Navneet Nayak
PES2UG23CS371
Section 'F'

Questions

Moons Dataset Questions

1. Inferences about the Linear Kernel's performance. The Linear Kernel performed poorly on the Moons dataset. Because the dataset has a non-linear crescent shape, a straight-line decision boundary is fundamentally incapable of separating the two classes effectively. This resulted in a low accuracy score of 0.87 and significant mis-classifications, as seen in the classification report and the decision boundary plot.
2. Comparison between RBF and Polynomial kernel decision boundaries. Both RBF and Polynomial kernels create non-linear decision boundaries suitable for the Moons dataset. O The RBF kernel created a highly flexible, wave-like boundary that closely followed the curve of the data points. O The Polynomial kernel created a single, smooth, C-shaped curve.
3. For this dataset, the RBF kernel's decision boundary appears to be a better fit, as it more accurately captures the distinct separation between the two crescent shapes.

Banknote Dataset Questions

1. Based on the classification reports, the RBF kernel was the most effective. It achieved the highest overall F1-score of 0.93, indicating a superior balance of precision and recall compared to the Linear and Polynomial kernels.
2. The Polynomial kernel likely under-performed because its fixed-degree curve was not the optimal shape for the decision boundary in the Banknote dataset's feature space. If the underlying data separation is more complex or doesn't follow a simple polynomial curve, this kernel can struggle compared to the highly flexible RBF kernel or even a simple linear one if the data is largely linearly separable.

Hard vs. Soft Margin Questions

1. The soft margin (C=0.1) is visibly wider.
2. A soft margin model allows mistakes because its primary goal is to create a wider, more generalized margin rather than perfectly classifying every single training point. This tolerance for error, controlled by a small C value, helps the model avoid overfitting to noise in the training data.

3. The hard margin model (C=100) is more likely to be overfitting. A large C value heavily penalizes misclassification, forcing the model to create a narrow, complex boundary that fits the training data as perfectly as possible. This makes the model highly sensitive to the specific training points, including outliers and noise, which hurts its ability to generalize.
4. I would trust the soft margin model more for new data. Its wider margin is less influenced by the specific noise and outliers of the training set, making it more robust. A model that generalizes better is more reliable when making predictions on unseen data, which is the primary goal of a predictive model.

Screeshots

Moons Dataset

1. Classification Report for SVM with LINEAR Kernel with SRN

```
SVM with LINEAR Kernel <PES2UG23CS371>
              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
```

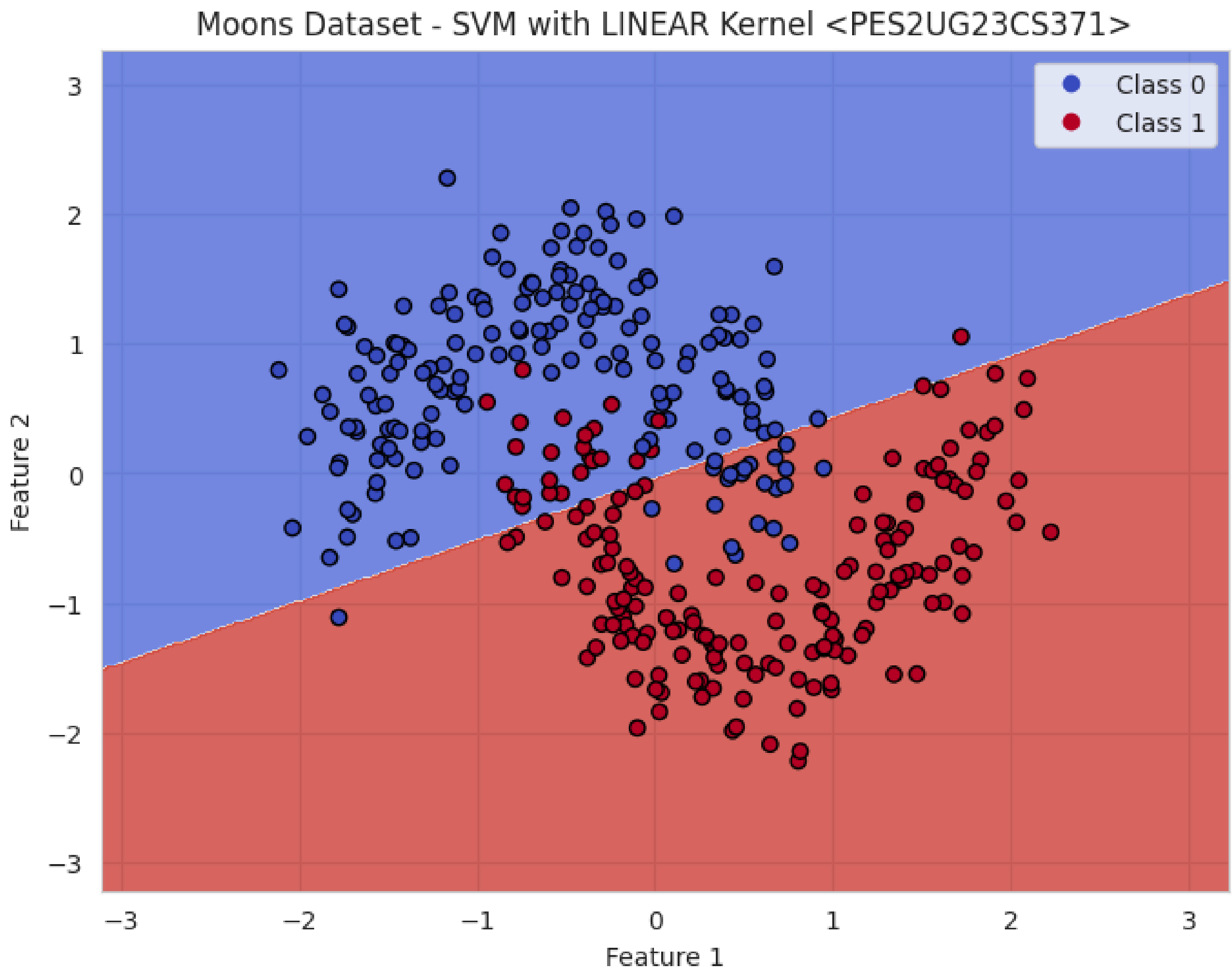
2. Classification Report for SVM with RBF Kernel with SRN

```
SVM with RBF Kernel <PES2UG23CS371>
              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
```

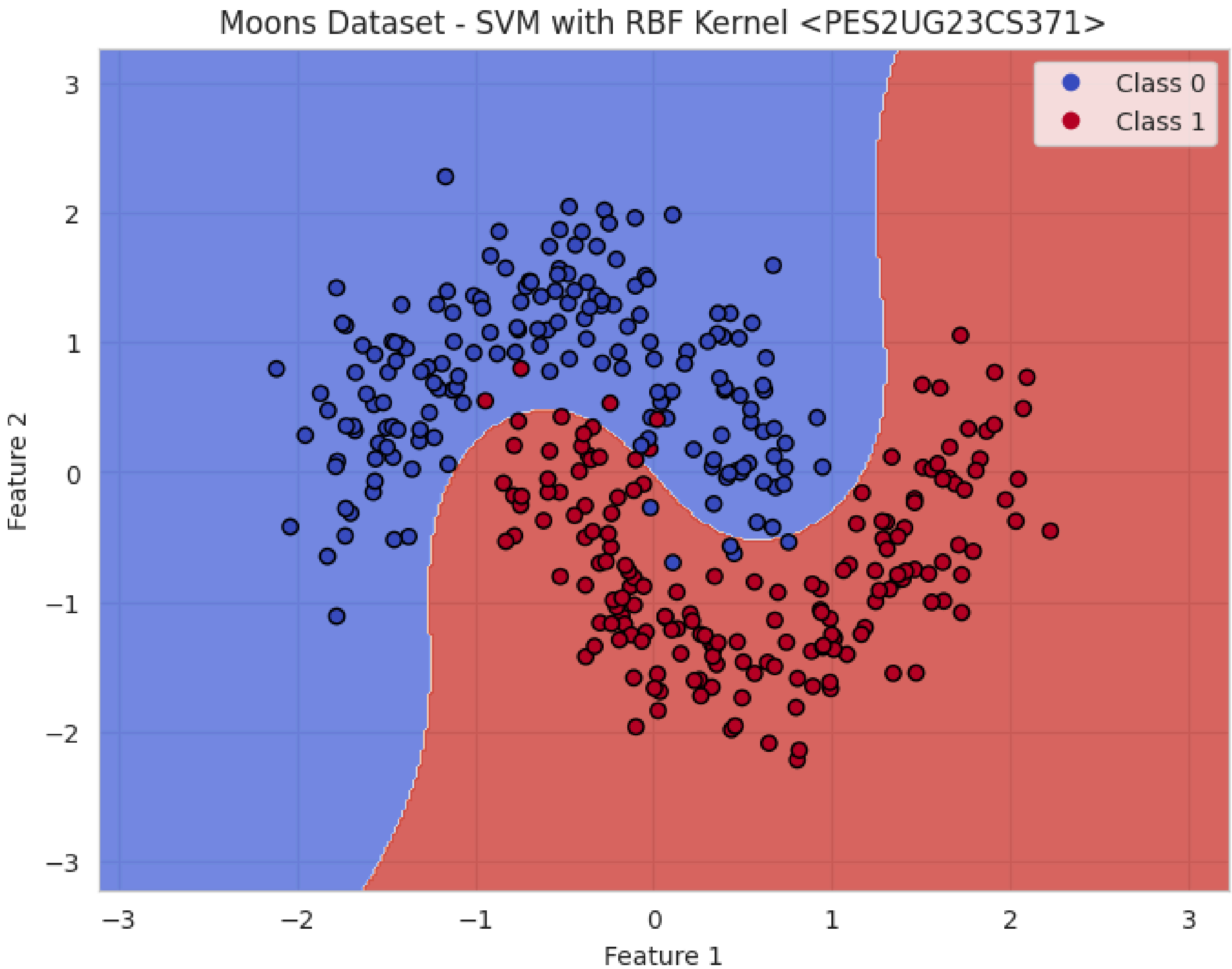
3. Classification Report for SVM with POLY Kernel with SRN

```
SVM with POLY Kernel <PES2UG23CS371>
              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
```

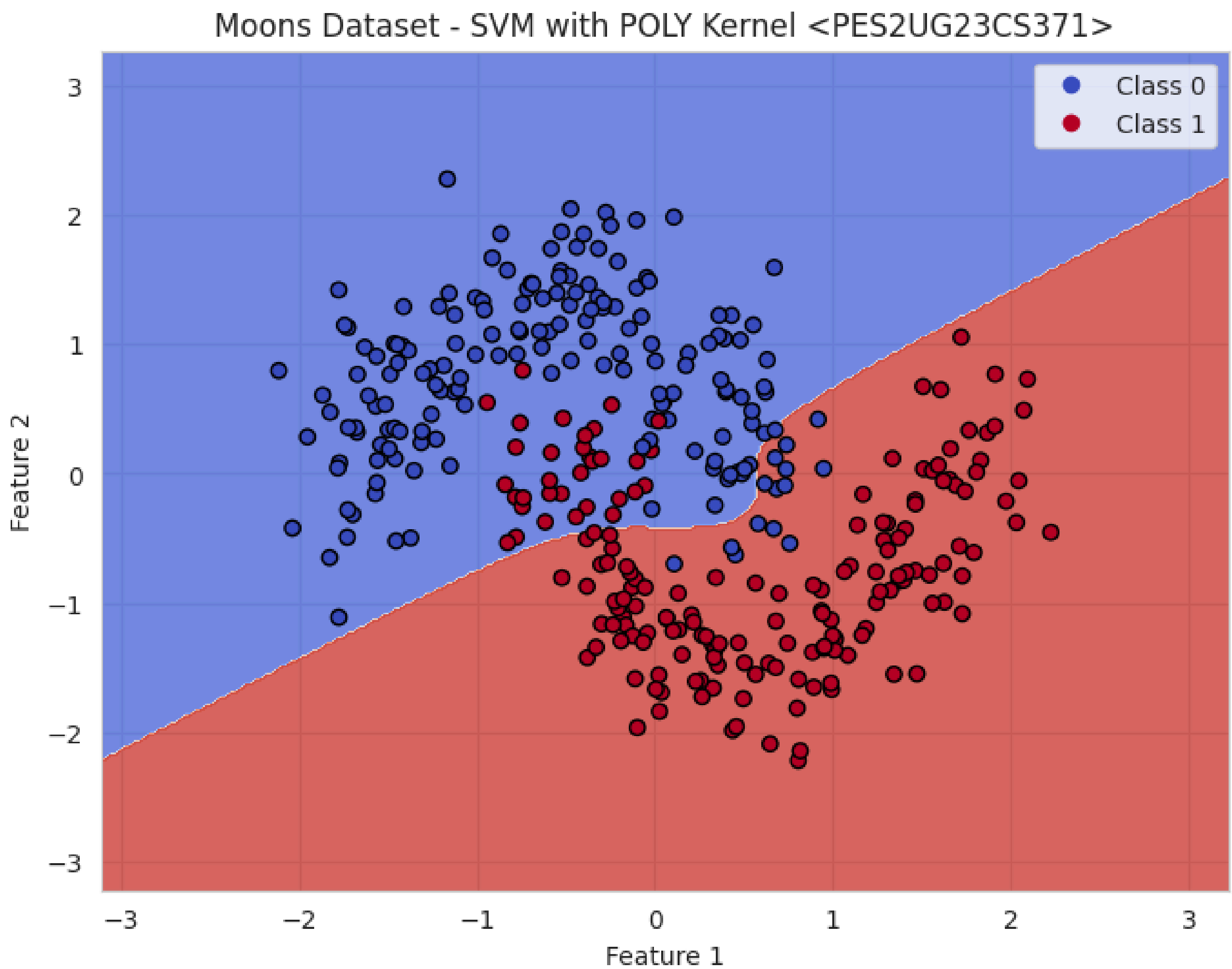
4. Moons Dataset - SVM with LINEAR Kernel



5. Moons Dataset - SVM with RBF Kernel



6. Moons Dataset - SVM with POLY Kernel



Banknote Dataset

1. Classification Report for SVM with LINEAR Kernel

```
SVM with LINEAR Kernel <PES2UG23CS371>
              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
```

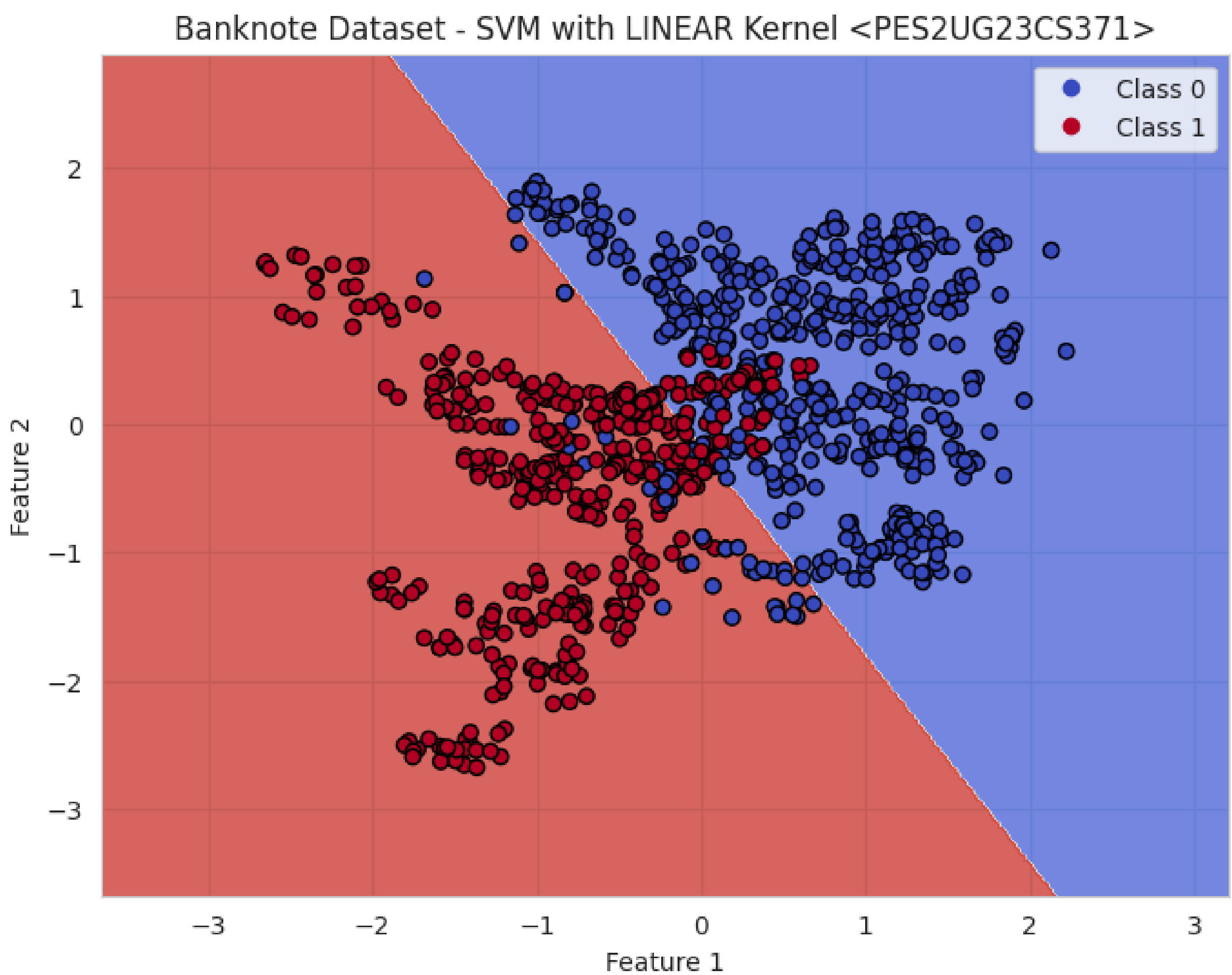
2. Classification Report for SVM with RBF Kernel

```
SVM with RBF Kernel <PES2UG23CS371>
              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
```

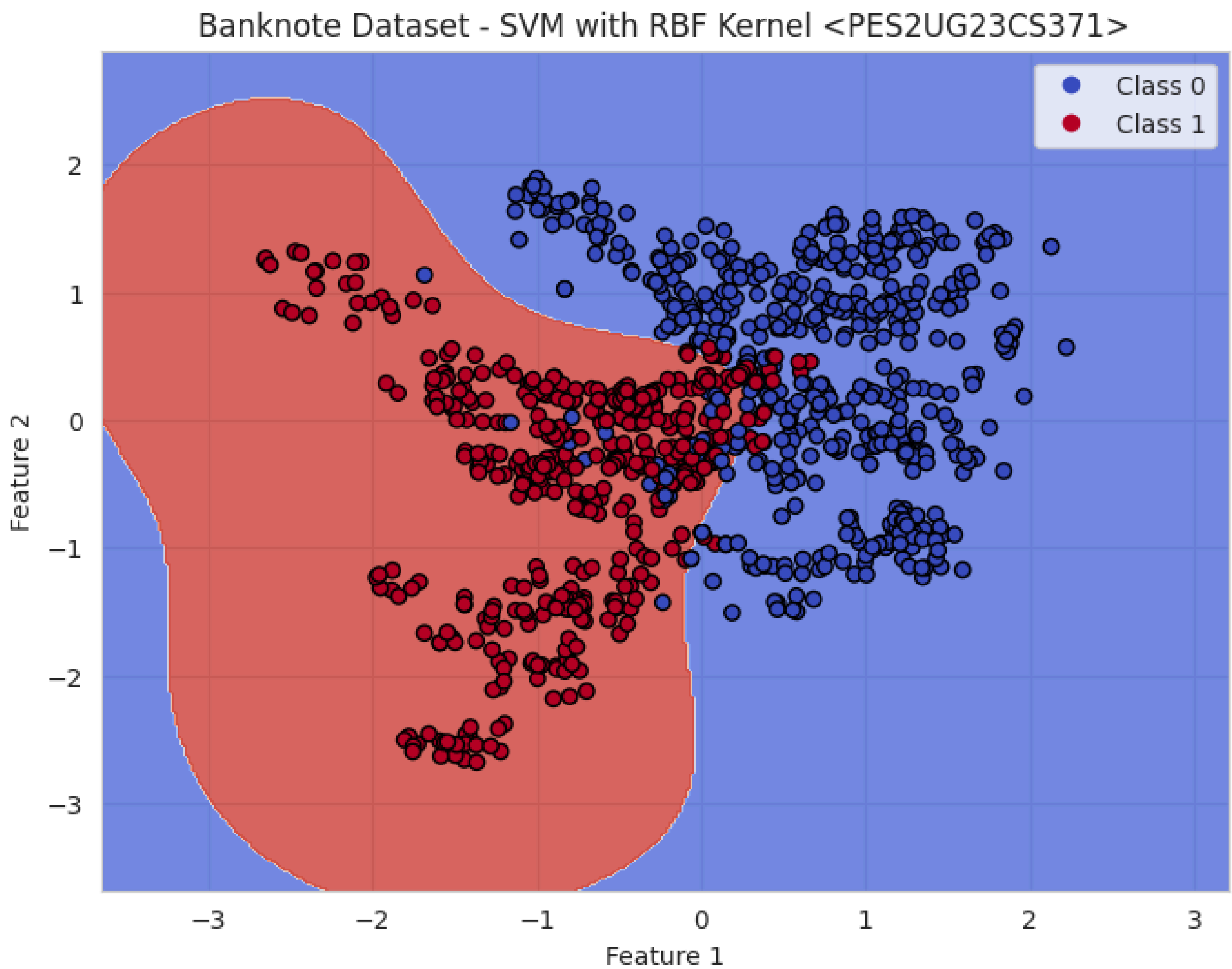
3. Classification Report for SVM with POLY Kernel

```
SVM with POLY Kernel <PES2UG23CS371>
              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
```

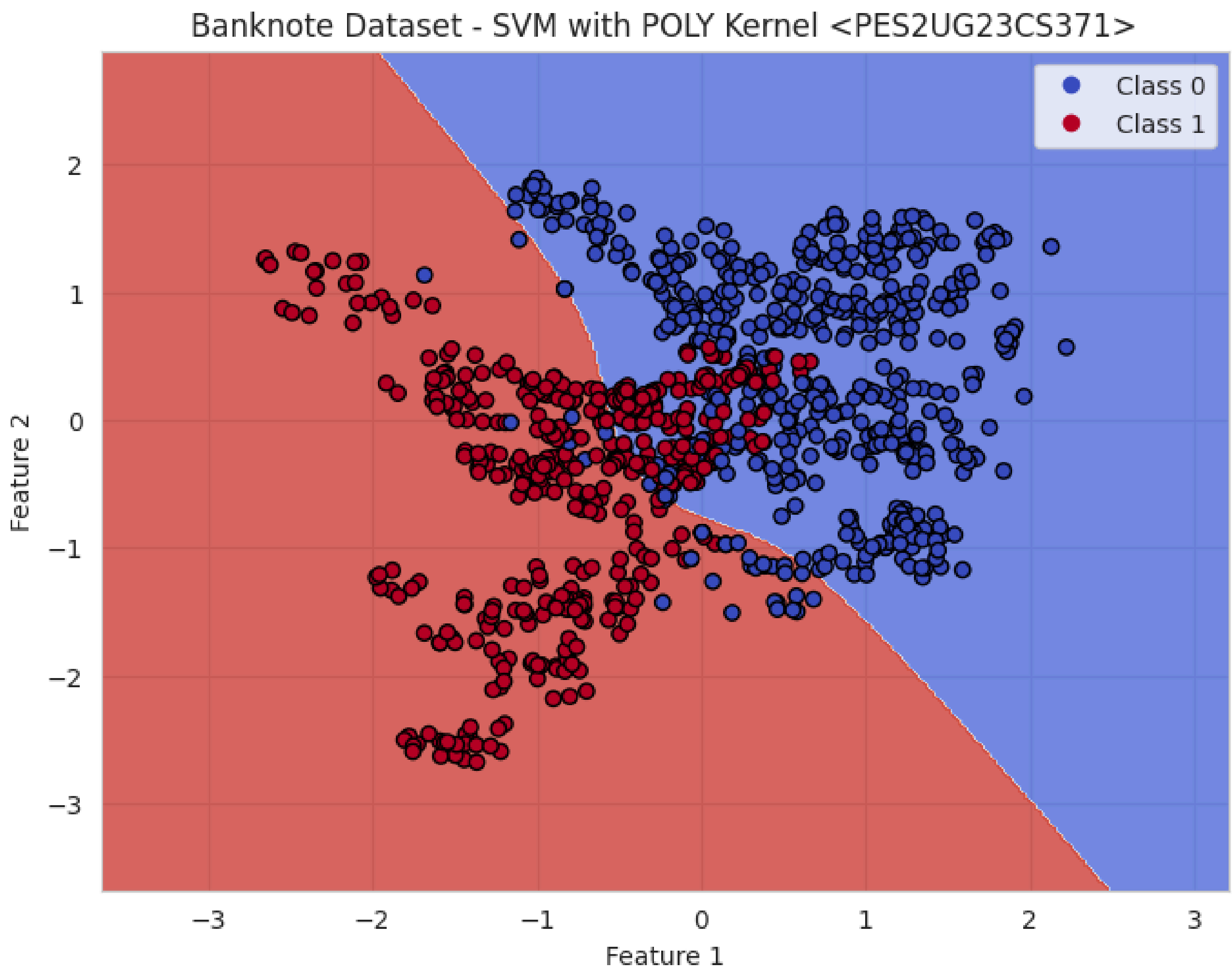
4. Banknote Dataset - SVM with LINEAR Kernel



5. Banknote Dataset - SVM with RBF Kernel

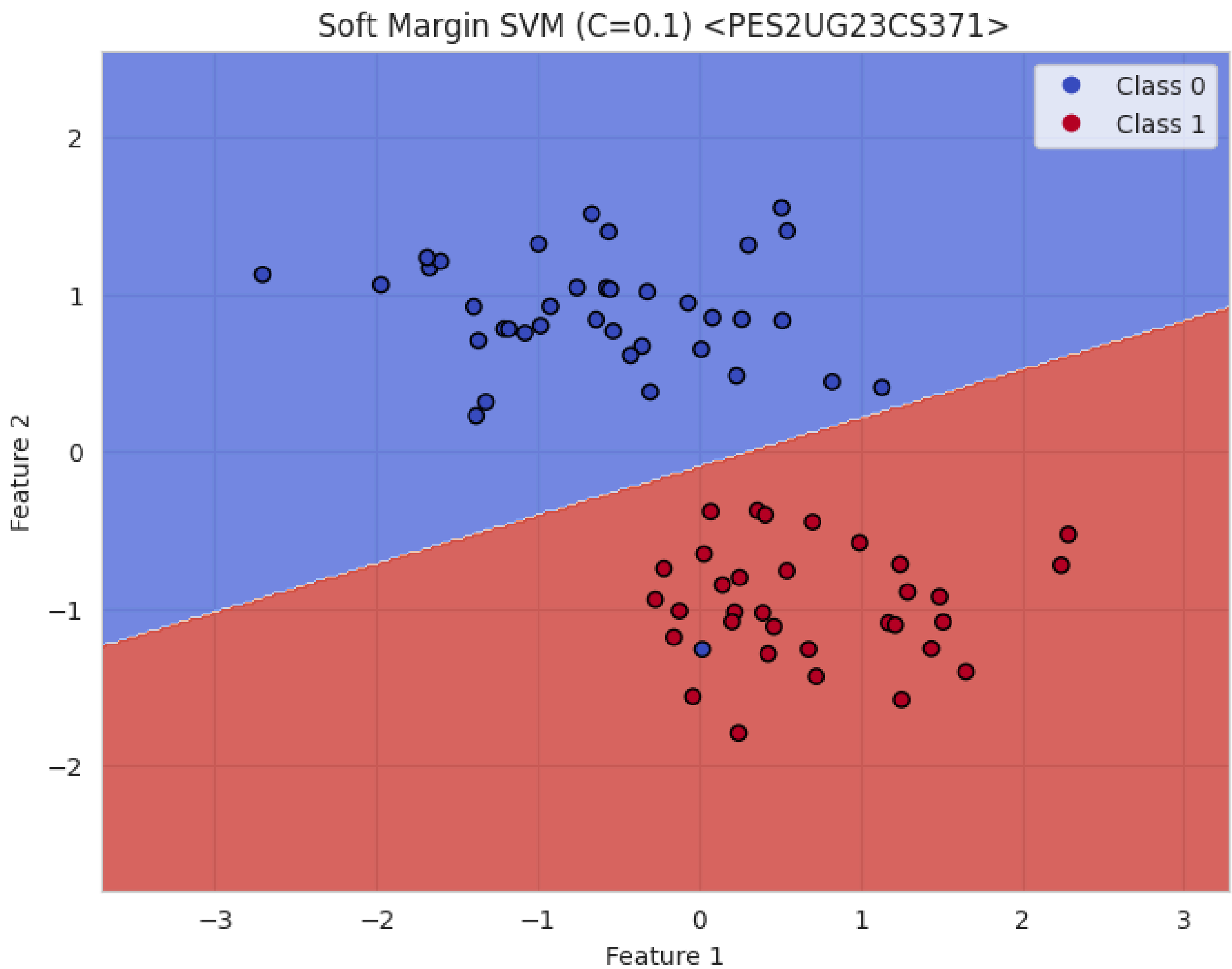


6. Banknote Dataset - SVM with POLY Kernel

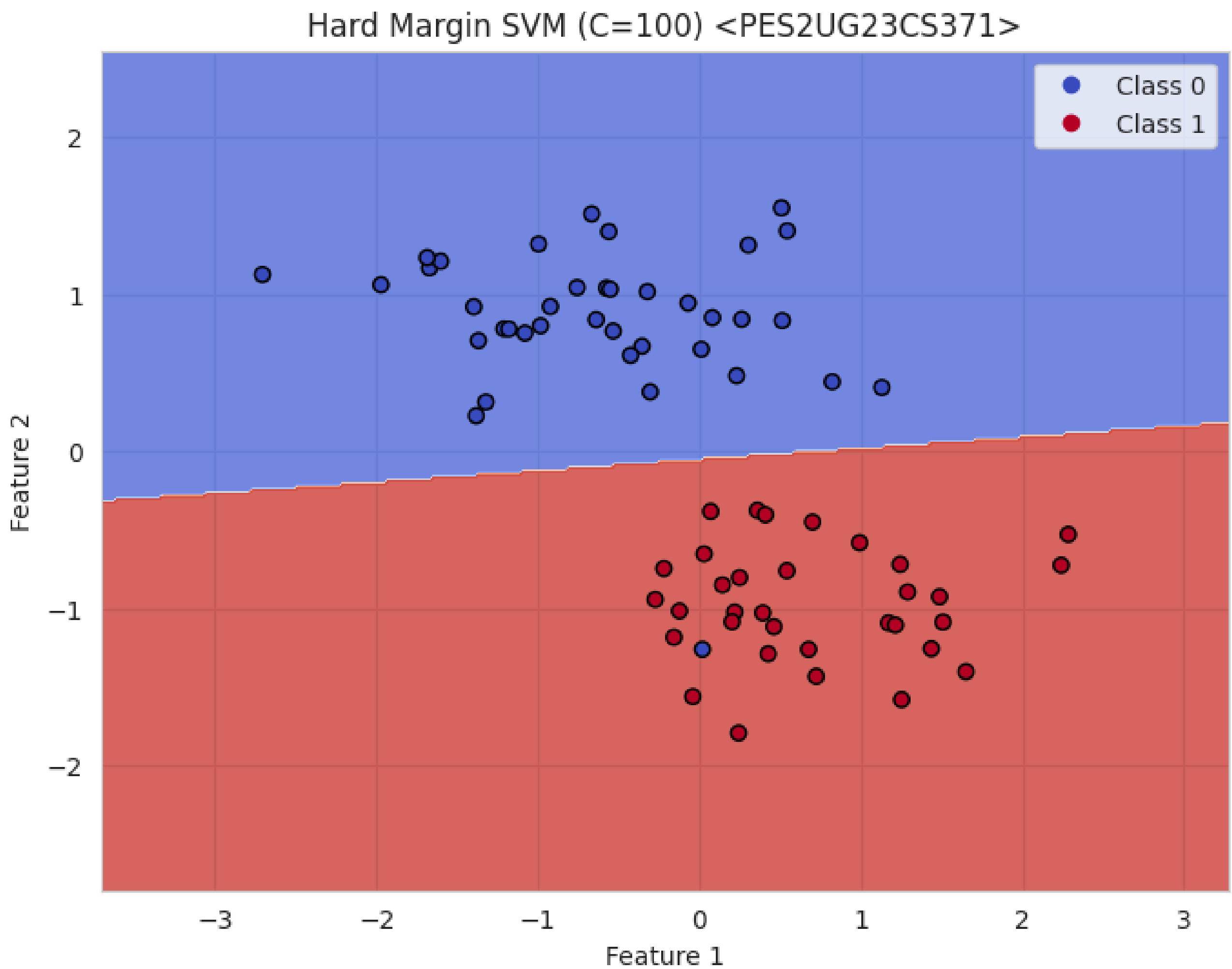


Margin Analysis

1. Soft Margin SVM (C=0.1)



2. .Hard Margin SVM (C=100)



Hard Margin SVM (C=100) <PES2UG23CS371>
Class 0
Class 1
Feature 1
Feature 2
−3
−2
−1
0
1
2
3