ML_LAB_WEEK_10_SVM_LAB

Name: Meghana Saisri Bisa
SRN: PES2UG23CS337
Section: F

PART 1: Moons Dataset Questions

1. Inferences about the Linear Kernel's performance
    - The Linear Kernel performs poorly on the Moons dataset.
    - Reason: The data is not linearly separable (two interlocking moons), so a straight line cannot separate the classes properly.
    - Result: Low accuracy and many misclassified points.

2. Comparison between RBF and Polynomial kernel decision boundaries
    - RBF Kernel: Creates a smooth, curved boundary that closely follows the shape of the moons. Handles non-linear patterns very well.
    - Polynomial Kernel: Also creates a curved boundary, but sometimes it overfits or underfits depending on the degree.
    - Conclusion: RBF generally captures the moons' shape more naturally and accurately.

PART 2: Banknote Dataset Questions

1. Which kernel was most effective for this dataset?
    - RBF Kernel is usually the most effective.
    - Reason: It handles non-linear patterns in the features better than a simple linear or polynomial kernel for this dataset.

2. Why might the Polynomial kernel have underperformed here?
    - Polynomial can overfit if the degree is too high or underfit if the degree is too low.
    - The Banknote dataset is mostly linearly separable with some non-linear noise, so RBF works better. Polynomial struggles to generalize.

PART 3: Hard vs Soft Margin Questions

1. Which margin (soft or hard) is wider?
    - Soft Margin (C=0.1) produces a wider margin.
    - Reason: It allows some misclassifications to create a bigger separation between classes.

2. Why does the soft margin model allow "mistakes"?
    - It tolerates some points inside or across the margin to prevent overfitting.
    - Goal: Balance accuracy on training data and generalization to new data.

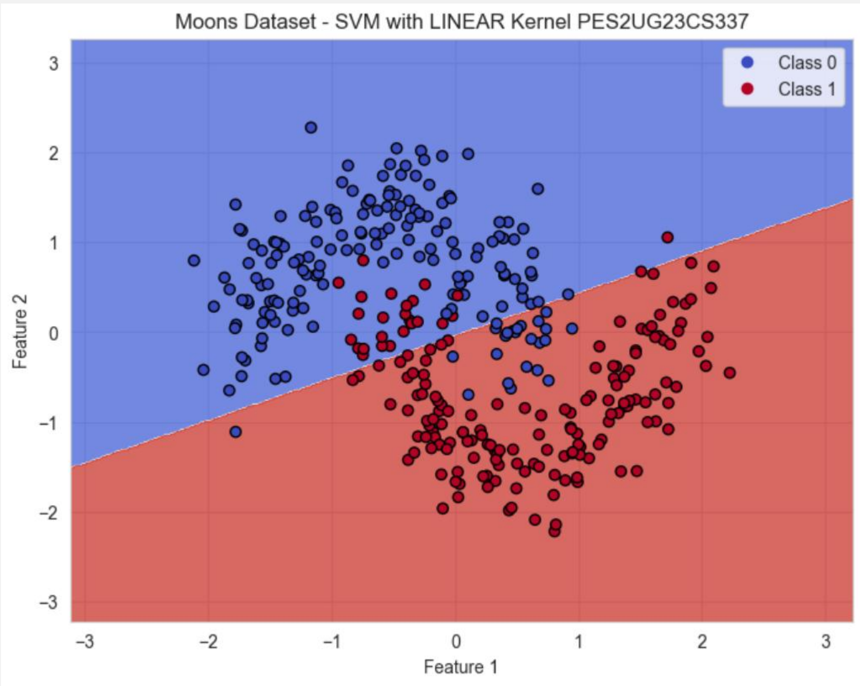
3. Which model is more likely to be overfitting and why?
    - Hard Margin (C=100) is more likely to overfit.
    - Reason: It tries to classify all training points perfectly, even outliers, which reduces generalization.

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

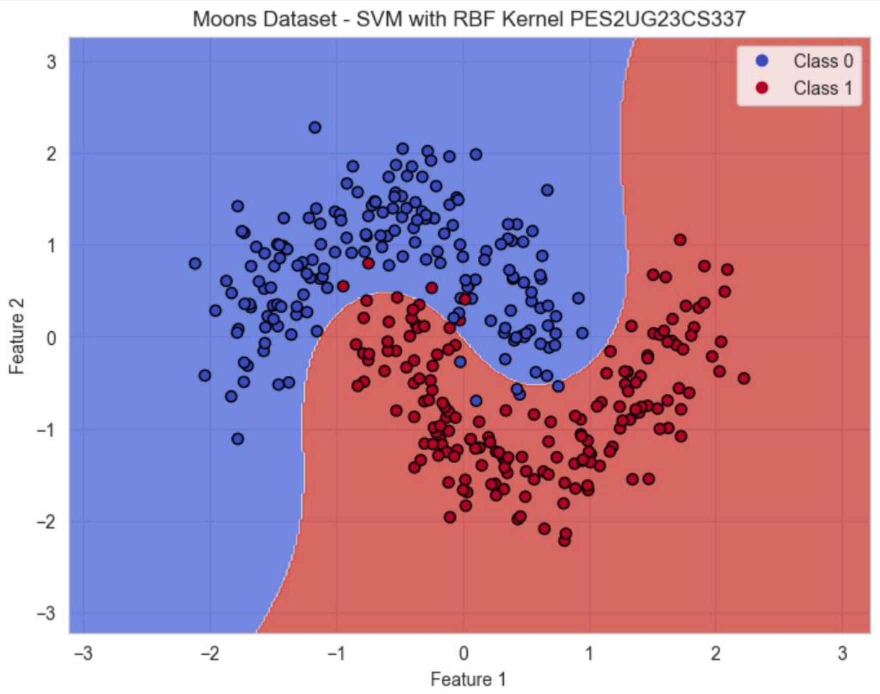
- Soft Margin (C=0.1) is safer for new, unseen data.
- Reason: It generalizes better, ignores small noise/outliers, and reduces risk of overfitting.
- In real-world scenarios, start with a low C value for noisy data.

TRAINING RESULTS

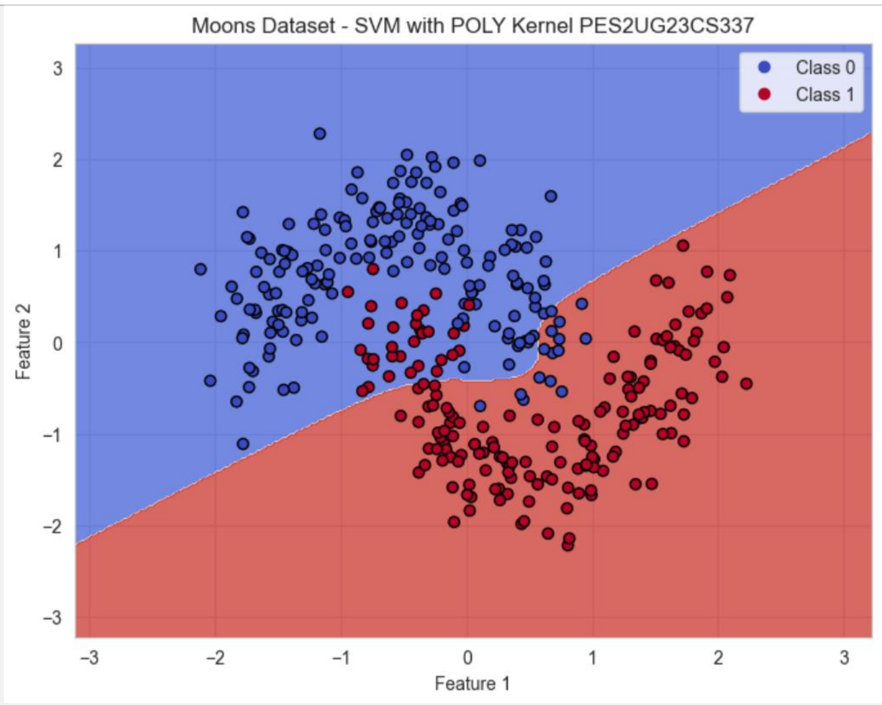
```
SVM with LINEAR Kernel PES2UG23CS337
              precision    recall  f1-score   support

           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

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



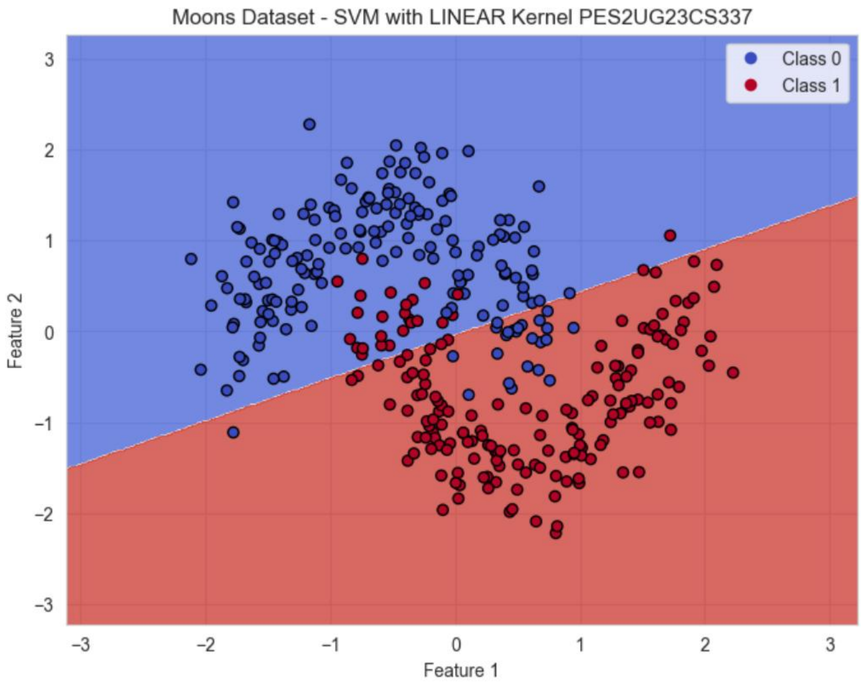
```
SVM with RBF Kernel PES2UG23CS337
              precision    recall  f1-score   support

           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

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

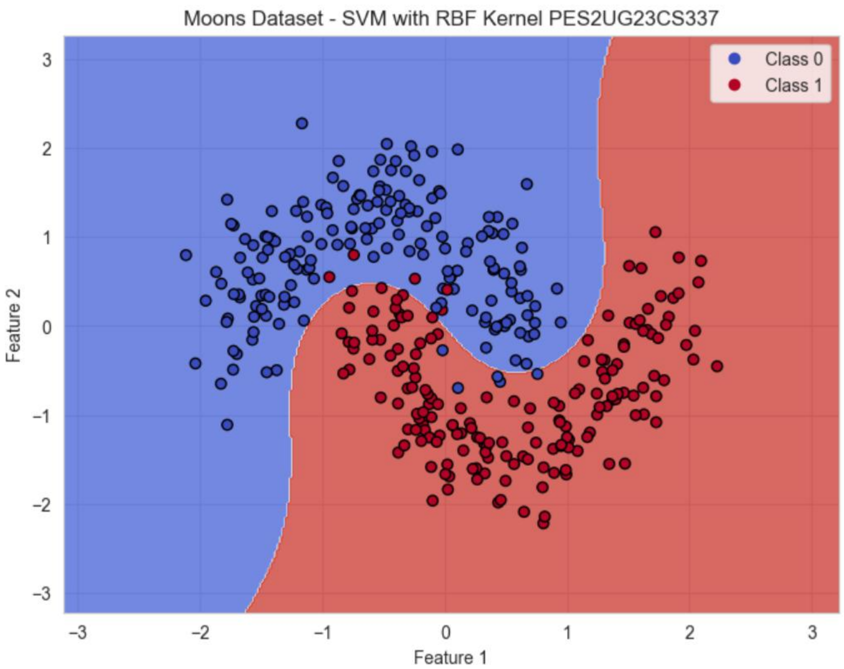


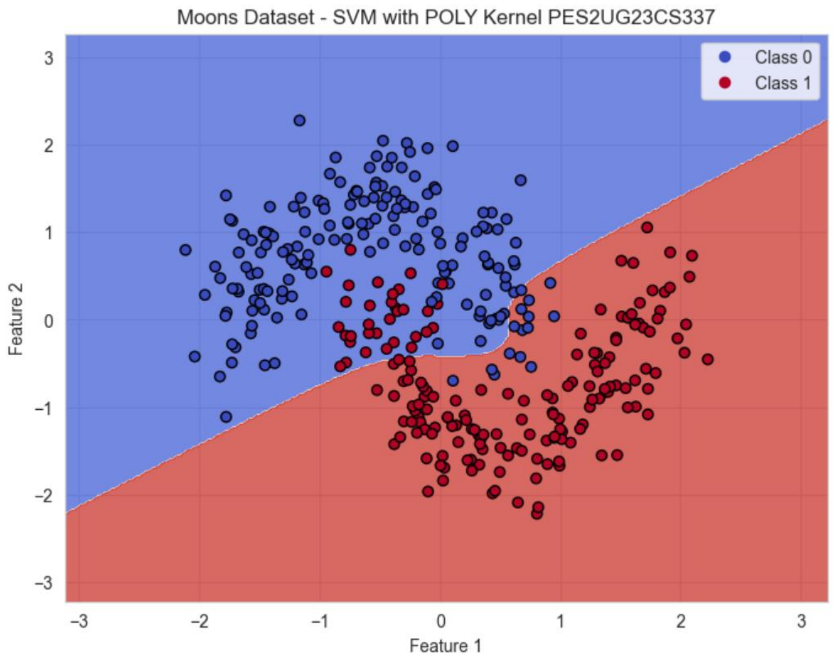

```
SVM with POLY Kernel PES2UG23CS337
              precision    recall  f1-score   support

           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
```

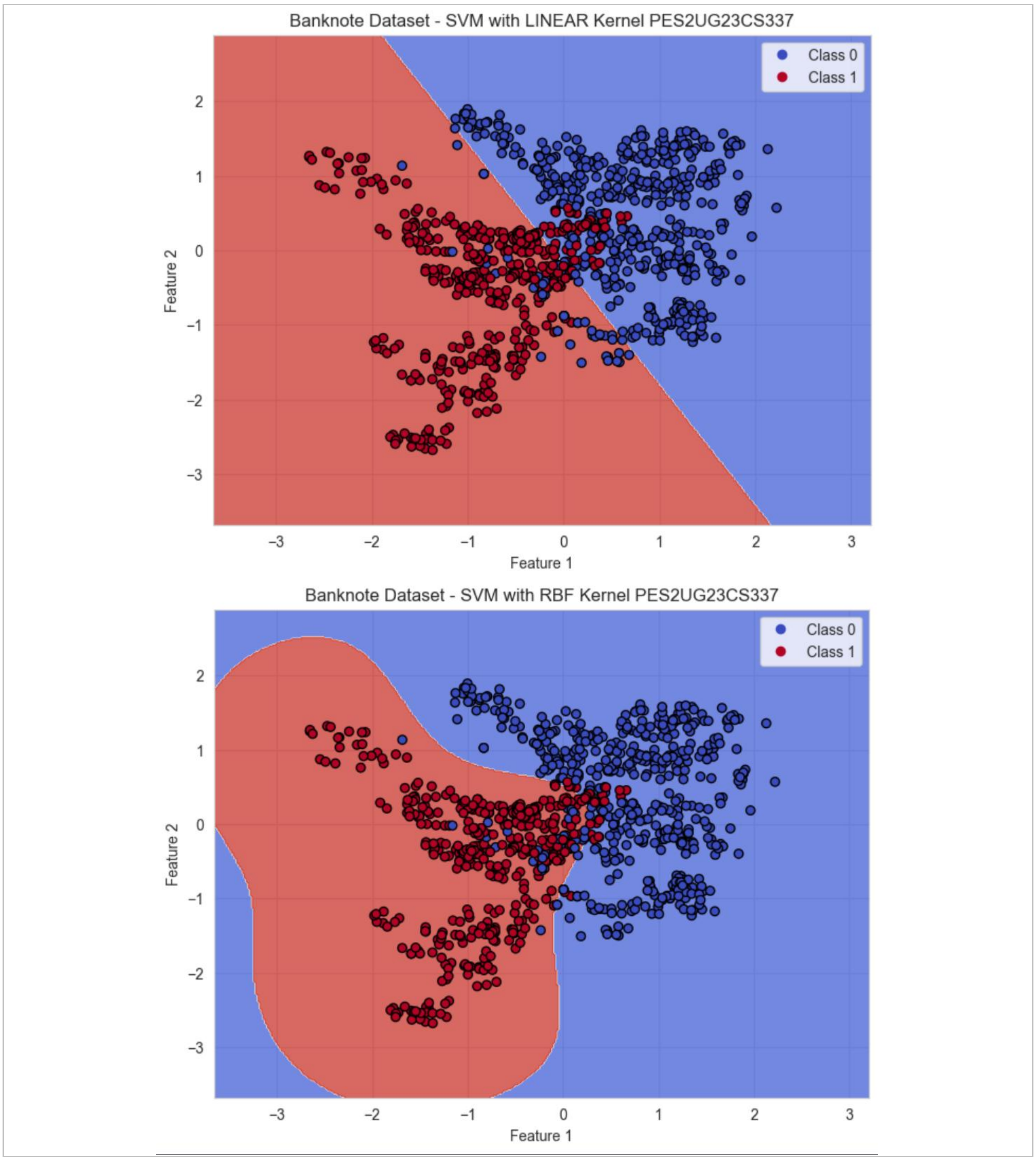


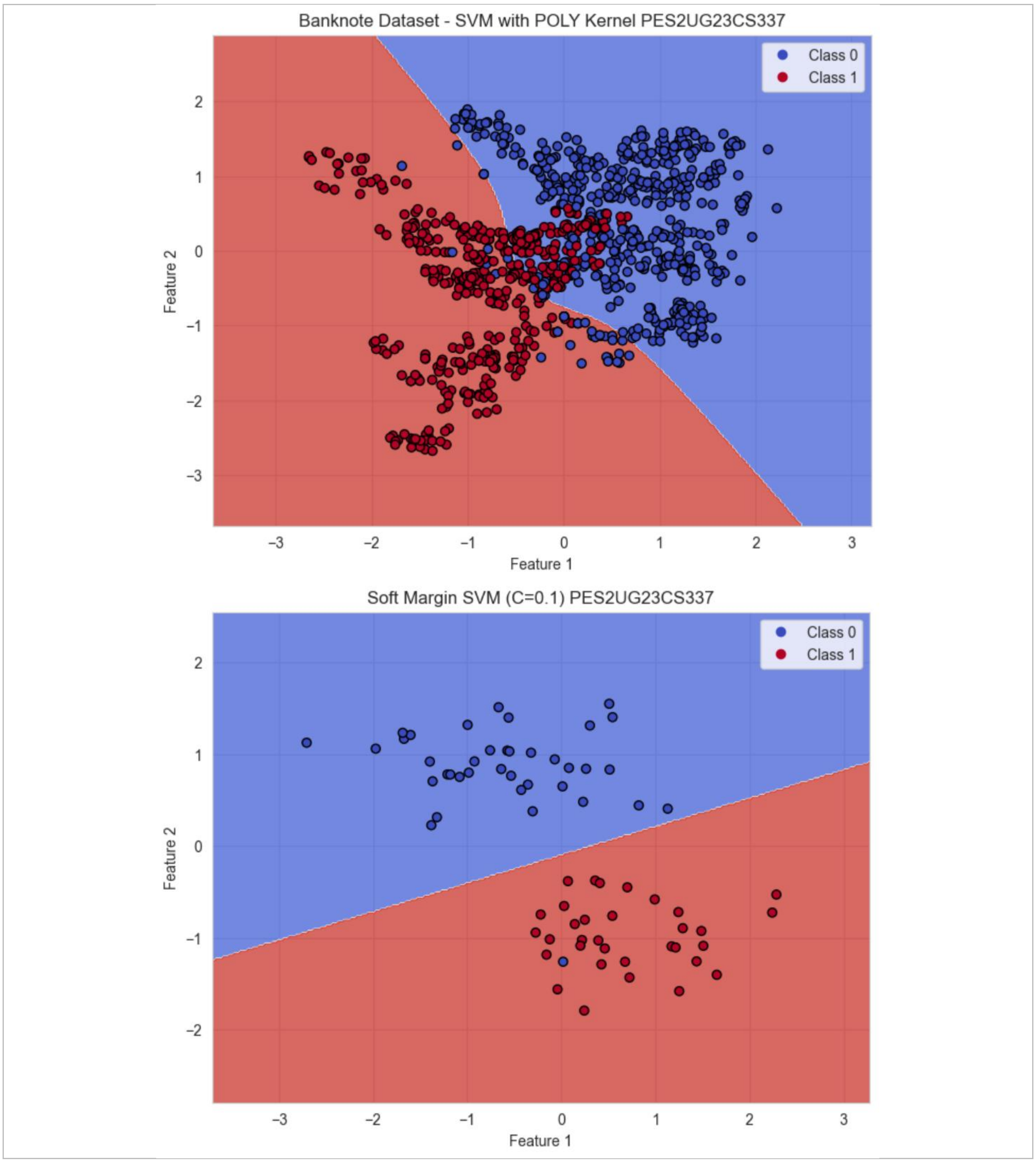
DECISION BOUNDARY VISUALIZATIONS

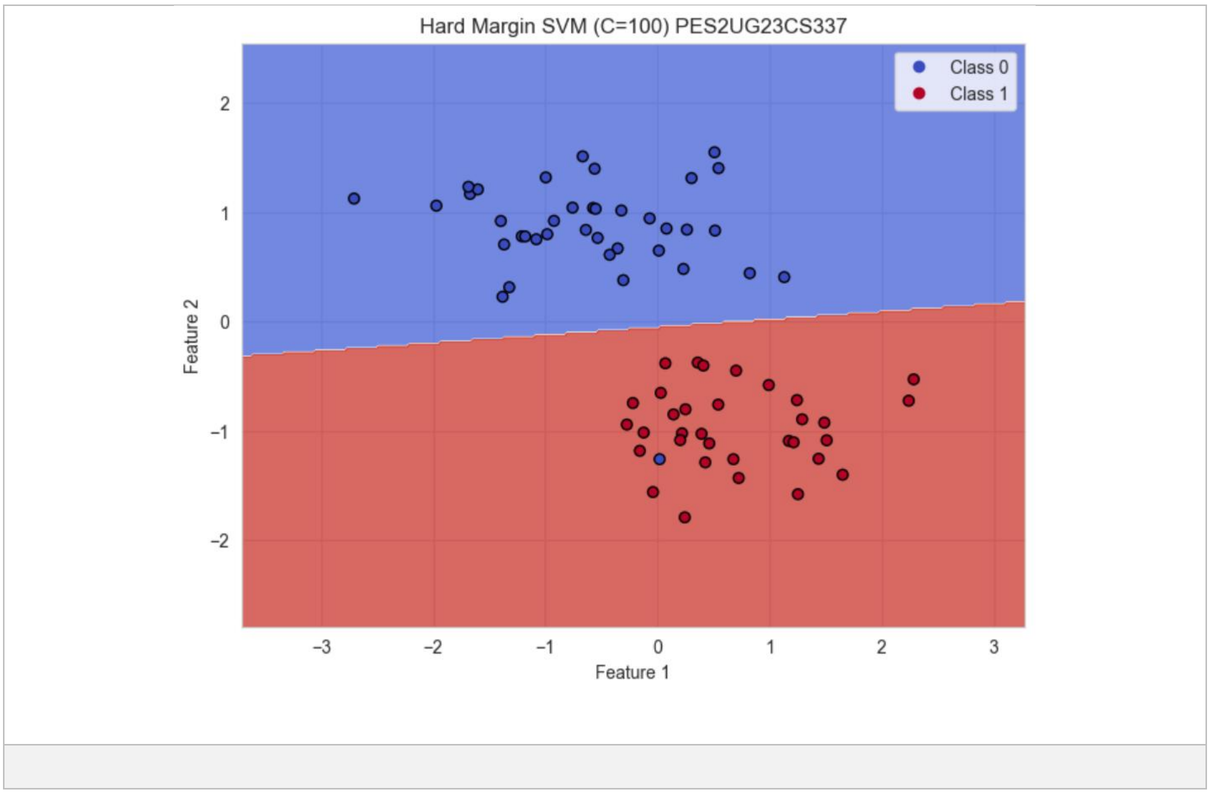



Moons Dataset - SVM with RBF Kernel PES2UG23CS337
Class 0
Class 1
Feature 1
Feature 2


Moons Dataset - SVM with POLY Kernel PES2UG23CS337
Class 0
Class 1
Feature 1
Feature 2


Banknote Dataset - SVM with LINEAR Kernel PES2UG23CS337
Class 0
Class 1
Feature 2
Feature 1
Banknote Dataset - SVM with RBF Kernel PES2UG23CS337
Class 0
Class 1
Feature 2
Feature 1


Banknote Dataset - SVM with POLY Kernel PES2UG23CS337
Class 0
Class 1
Feature 2
Feature 1
Soft Margin SVM (C=0.1) PES2UG23CS337
Class 0
Class 1
Feature 2
Feature 1


Hard Margin SVM (C=100) PES2UG23CS337
Class 0
Class 1
Feature 2
Feature 1
2
1
0
−1
−2
−3
−2
−1
0
1
2
3