ML Lab

Lab 10 - Submission

Name: Gauthamdev R Holla

SRN: PES2UG23CS197

Branch: CSE

Sem: V

Section: C

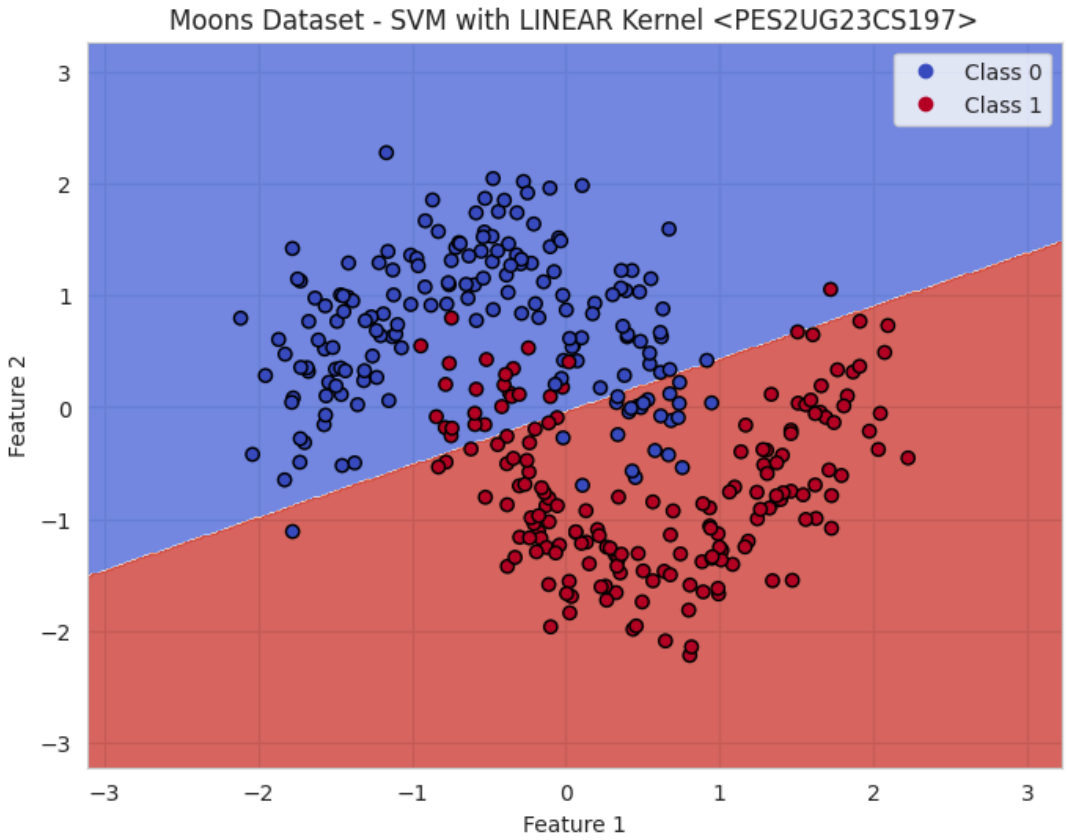
---

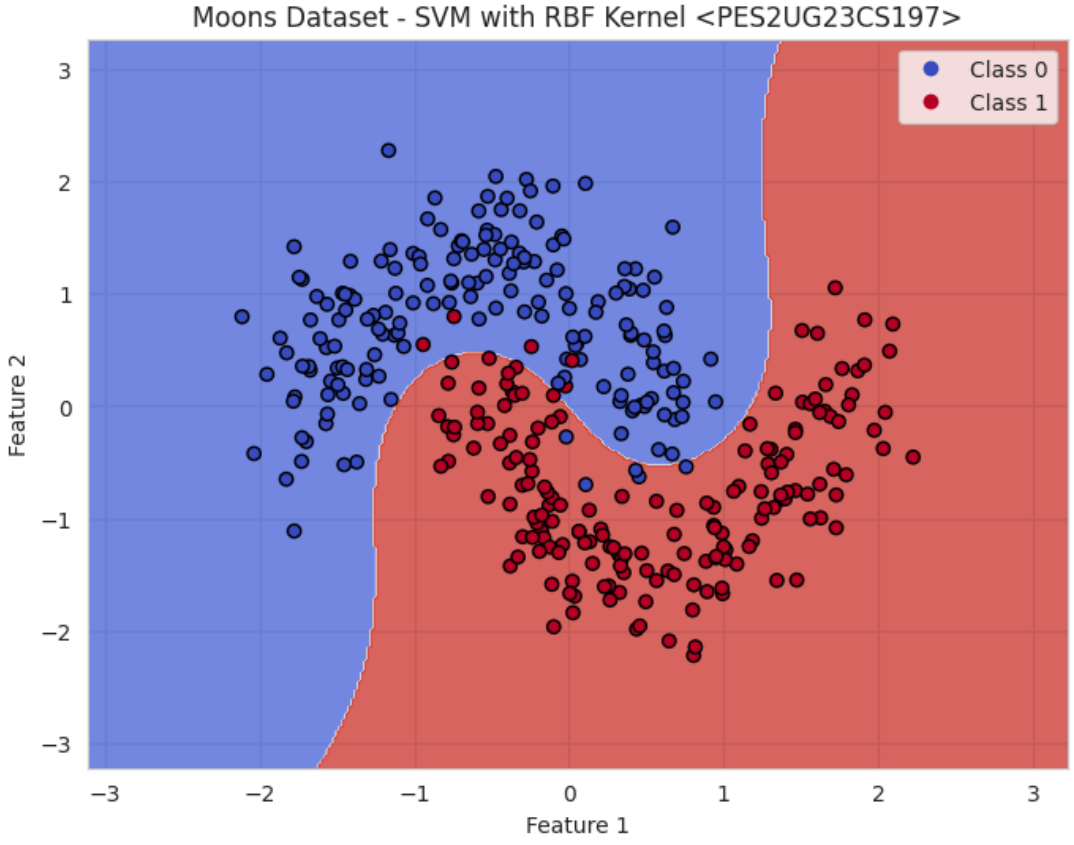
1) Moons Dataset

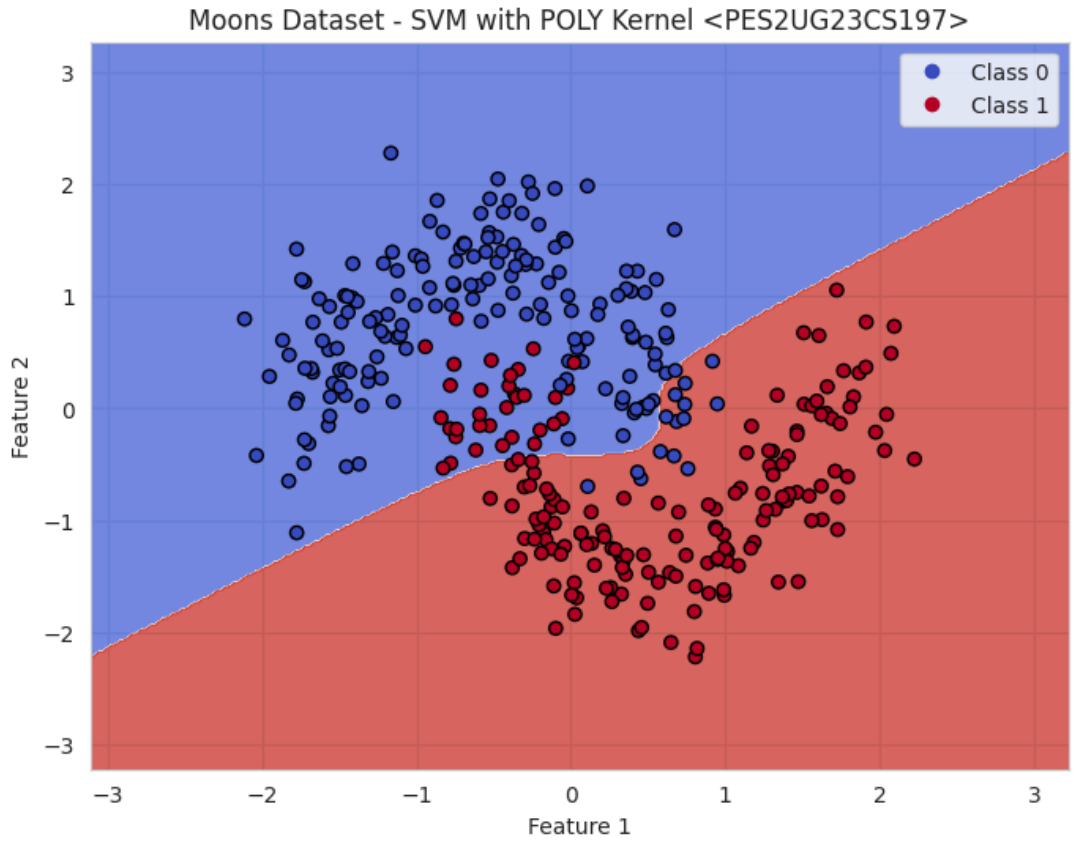
```
SVM with LINEAR Kernel PES2UG23CS197
              precision    recall  f1-score   support

     Class 0       0.85      0.89      0.87        75
     Class 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 PES2UG23CS197
              precision    recall  f1-score   support

     Class 0       0.95      1.00      0.97        75
     Class 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 PES2UG23CS197
              precision    recall  f1-score   support

     Class 0       0.85      0.95      0.89        75
     Class 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

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


Moons Dataset - SVM with LINEAR Kernel <PES2UG23CS197>
Class 0
Class 1
Feature 1
Feature 2
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3


Moons Dataset - SVM with RBF Kernel <PES2UG23CS197>
Class 0
Class 1
Feature 1
Feature 2
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3



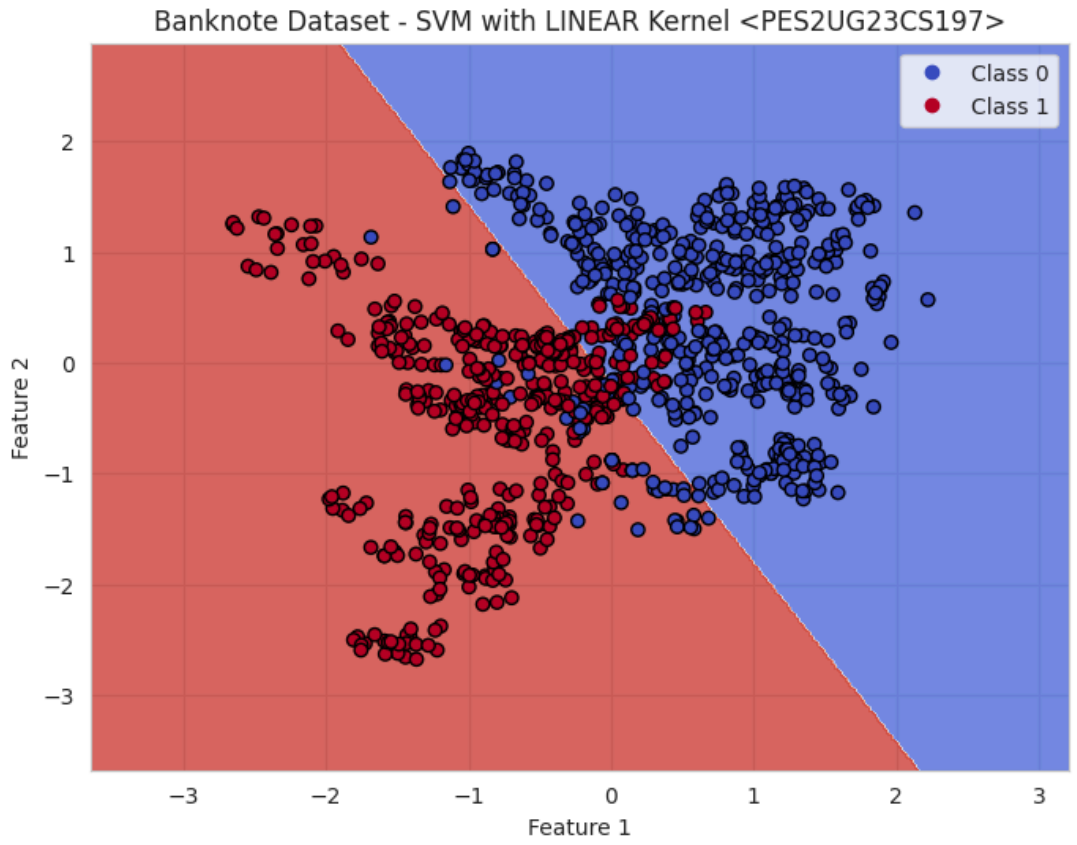
---

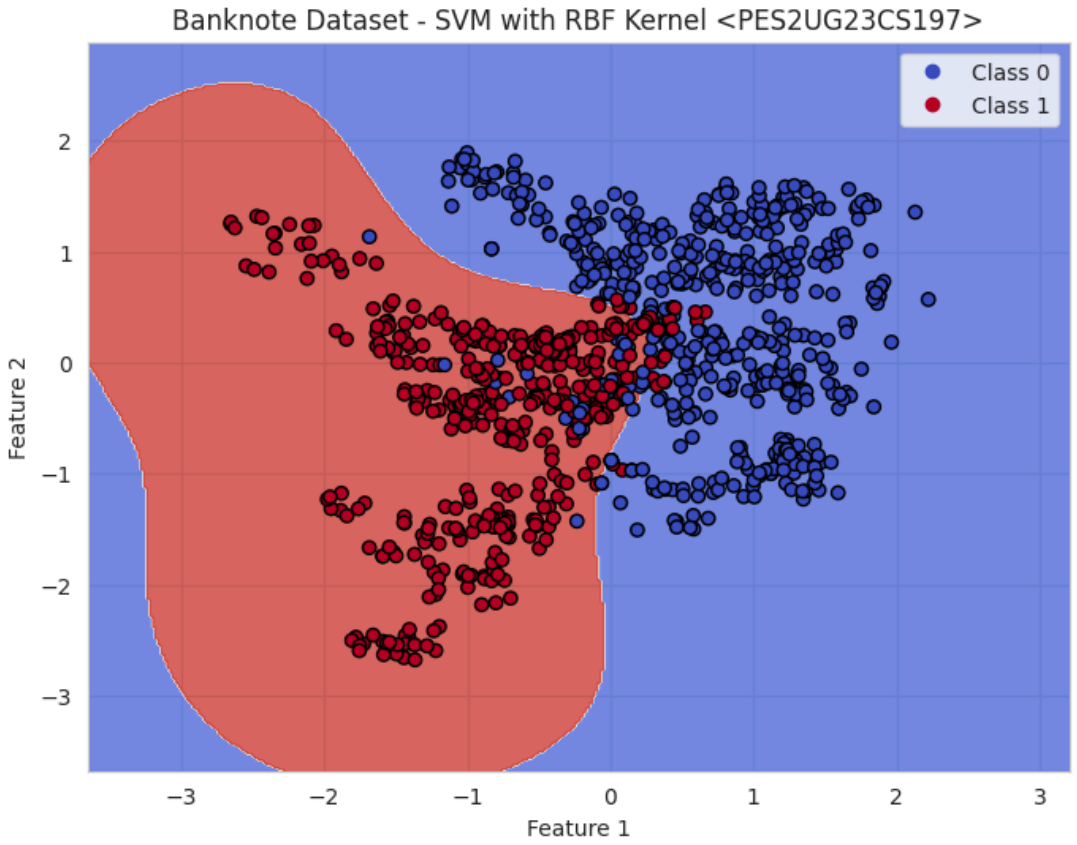
2) Banknote Dataset

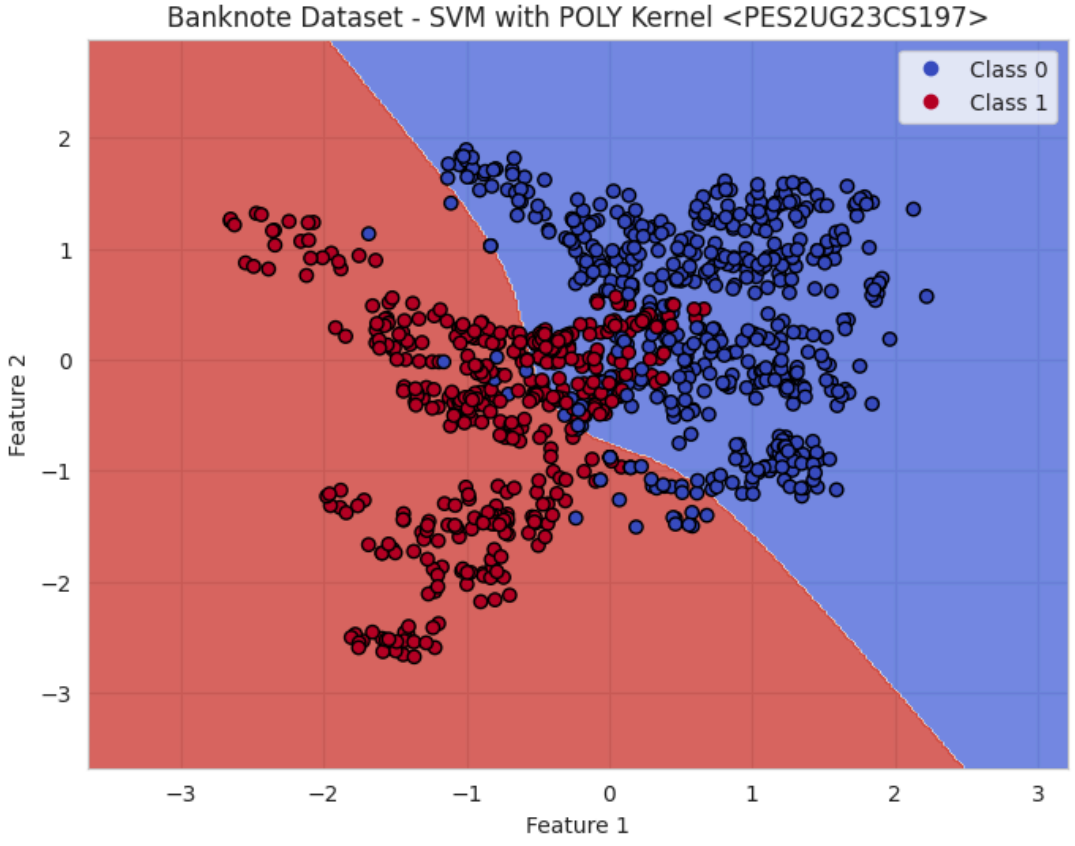
```
SVM with LINEAR Kernel PES2UG23CS197
              precision    recall  f1-score   support

      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 PES2UG23CS197
              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

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

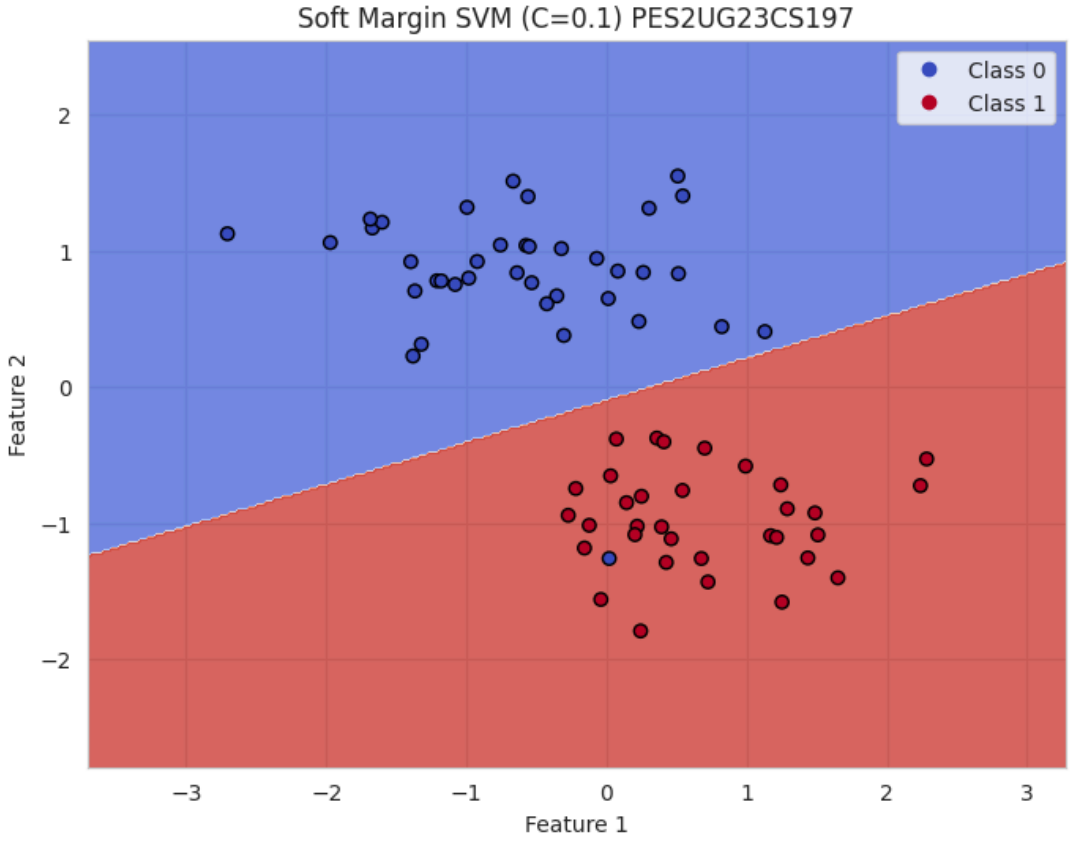
```
SVM with POLY Kernel PES2UG23CS197
              precision    recall  f1-score   support

      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

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

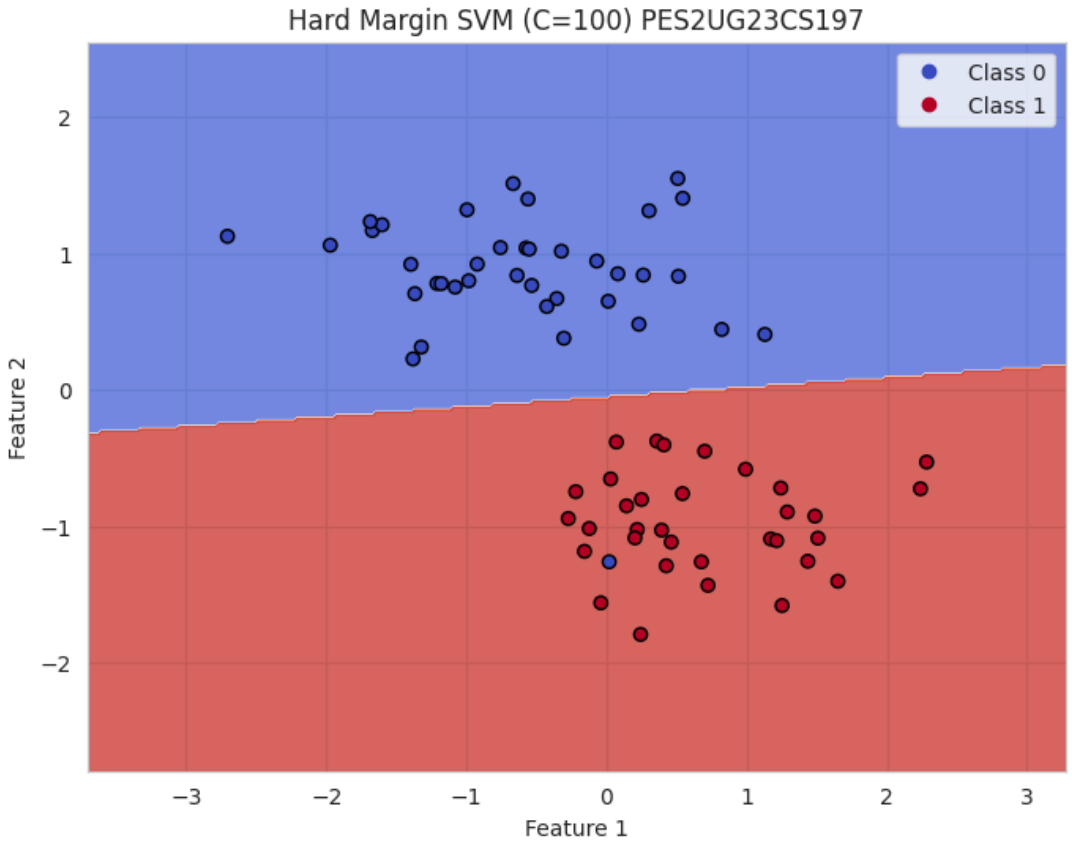



Banknote Dataset - SVM with RBF Kernel <PES2UG23CS197>
Class 0
Class 1
Feature 2
Feature 1
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3


Banknote Dataset - SVM with POLY Kernel <PES2UG23CS197>
Class 0
Class 1
Feature 2
Feature 1
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

3) Margin Analysis





---

4) Analysis Questions

Moons Dataset Questions:

1. Inferences about the Linear Kernel's performance.

- Overall accuracy of 87%
- Good precision and recall.
- Moons dataset is non-linearly separable. So, it misclassifies points near the overlapping regions of the two classes.

2. Comparison between RBF and Polynomial kernel decision boundaries.

- Both RBF and Polynomial kernels outperform the Linear Kernel.
- RBF Kernel Accuracy = 97%
- Polynomial Kernel Accuracy = 89%
- RBF Kernel has flexibility in modeling nonlinear relationships. So, it adapts well to the Moons dataset.
- Polynomial Kernel is less adaptive than RBF but better than Linear.

Banknote Dataset Questions:

1. Which kernel was most effective for this dataset?

- RBF Kernel is more effective with 93% accuracy.
- Outperforms both Linear and Polynomial. This indicates that the data has non-linear patterns.
- It separates "Forged" and "Genuine" classes more accurately due to its ability to model complex decision boundaries.

2. Why might the Polynomial kernel have underperformed here?

- The Banknote dataset didn't exhibit polynomial-like separation, making it less effective.
- So, the Polynomial kernel may have introduced unnecessary complexity, leading to overfitting or poor generalization.
- Unlike the Moons dataset, the Banknote dataset features are better separated with simpler, smoother boundaries.

Performance Discussion Hard vs. Soft Margin Questions:

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

Soft Margin SVM produces a wider margin.

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

The primary goal is generalization. It accepts a few mistakes to perform well on future inputs too.

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

Hard Margin as fit forces the model to avoid misclassifications at all costs, which can lead to memorizing the training data rather than learning general patterns.

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

Soft Margin as it is better suited to handle variability and uncertainty. Low C value is generally recommended when working with noisy datasets, as it encourages the model to focus on broader trends.
