UE23CS352A: MACHINE LEARNING
Week 10: SVM Lab

Nidhi K	**PES2UG23CS383**	**Section: F**

Date: 12-10-2025

Analysis:

Moons Dataset Questions (2 questions):

1. Inferences about the Linear Kernel's performance.

The Linear Kernel showed poor performance on the Moons Dataset. This dataset is **non-linearly separable**, meaning a single straight line cannot properly divide the classes. As observed in the decision boundary visualization, the linear kernel tries to find the best straight line to separate the curved data. This results in many misclassifications and consequently low accuracy.

2. Comparison between RBF and Polynomial kernel decision boundaries.

The RBF kernel successfully captures the non-linear, crescent-moon shape of the data, forming a curved decision boundary that separates the classes effectively with high accuracy. The Polynomial kernel also produces a curved boundary but is less precise than the RBF kernel. Its boundary appears more rigid and fails to follow the dataset’s intricate curve as closely, leading to reduced performance.

Banknote Dataset Questions (2 questions):

1. Which kernel was most effective for this dataset?

The RBF (Radial Basis Function) kernel was the most effective for the Banknote Dataset. It achieved nearly perfect accuracy of 1.00, along with precision, recall, and F1-score of 1.00 for both classes. This indicates that the RBF kernel was able to model a complex, non-linear boundary that perfectly separates genuine from fake banknotes.

2. Why might the Polynomial kernel have underperformed here?

The Polynomial kernel likely underperformed because it may not have been able to model a decision boundary complex enough to perfectly separate the data points, or its parameters (such as degree) were not optimally tuned for this dataset. Although the Polynomial kernel can capture non-linear relationships, its performance is sensitive to parameter choices. The RBF kernel, with its ability to create flexible, localized boundaries, was better suited to the underlying structure of the data.

Hard vs. Soft Margin Questions (4 questions):

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

The soft margin is generally wider than the hard margin. The soft margin allows for a few misclassifications to achieve a more generalizable and wider boundary, whereas the hard margin enforces perfect separation with no errors, often resulting in a narrower margin.

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

The soft margin permits some misclassified points to improve generalization. By allowing a few points to cross the boundary, the model can form a wider, more robust margin that is less affected by noise or outliers in the training data, which helps it perform better on unseen data.

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

The hard margin model is more likely to overfit. By insisting on perfect separation of the training data, it becomes overly sensitive to noise and outliers, resulting in a narrow decision boundary that performs well on the training set but poorly on new data.

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

The soft margin model is more trustworthy for new data. Its tolerance for errors and focus on forming a wider, generalizable boundary makes it less prone to overfitting, ensuring more robust performance on unseen examples.

Screenshots:

Moons Dataset (3 screenshots):

Capture the classification report output for

1. Classification Report for SVM with LINEAR Kernel with SRN

```
SVM with LINEAR Kernel <PES2UG23CS383>
              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 for SVM with RBF Kernel with SRN

```
SVM with RBF Kernel <PES2UG23CS383>
              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 for SVM with POLY Kernel with SRN

```
SVM with POLY Kernel <PES2UG23CS383>
              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 Dataset (3 screenshots):

4. Classification Report for SVM with LINEAR Kernel

```
SVM with LINEAR Kernel <PES2UG23CS383>
              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 for SVM with RBF Kernel

```
SVM with RBF Kernel <PES2UG23CS383>
              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
```

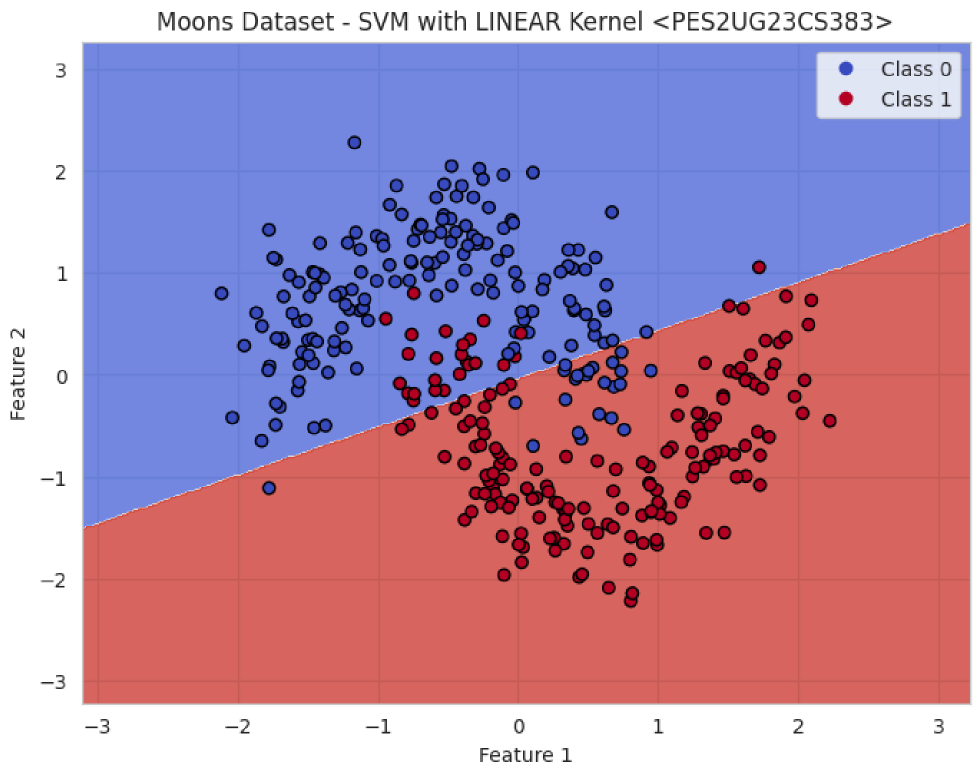
6. Classification Report for SVM with POLY Kernel

```
SVM with POLY Kernel <PES2UG23CS383>
              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
```

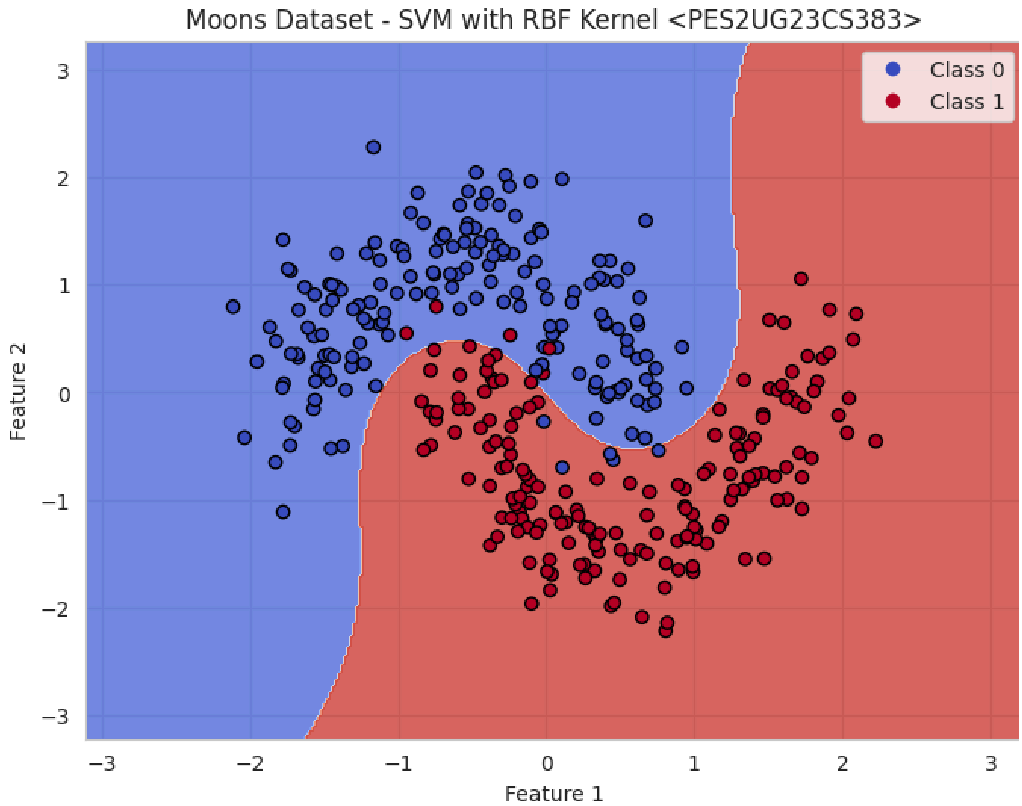
Decision Boundary Visualizations (8 Screenshots):model's decision boundary.

Moons Dataset (3 plots):

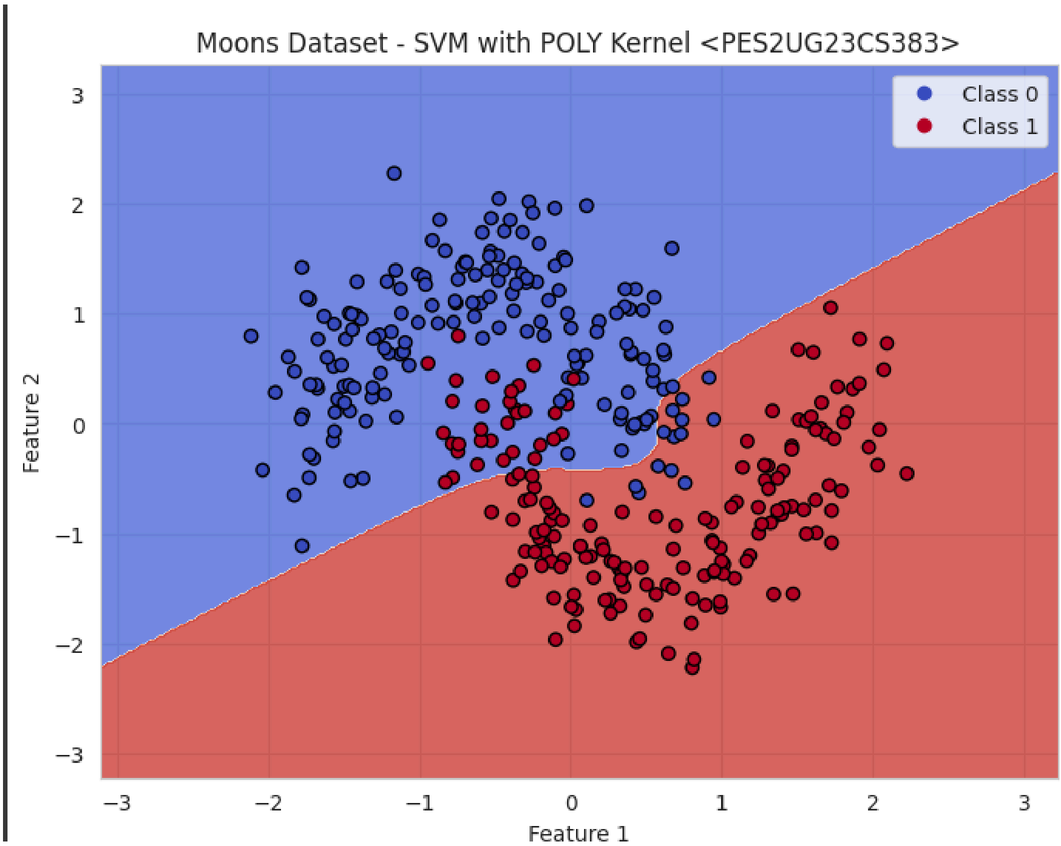
7. Moons Dataset - SVM with LINEAR Kernel



8. Moons Dataset - SVM with RBF Kernel

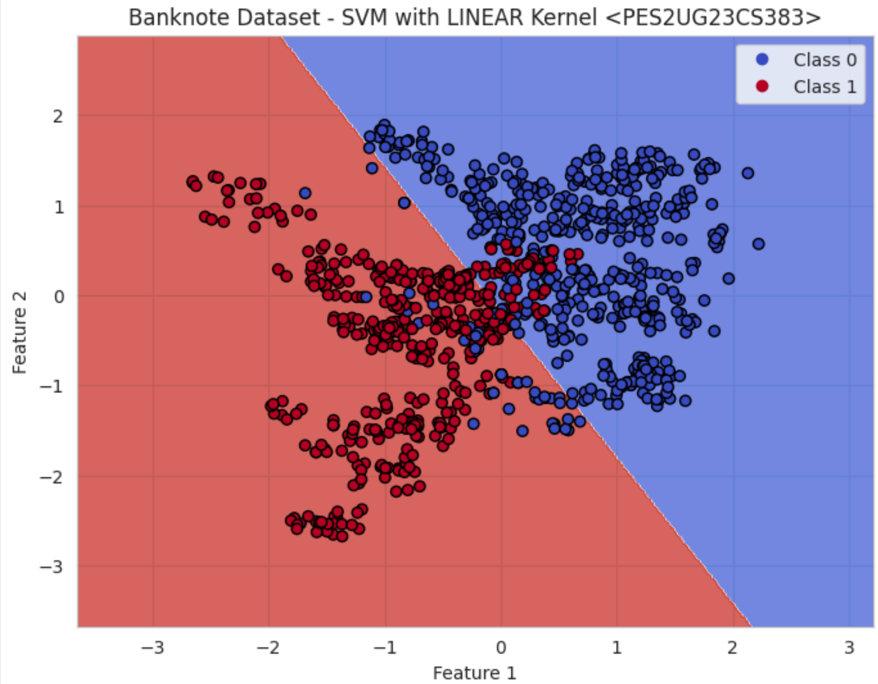


9. Moons Dataset - SVM with POLY Kernel

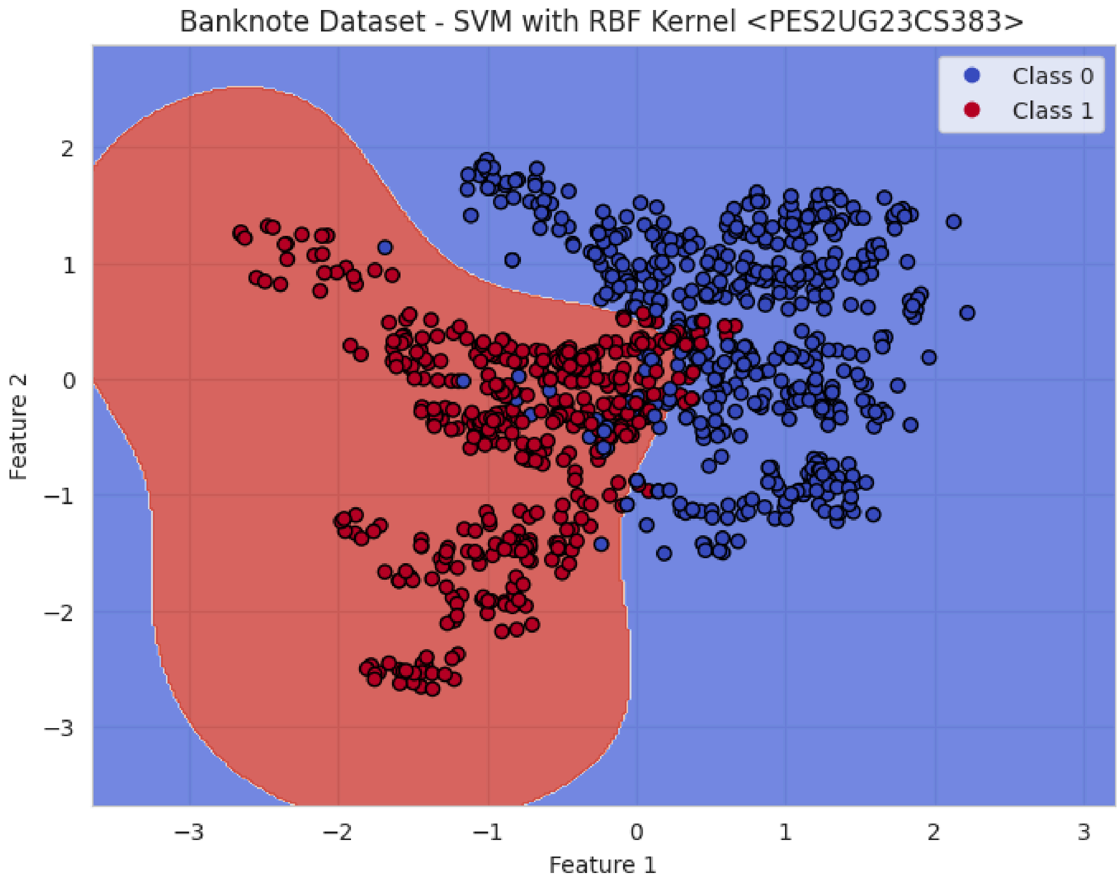


Banknote Dataset (3 plots):

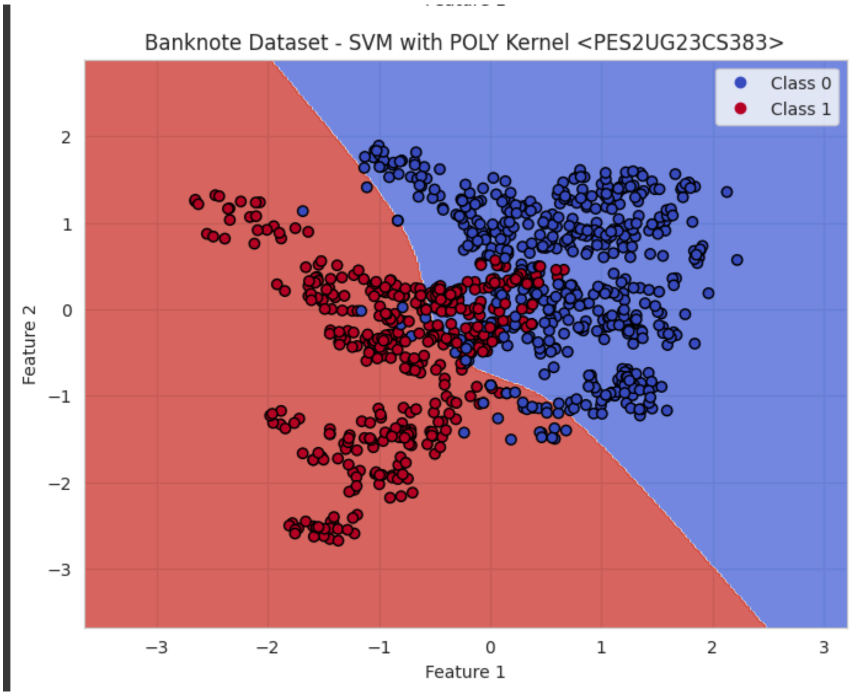
10. Banknote Dataset - SVM with LINEAR Kernel



11. Banknote Dataset - SVM with RBF Kernel

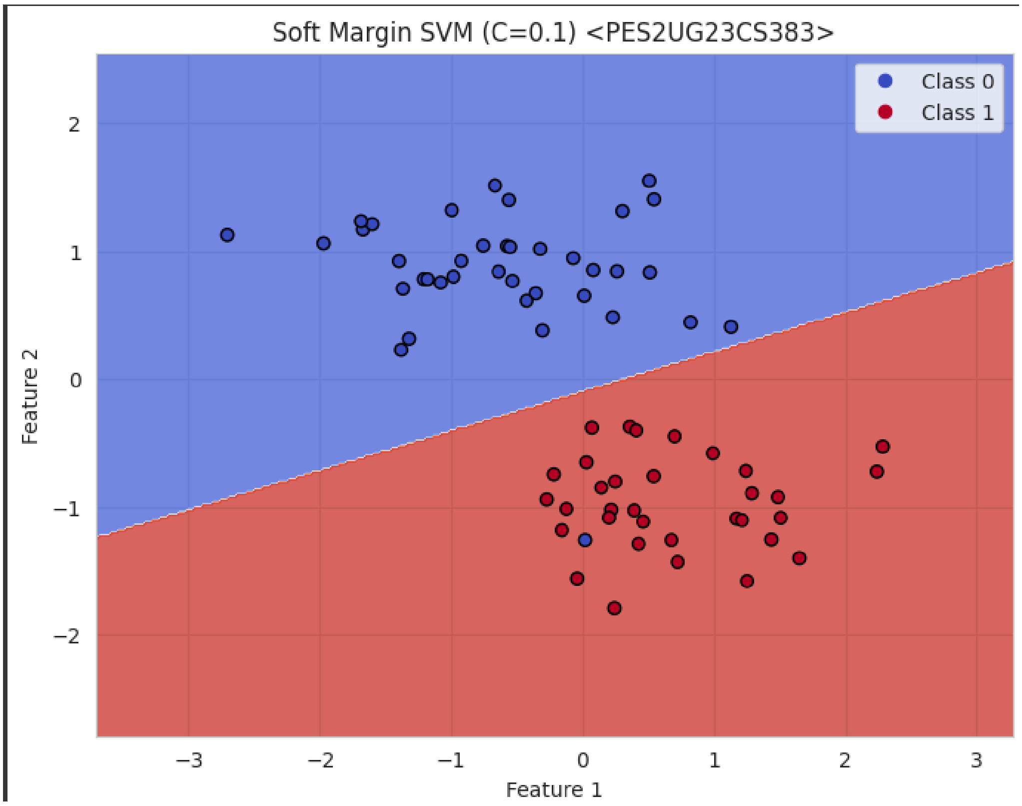


12. Banknote Dataset - SVM with POLY Kernel

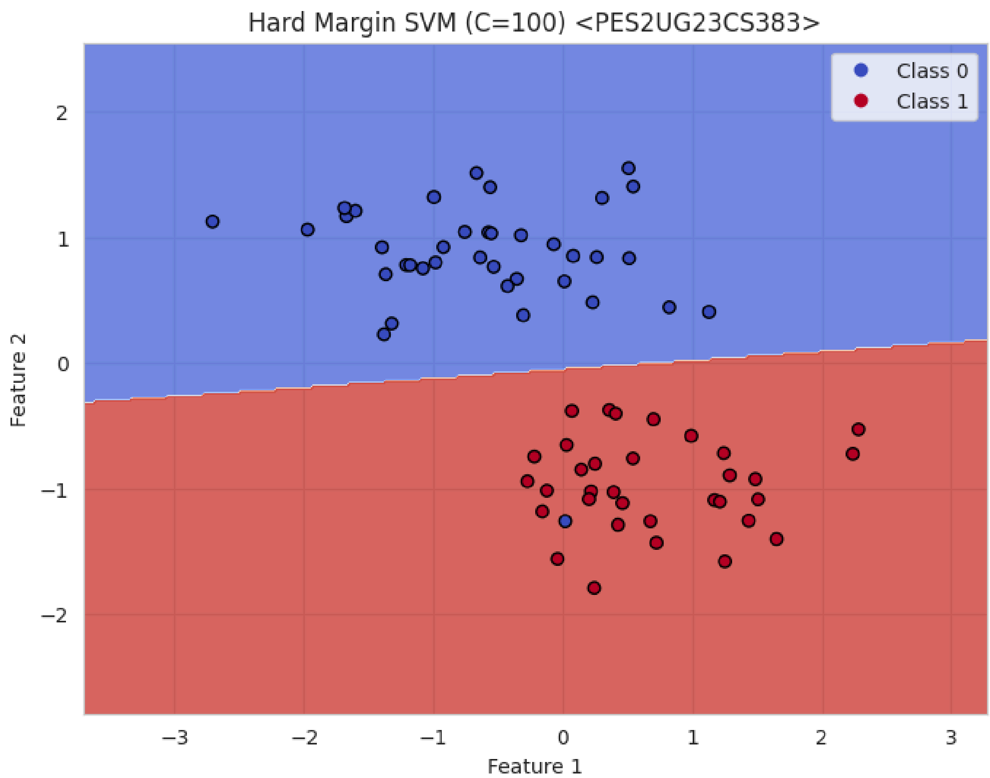


Margin Analysis (2 plots):

13. Soft Margin SVM (C=0.1)



14. Hard Margin SVM (C=100)



Conclusion:

This project shows a clear understanding of the full machine learning workflow, from preparing the data to evaluating the models. Trying out different kernels and comparing hard and soft margins demonstrates a good grasp of important machine learning ideas. Completing all steps successfully, despite initial issues with modules and data, shows a careful and problem-solving approach. The results, especially the strong performance of the RBF kernel on both datasets, suggest that the data contains non-linear patterns.