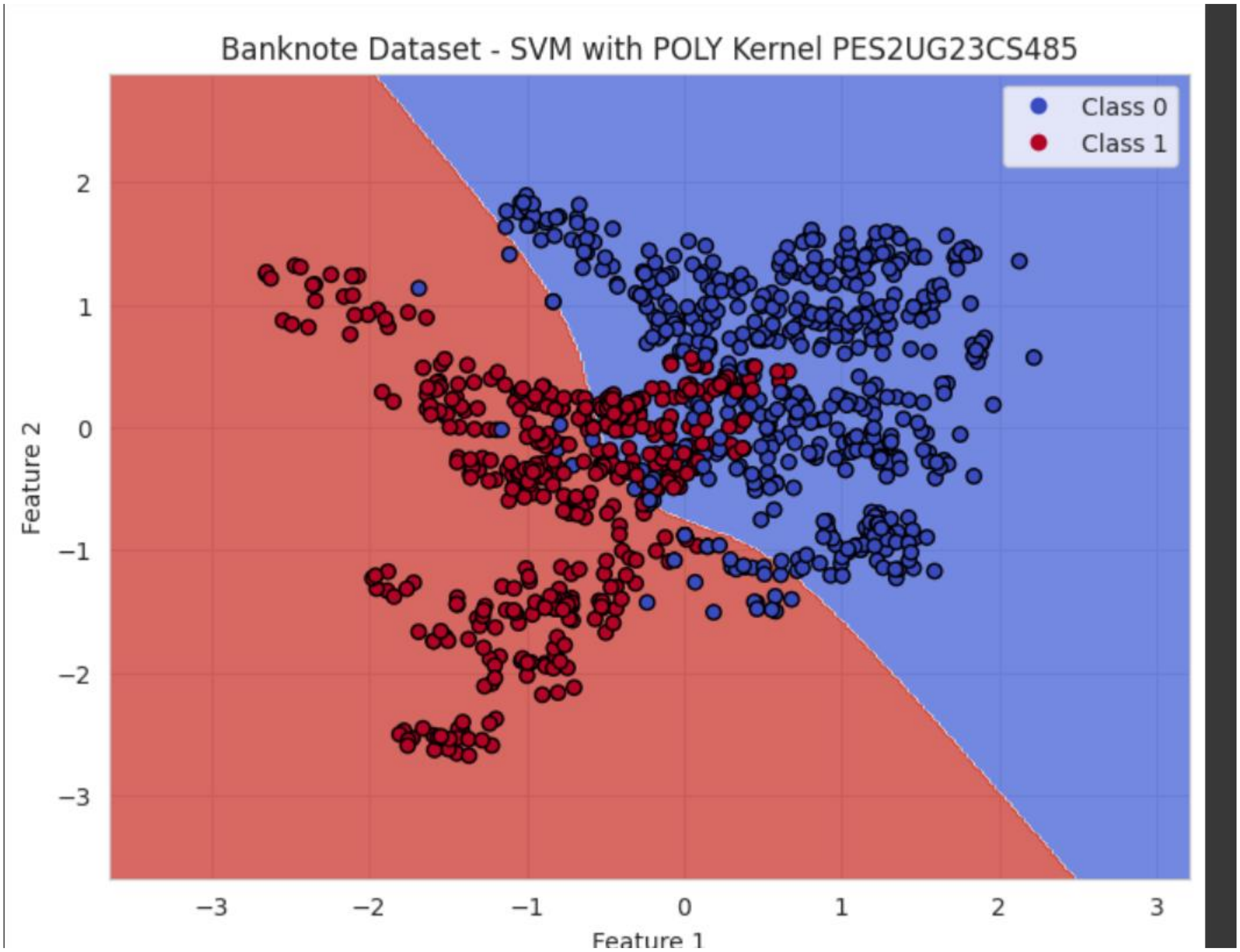

Moons Dataset - SVM with POLY Kernel PES2UG23CS485
Class 0
Class 1
Feature 2
Feature 1
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

BANKNOTE

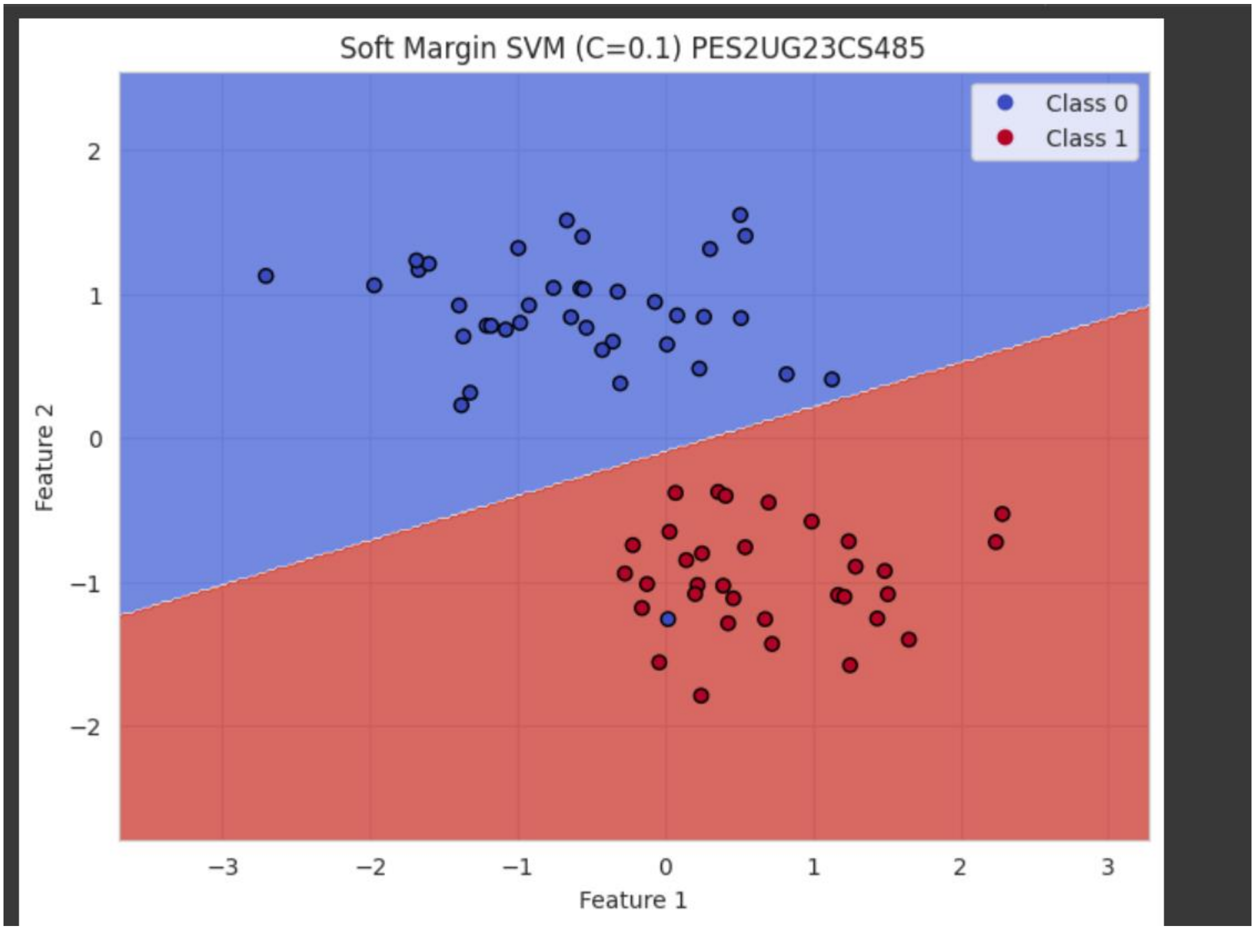

Banknote Dataset - SVM with LINEAR Kernel PES2UG23CS485
Class 0
Class 1
Feature 1
Feature 2
−3
−2
−1
0
1
2
3



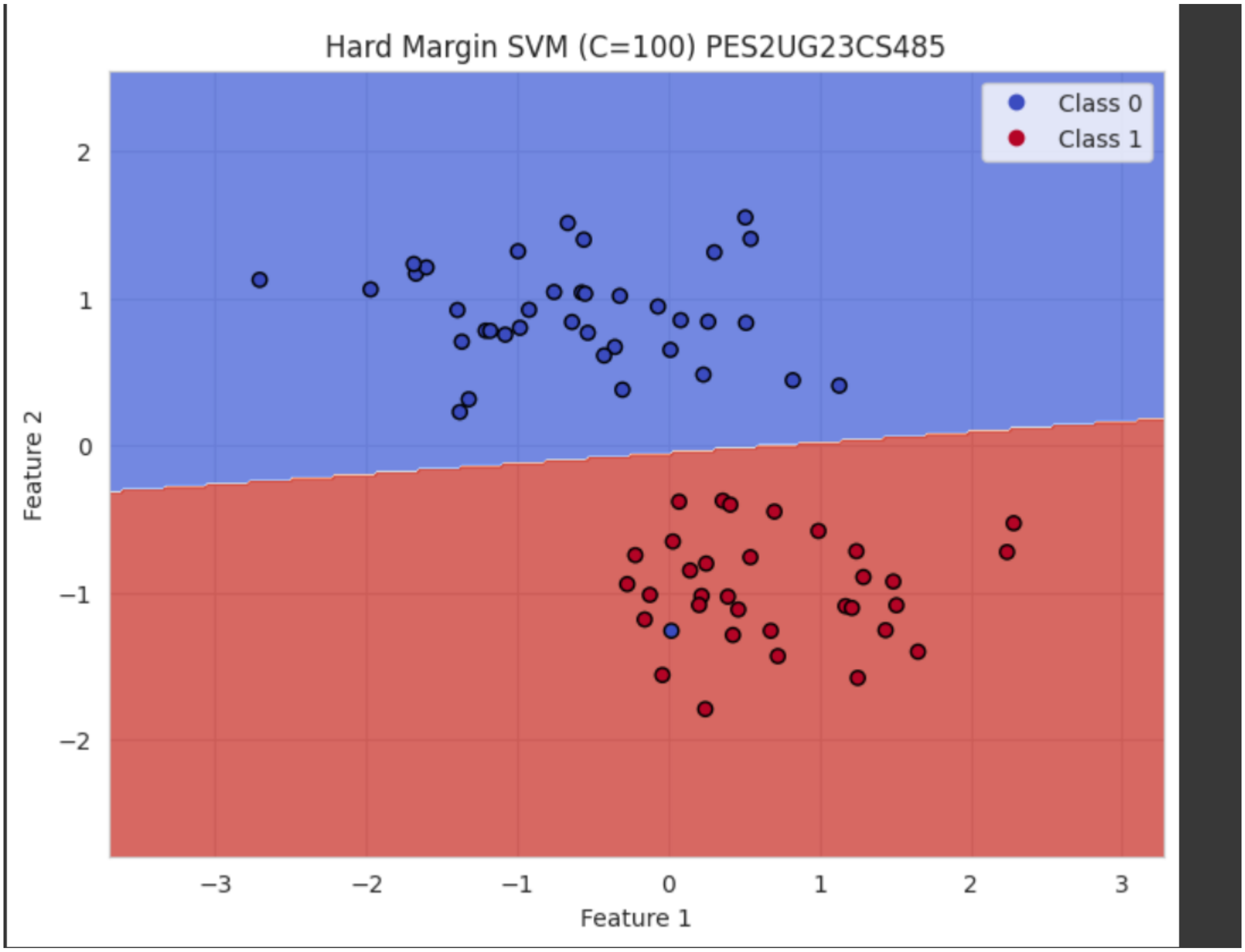
Banknote Dataset - SVM with RBF Kernel PES2UG23CS485
Class 0
Class 1
Feature 2
Feature 1
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3



MARGIN ANALYSIS (2 PLOT)



Soft Margin SVM (C=0.1) PES2UG23CS485
Class 0
Class 1
Feature 1
Feature 2
−3
−2
−1
0
1
2
3


Hard Margin SVM (C=100) PES2UG23CS485
Class 0
Class 1
Feature 2
Feature 1
2
1
0
−1
−2
−3
−2
−1
0
1
2
3