ML Lab Week 10 SVM Lab Instructions

DETAILS:
NAME: PREKSHA KAMALESH
SRN:PES2UG23CS902
SECTION:F

- **Moons Dataset Questions (2 questions):**
    1. Inferences aboutthe Linear Kernel's performance.

       Based on the classification report, the Linear Kernel achieved an accuracy of 87% on the Moons dataset. The precision, recall, and f1-score are also around 87% for both classes. The visualization shows a straight-line decision boundary, which is not ideal for separating the non-linearly separable moon shapes. This indicates that a linear model is not well-suited for this dataset, and a more complex kernel is needed to capture the data's structure effectively.
    2. Comparison between RBF and Polynomial kernel decision boundaries.

       Comparing the visualizations, the RBF kernel's decision boundary is curved and closely follows the shape of the two moons, effectively separating the two classes. The Polynomial kernel also creates a curved boundary, but it appears less smooth and doesn't conform to the moon shapes as naturally as the RBF kernel. Therefore, the RBF kernel seems to capture the shape of the data more naturally and provides a better separation for this non-linear dataset.

- **Banknote Dataset Questions (2 questions):**
    1. Which kernel was most effective for this dataset?

       Based on the classification reports for the Banknote dataset, the RBF kernel appears to be the most effective. It achieved the highest accuracy (93%) and generally higher precision, recall, and f1-scores for both classes compared to the Linear and Polynomial kernels. The visualization also shows a clear separation of the two classes with the RBF kernel.
    2. Why might the Polynomial kernel have underperformed here?

       The lower performance of the Polynomial kernel on the Banknote dataset compared to the Moons dataset could be due to several factors. The Banknote dataset, while having some non-linear separation, might not have a structure that is well-captured by a simple polynomial transformation in the 2D space of the selected features ('variance' and 'skewness'). The Moons dataset, on the other hand, has a more distinct curved structure that the polynomial kernel can better approximate. Additionally, the optimal degree for the polynomial kernel can vary between datasets, and the default degree might not be suitable for the Banknote dataset.

- **Hard vs. Soft Margin Questions (4 questions):**
    1. Which margin (soft or hard) is wider?

       The "Soft Margin" model with C=0.1 produces a wider margin. The decision boundary is less influenced by individual data points, resulting in a larger separation between the classes.
    2. Why does the soft margin model allow "mistakes"?

       The Soft Margin SVM allows these "mistakes" (points inside the margin or misclassified) because its primary goal is to find a balance between maximizing the margin width and minimizing the classification errors. The C parameter controls this trade-off. A smaller C value (like 0.1) makes the model more tolerant of errors, prioritizing a wider margin for better generalization.

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

   The "Hard Margin" model with C=100 is more likely to be overfitting to the training data. With a large C value, the model heavily penalizes misclassifications and tries to fit every training point perfectly, even outliers. This results in a narrower margin that is highly sensitive to the training data and may not generalize well to unseen data.

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

   I would trust the "Soft Margin" model (C=0.1) more to classify a new, unseen data point correctly. The wider margin and tolerance for some errors make it more robust to noise and better at generalizing to new data. In a real-world scenario with noisy data, I would generally prefer to start with a low value of C to create a softer margin and avoid overfitting.

Screenshots Provide clearly labeled screenshots for all the results generated by your notebook. You must include a total of 14 screenshots, divided as follows:

- **Training Results (6 Screenshots):** Capture the classification report output for each model.
  - **Moons Dataset (3 screenshots):**

3. Classification Report for SVM with LINEAR Kernel with SRN

```
SVM with LINEAR Kernel <PES2UG23CS902>
              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

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

4. Classification Report for SVM with RBF Kernel with SRN

```
SVM with RBF Kernel <PES2UG23CS902>
              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

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

5. Classification Report for SVM with POLY Kernel with SRN

```
SVM with POLY Kernel <PES2UG23CS902>
              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):**

6. Classification Report for SVM with LINEAR Kernel

```
SVM with LINEAR Kernel <PES2UG23CS902>
              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

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

7. Classification Report for SVM with RBF Kernel

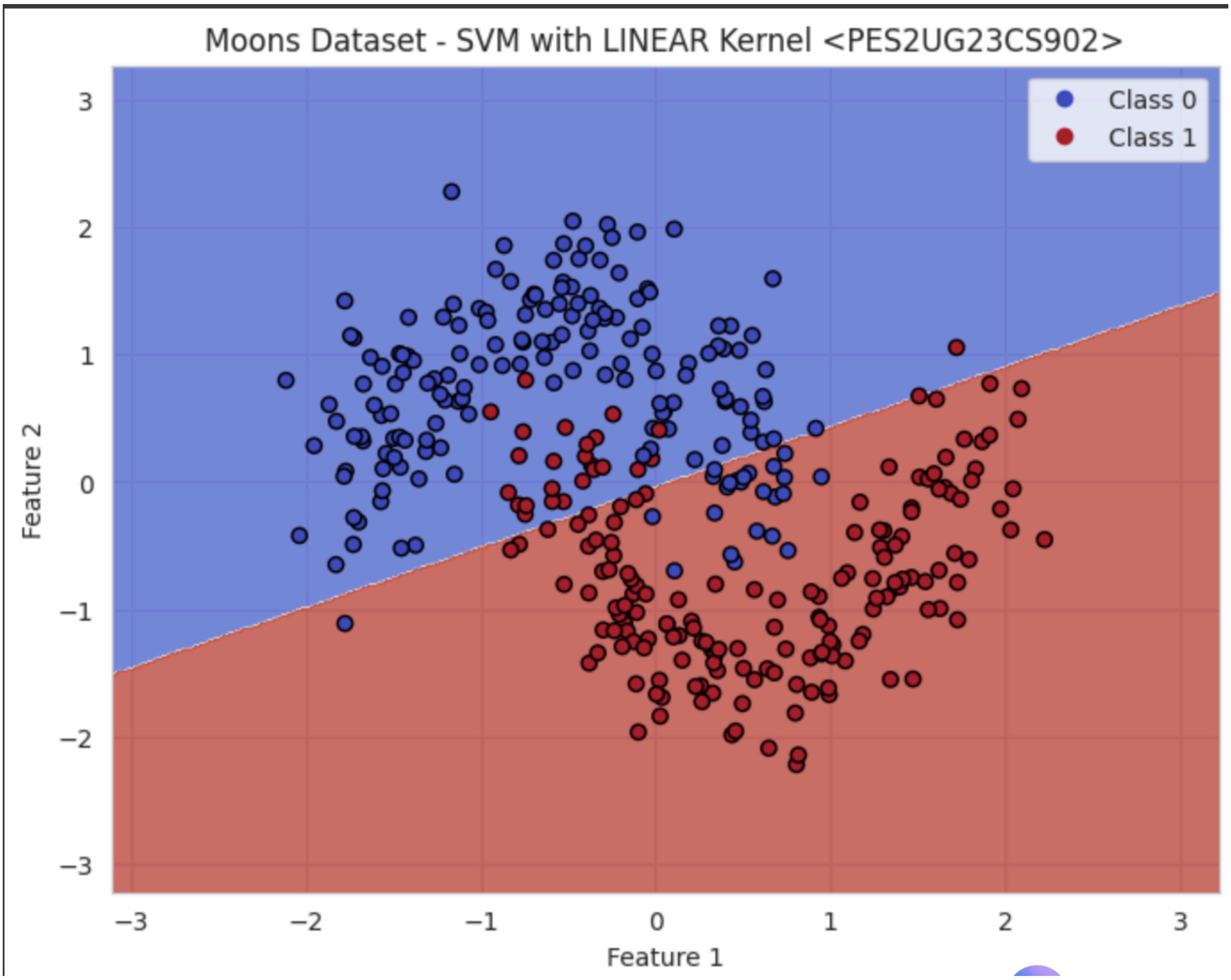
```
SVM with RBF Kernel <PES2UG23CS902>
              precision    recall  f1-score   support

      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

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

8. Classification Report for SVM with POLY Kernel

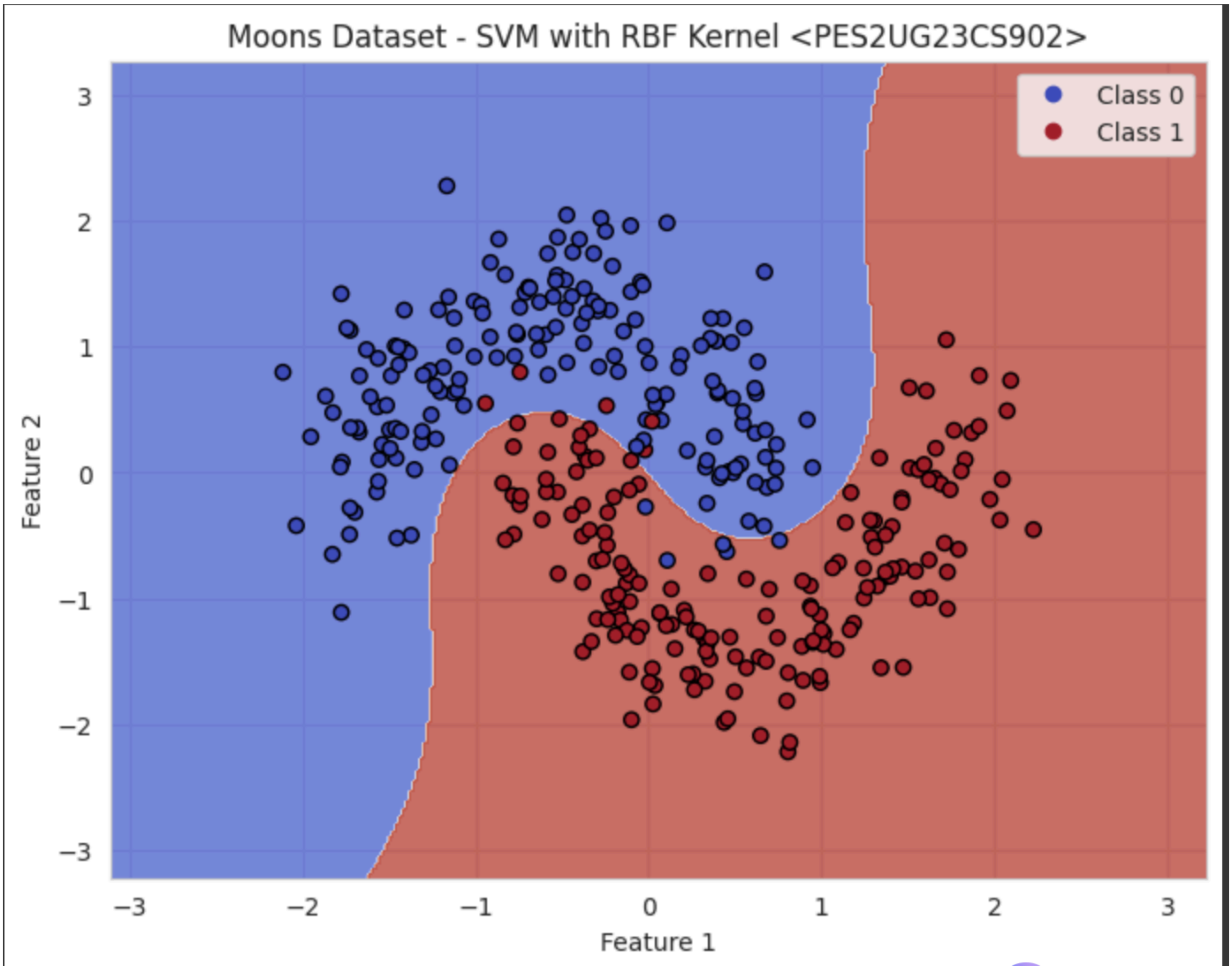
```
SVM with POLY Kernel <PES2UG23CS902>
              precision    recall  f1-score   support

      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):**

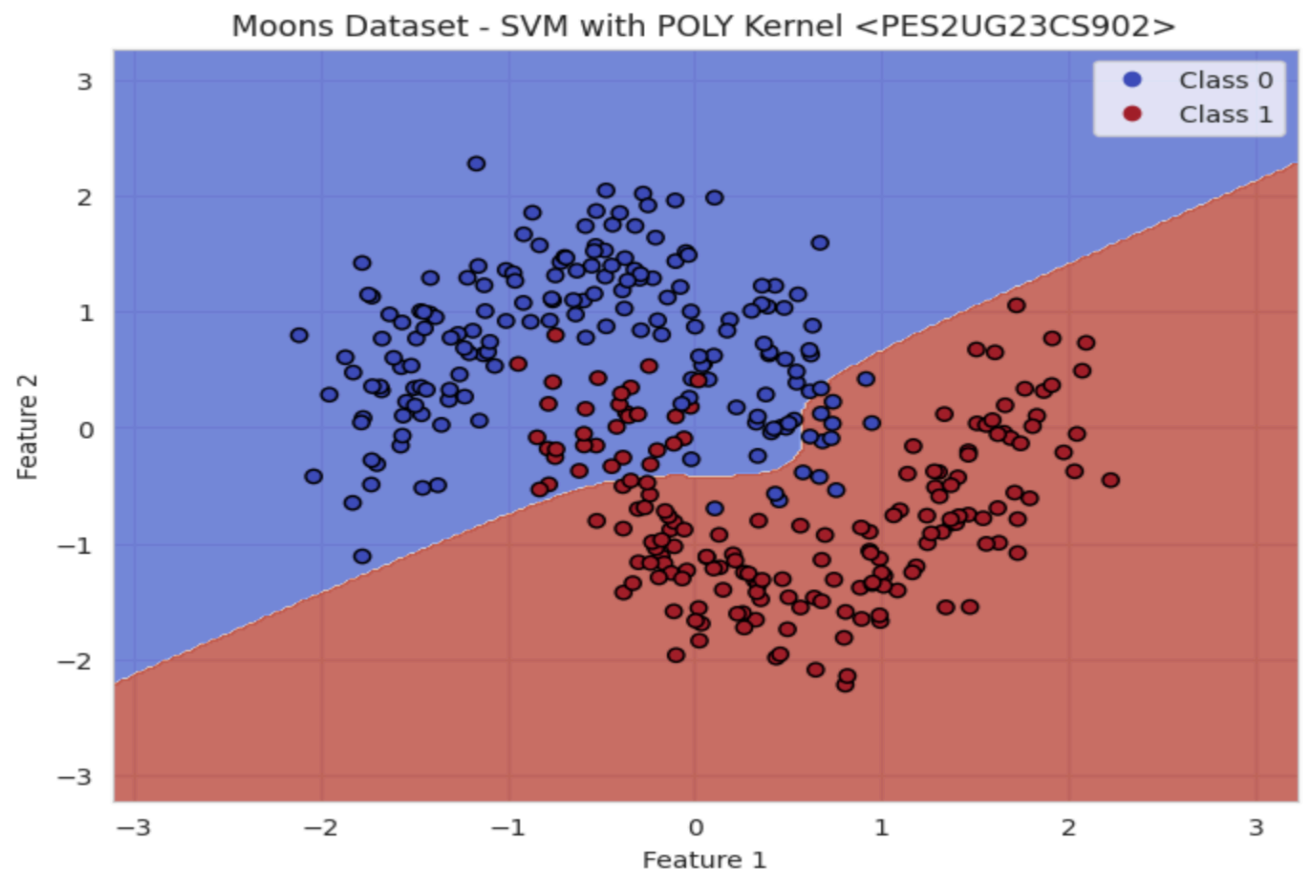
9. Moons Dataset - SVM with LINEAR Kernel



10. Moons Dataset - SVM with RBF Kernel

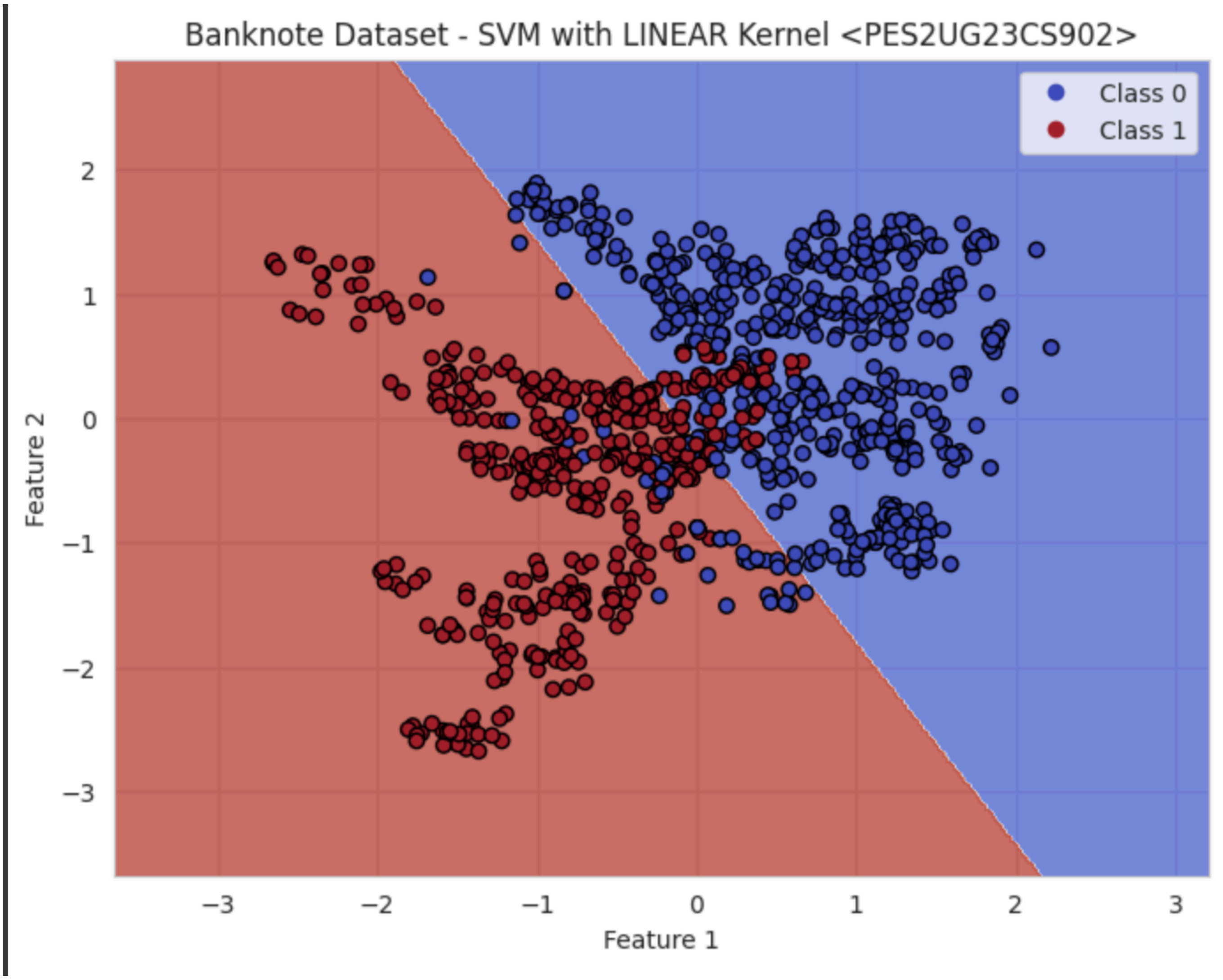


11. Moons Dataset - SVM with POLY Kernel

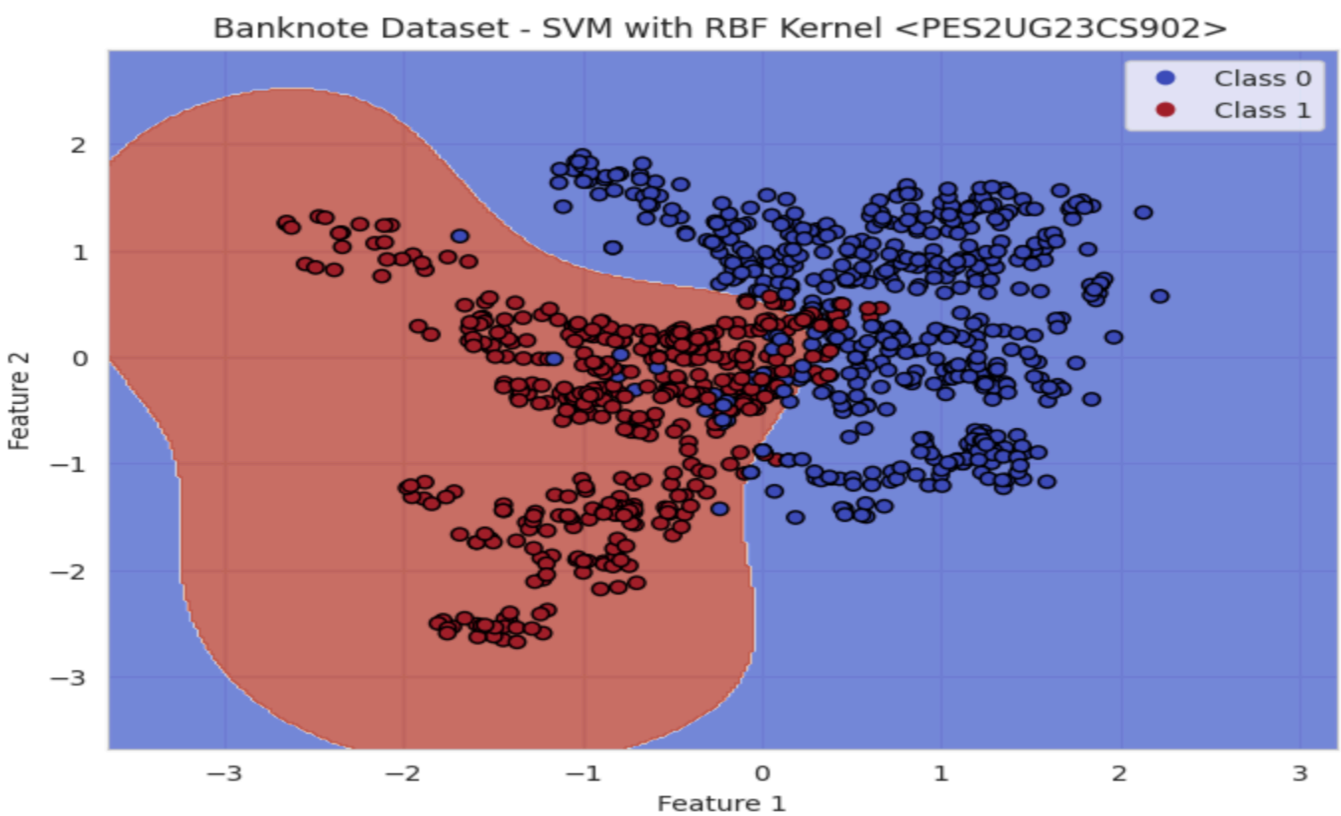


- **Banknote Dataset (3 plots):**

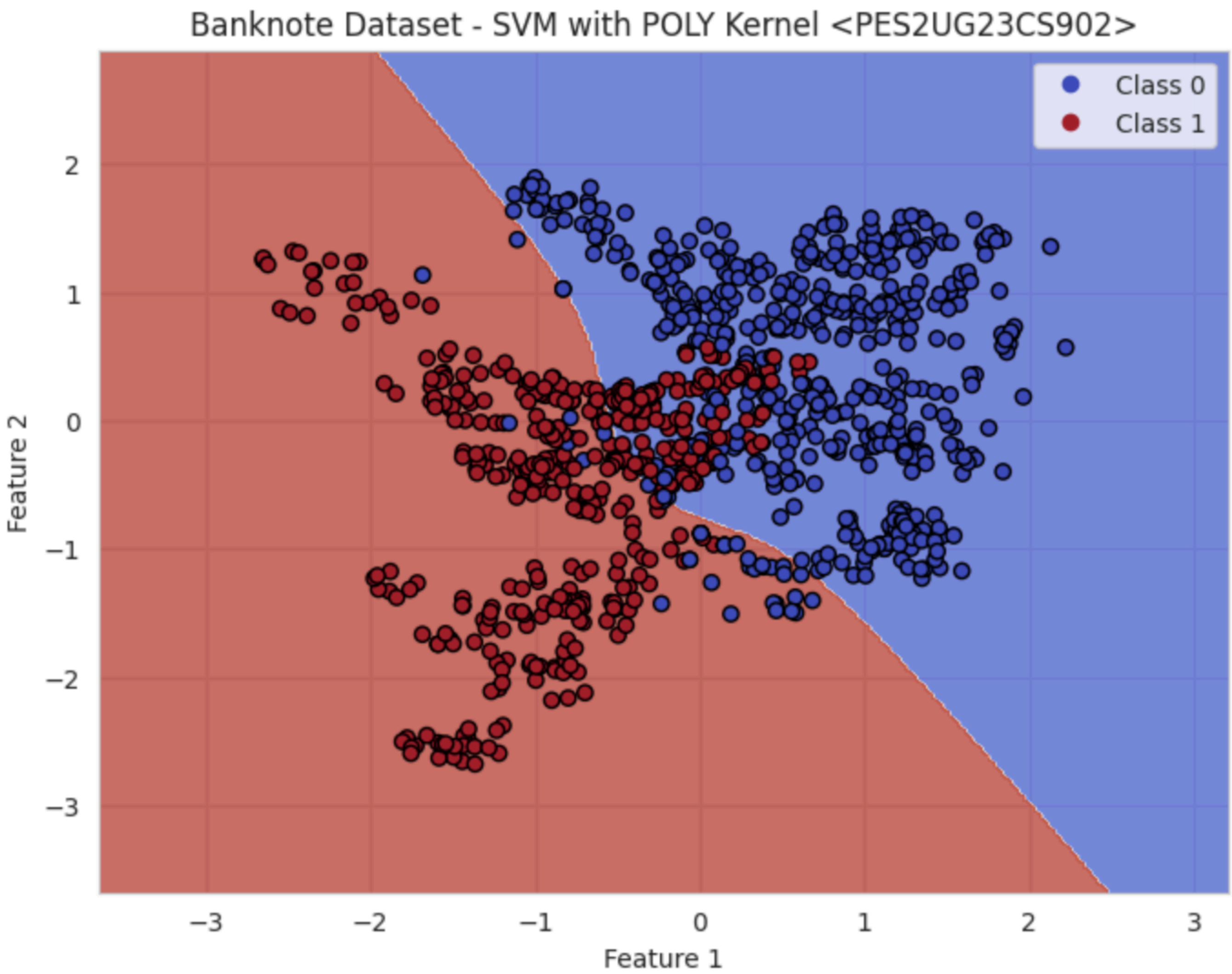
12. Banknote Dataset - SVM with LINEAR Kernel



13. Banknote Dataset - SVM with RBF Kernel

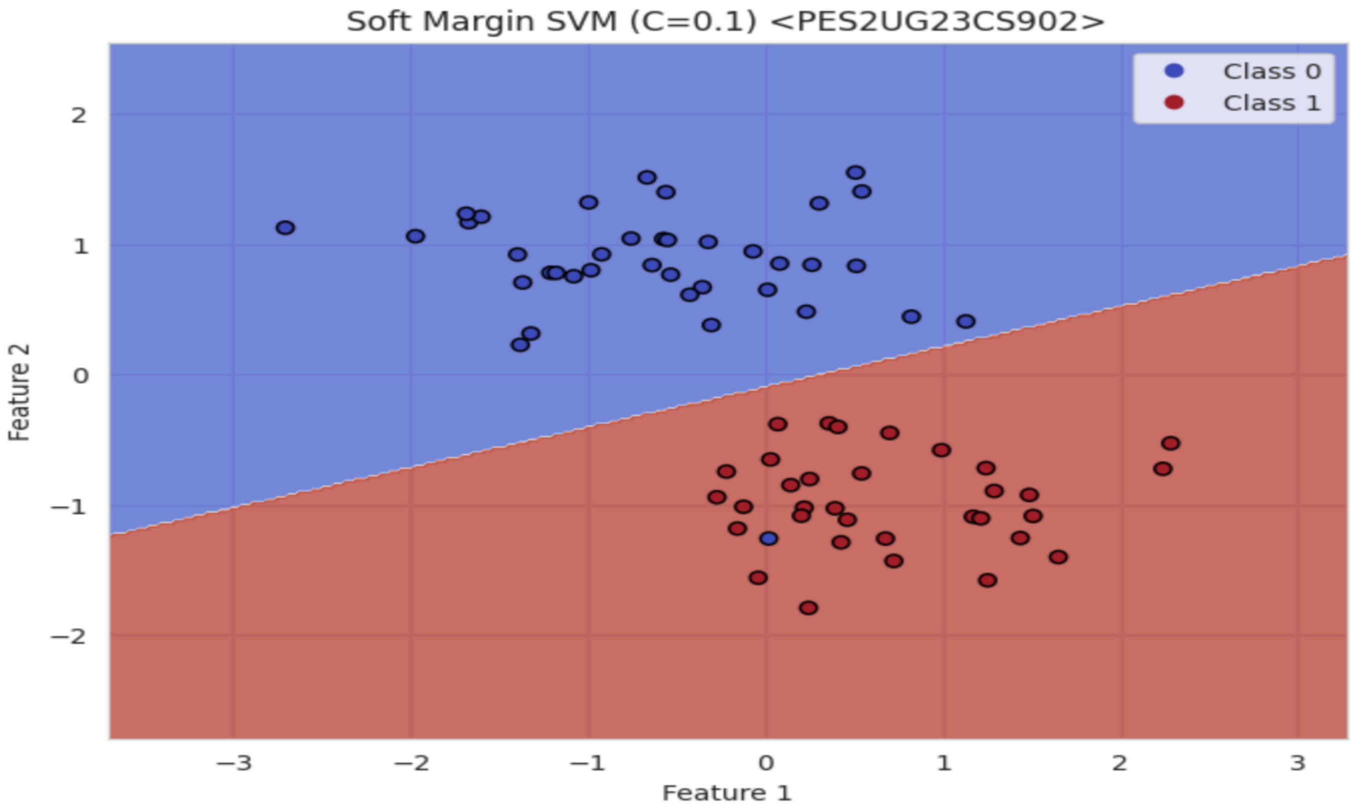


14. Banknote Dataset - SVM with POLY Kernel



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

Soft Margin SVM (C=0.1)



15. Hard Margin SVM (C=100)