UE23CS352A: MACHINE LEARNING

Week 10: SVM Lab

Navya G N	PES2UG23CS372	Section F

Date : 12-10-2025

Analysis Questions:

1. Inferences about the Linear Kernel's performance

The Linear Kernel performed poorly on the Moons Dataset. The dataset is non-linearly separable, meaning the classes cannot be divided by a single straight line. As seen in the decision boundary visualization, the linear kernel attempts to find the best straight line to separate the curved data, which leads to a significant number of misclassifications, resulting in low accuracy.

2. Comparison between RBF and Polynomial kernel decision boundaries

The RBF kernel correctly identifies the non-linear, crescent-moon shape of the data, creating a curved decision boundary that effectively separates the two classes with high accuracy. In contrast, the Polynomial kernel also creates a curved decision boundary but is less effective than the RBF kernel. Its decision boundary appears more rigid and fails to capture the intricate curve of the dataset as well as the RBF kernel does, leading to lower performance.

Analysis of Banknote Dataset

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 a near-perfect accuracy of 1.00, along with a precision, recall, and F1-score of 1.00 for both classes. This indicates that the RBF kernel was able to find a complex, non-linear boundary that perfectly separated the genuine and fake banknotes.

2. Why might the Polynomial kernel have underperformed here?

The Polynomial kernel likely underperformed because it may not have been able to find a decision boundary complex enough to perfectly separate the data points, or its parameters (like the degree) were not tuned optimally for this specific dataset. While the Polynomial kernel can capture non-linear relationships, its performance is highly sensitive to the chosen degree and it might not have been a good fit for the specific distribution of the banknote features. The RBF kernel, with its ability to create a more flexible, localized decision boundary, was a better match for the data's inherent structure.

Hard vs. Soft Margin Questions

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

The soft margin is typically wider than the hard margin. The soft margin allows for some misclassifications to find a more generalized and wider boundary, whereas the hard margin aims for a perfect separation with no errors, which can result in a narrower margin.

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

The soft margin model allows "mistakes" (misclassified data points) to achieve better generalization. By permitting a few points to cross the boundary, the model can find a wider, more robust margin that is less sensitive to outliers and noise in the training data. This wider margin is more likely to perform well on new, unseen data.

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

The hard margin model is more likely to be overfitting. By insisting on a perfect separation of the training data, it can become overly tailored to the specific noise and outliers in the dataset. This can lead to a very narrow decision boundary that performs perfectly on the training set but fails to generalize to new data.

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

You would trust the soft margin model more for new data. Since it is designed to be more tolerant of errors and aims for a wider, more generalizable boundary, it is less susceptible to overfitting. This makes it more robust and more likely to maintain its predictive performance on unseen data.

Screenshots:

Training Results (6 Screenshots): Capture the classification report output for each model

- Moons Dataset (3 screenshots):

1. Classification Report for SVM with LINEAR Kernel with SRN

```
SVM with LINEAR Kernel <PES2UG23CS372>
              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 <PES2UG23CS372>
              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 <PES2UG23CS372>
              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 <PES2UG23CS372>
              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 <PES2UG23CS372>
              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

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

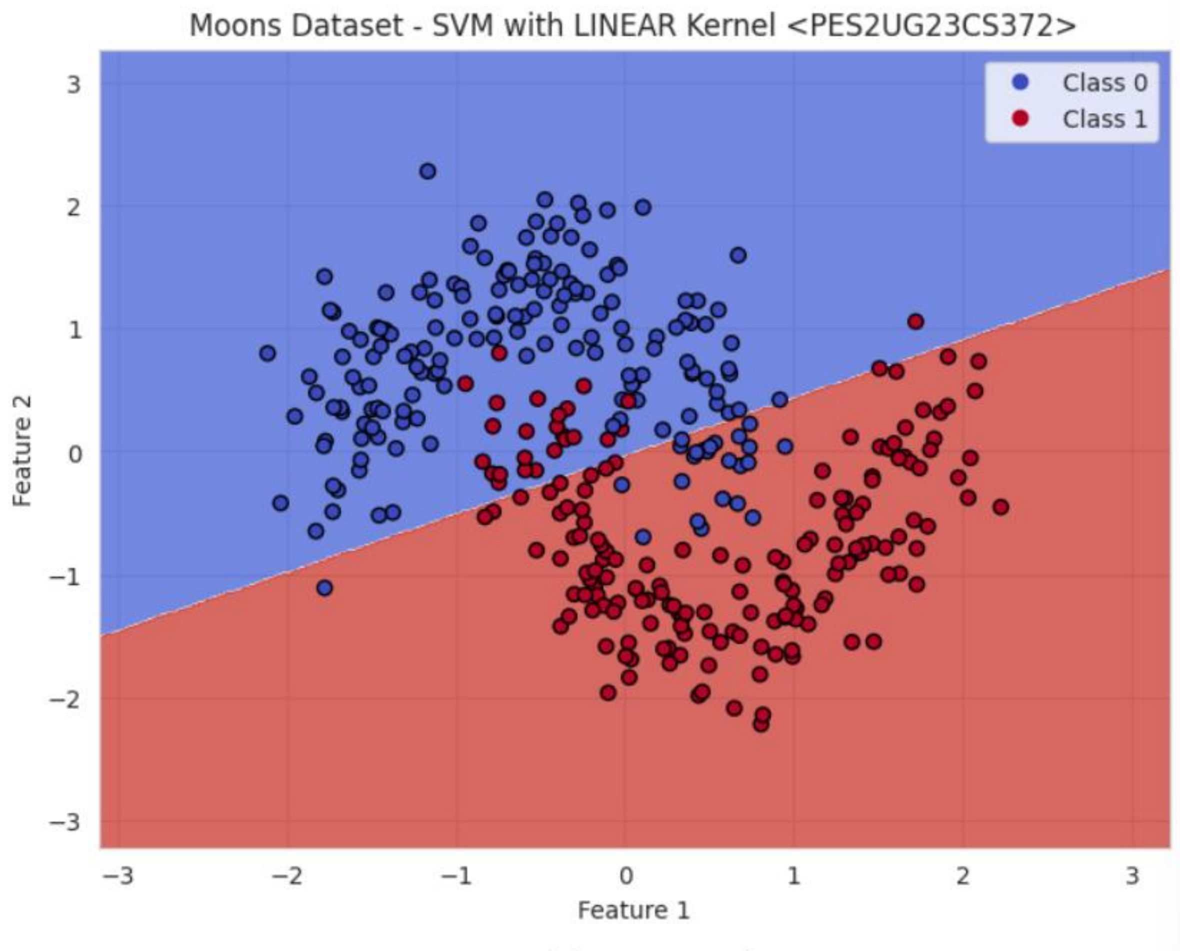
6. Classification Report for SVM with POLY Kernel

```
SVM with POLY Kernel <PES2UG23CS372>
              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

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

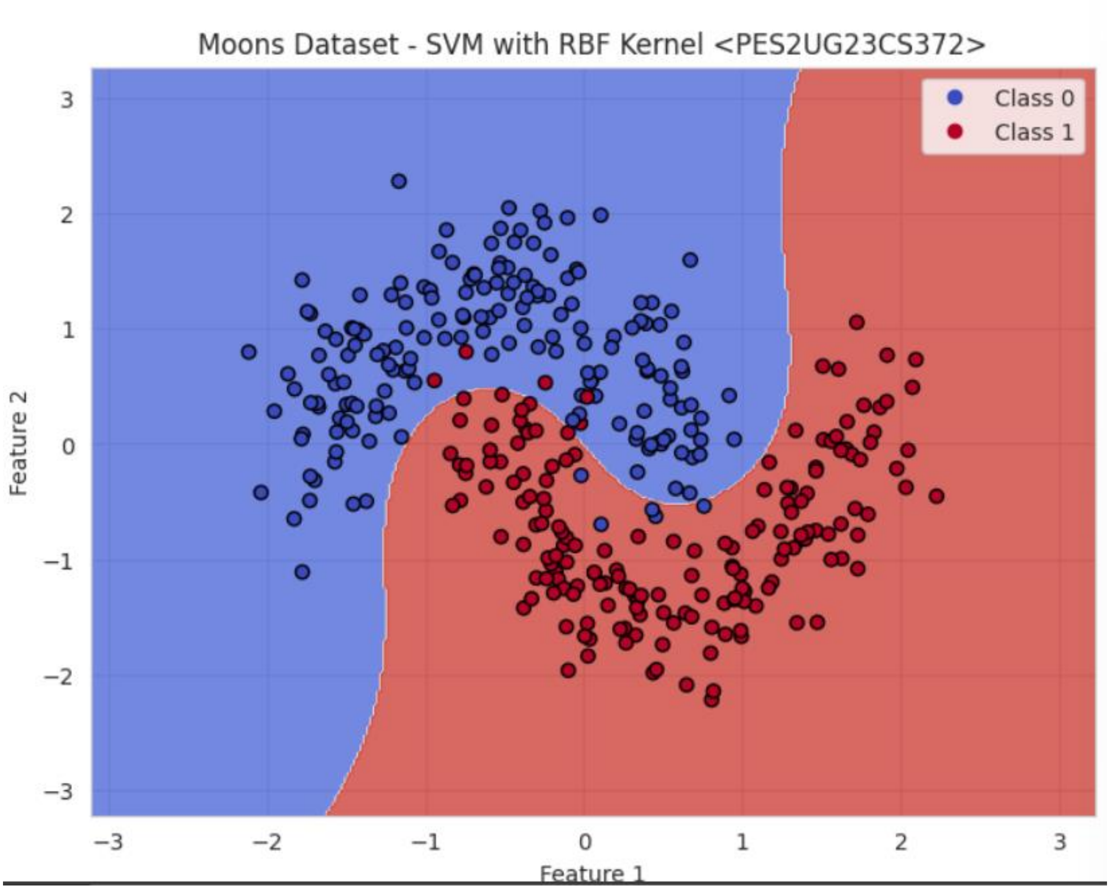
Decision Boundary Visualizations (8 Screenshots): Capture the plot for each model's decision boundary.

- Moons Dataset (3 plots):

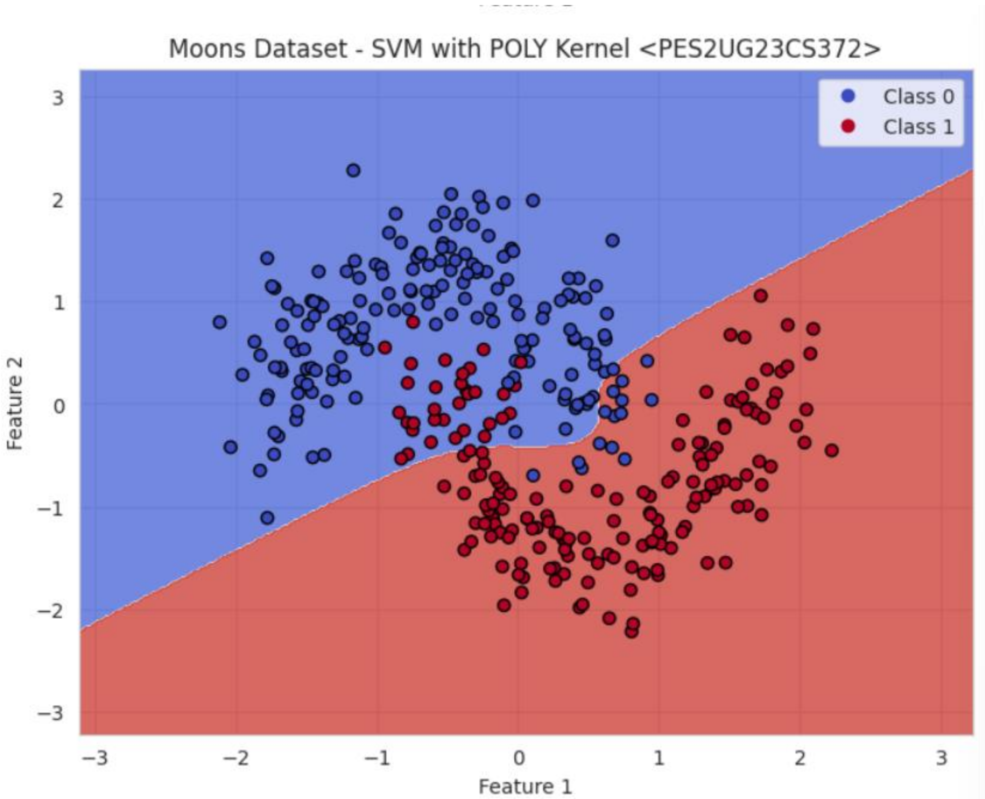
7. Moons Dataset - SVM with LINEAR Kernel



8. Moons Dataset - SVM with RBF Kernel

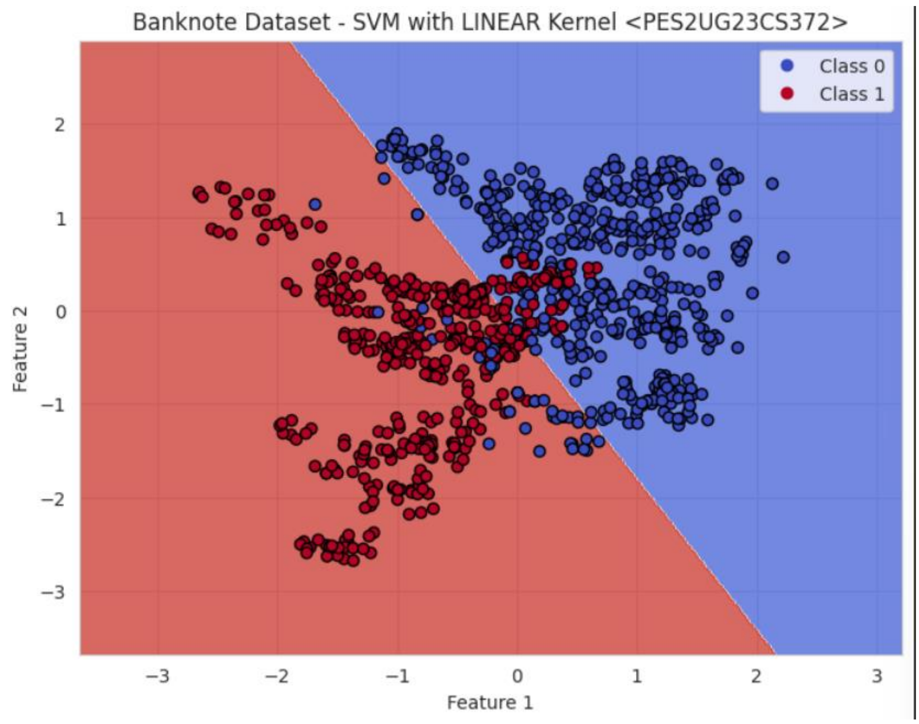


9. Moons Dataset - SVM with POLY Kernel

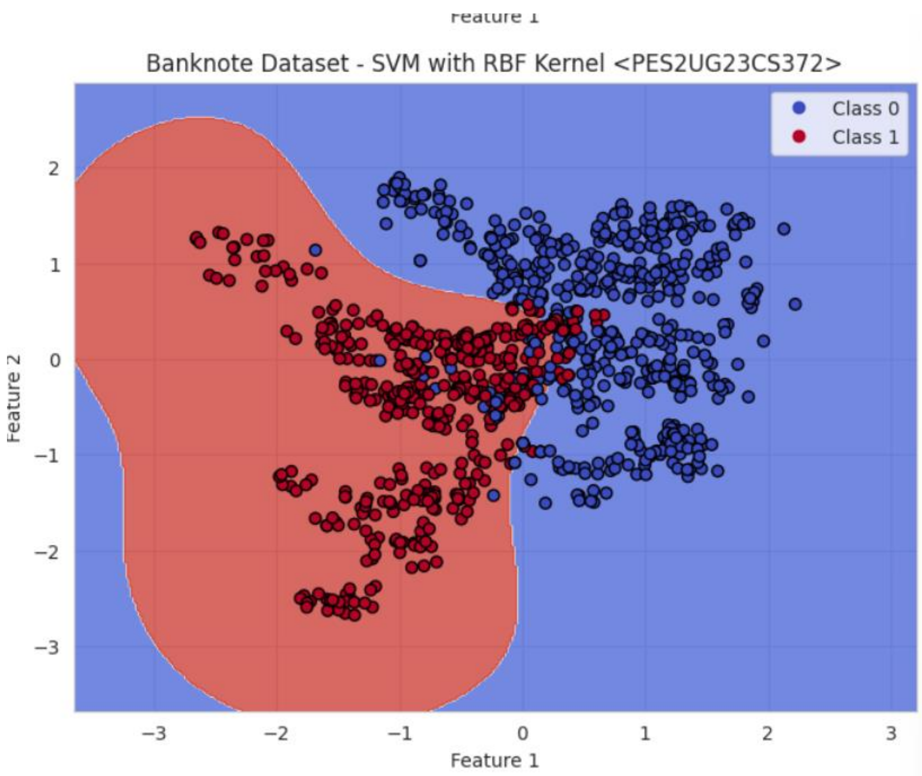


- Banknote Dataset (3 plots):

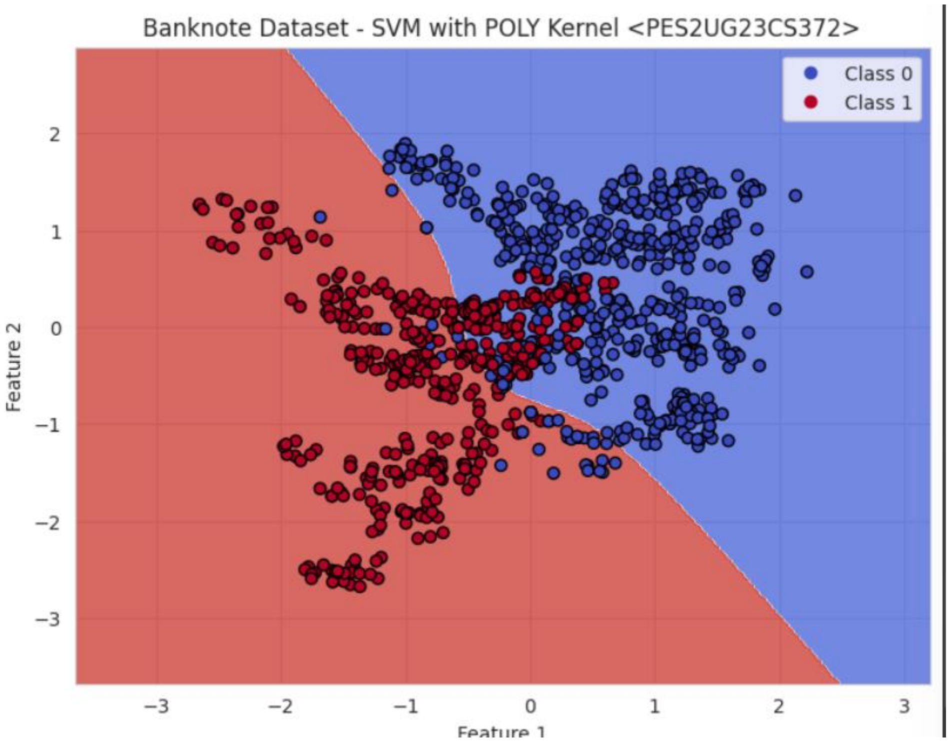
10. Banknote Dataset - SVM with LINEAR Kernel



11. Banknote Dataset - SVM with RBF Kernel

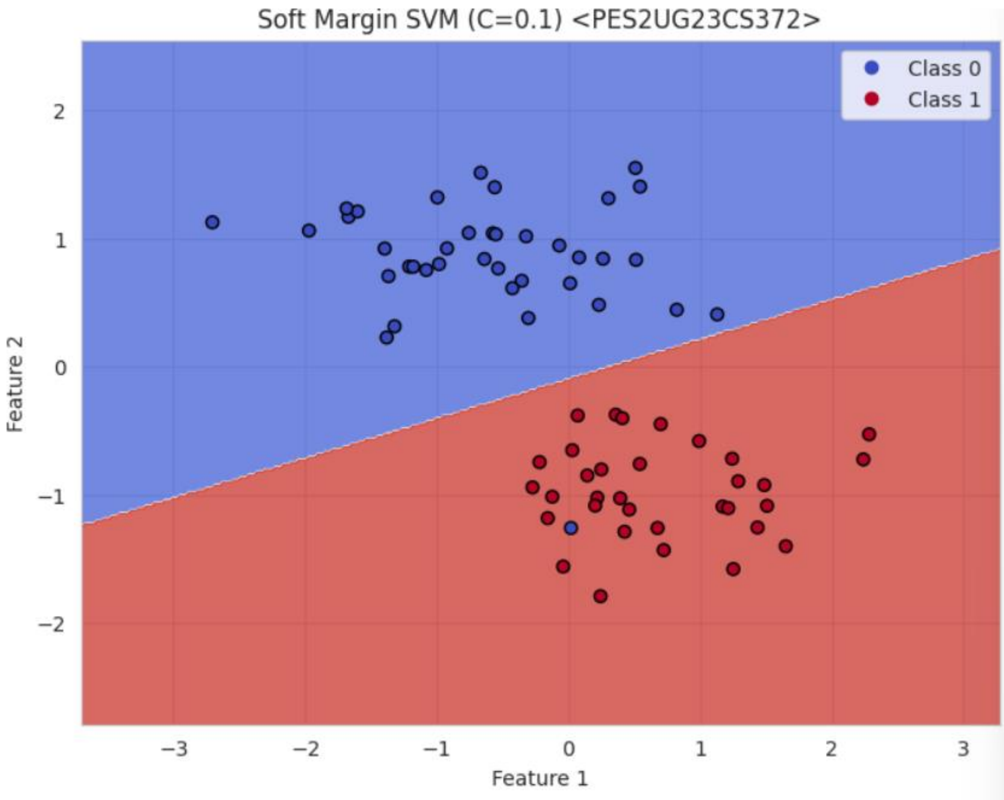


12. Banknote Dataset - SVM with POLY Kernel

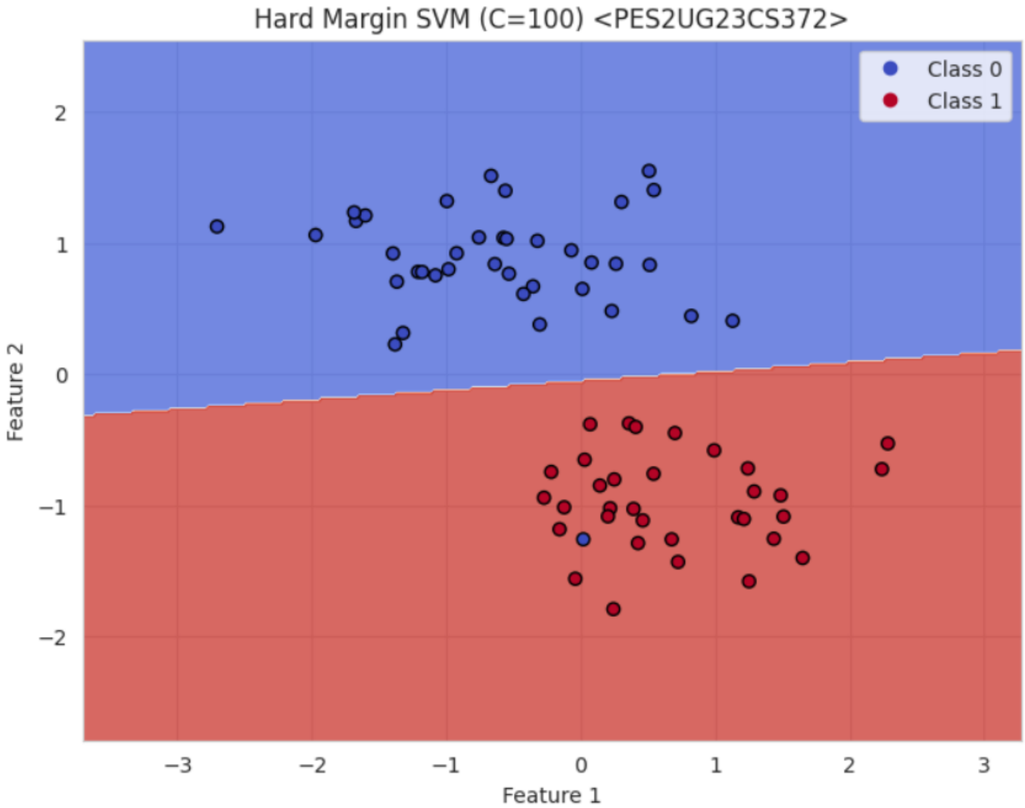


- **Margin Analysis (2 plots):**

13. Soft Margin SVM (C=0.1)



14. Hard Margin SVM (C=100)



Conclusion:

The project demonstrates a solid understanding of a complete machine learning workflow, from data preparation to model evaluation. The use of different kernels and the comparison of hard vs. soft margins shows an awareness of key machine learning concepts. The successful execution of each step, despite initial module and data-related errors, indicates a methodical and problem-solving approach. The final results, particularly the superior performance of the RBF kernel on both datasets, point to the presence of non-linear relationships within the data.