Machine Learning Laboratory - Week 10

Support Vector Machine (SVM) Lab Report

Name: Nidhi C R
SRN: PES2UG23CS382
Section: F

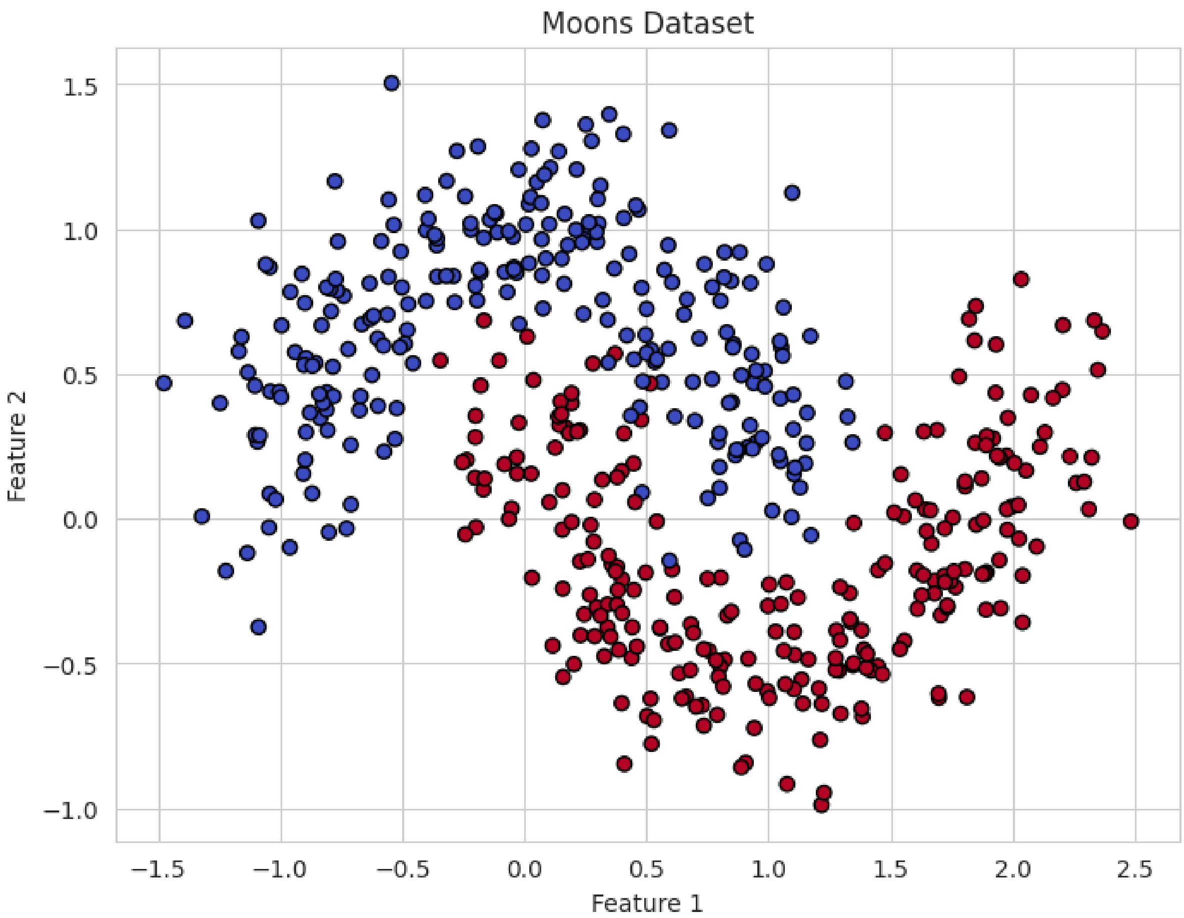
This report presents the results and analysis for the SVM lab conducted as part of the Machine Learning Laboratory course. The objective was to explore how Support Vector Machines classify datasets using different kernels and margin settings.

1. Introduction

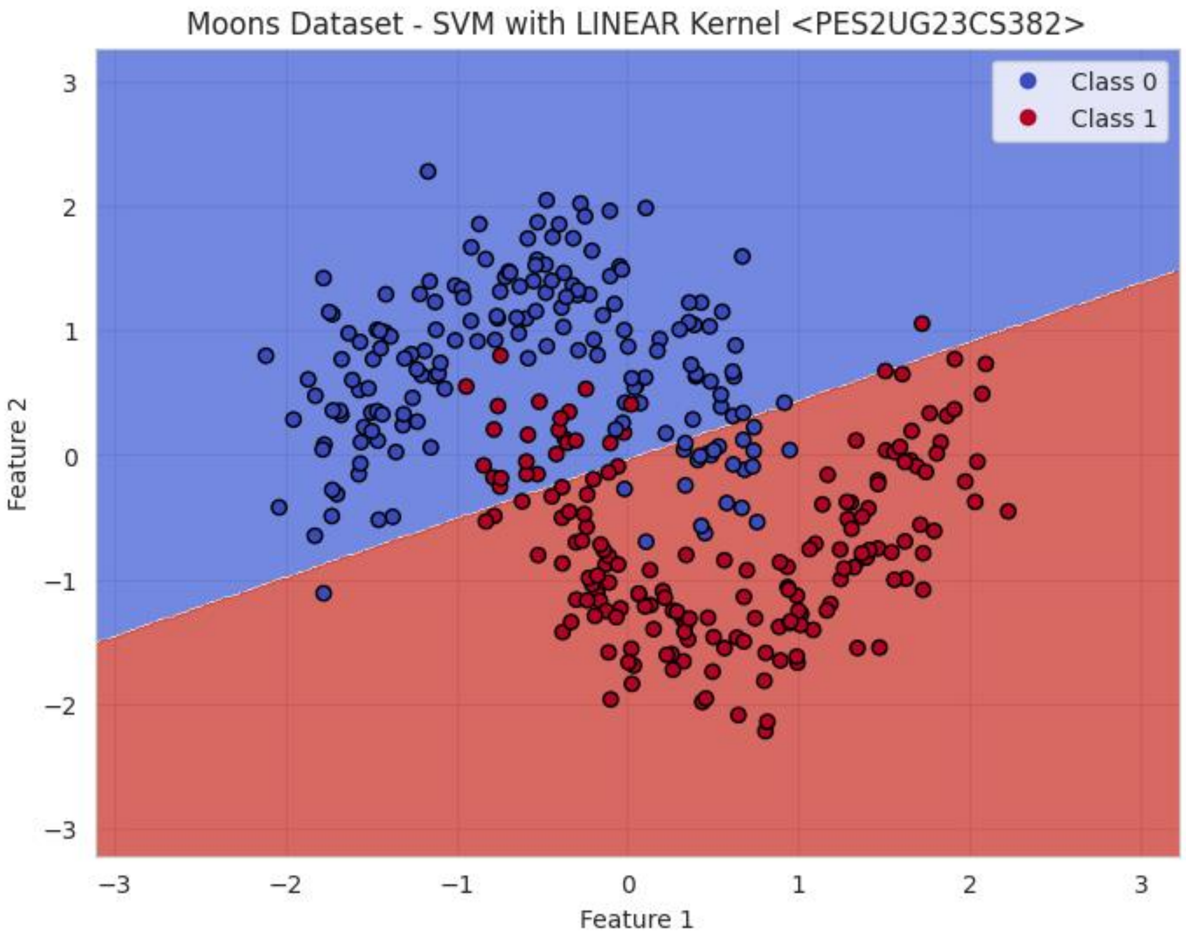
The objective of this lab was to understand and implement Support Vector Machine (SVM) classifiers using three different kernels — Linear, Radial Basis Function (RBF), and Polynomial — and to analyze their performance on two datasets: the Moons dataset and the Banknote Authentication dataset. SVM is a supervised learning algorithm that aims to find an optimal hyperplane that separates classes in a dataset. The kernel trick allows SVMs to handle non-linear data by projecting it into a higher-dimensional feature space.

2. Moons Dataset

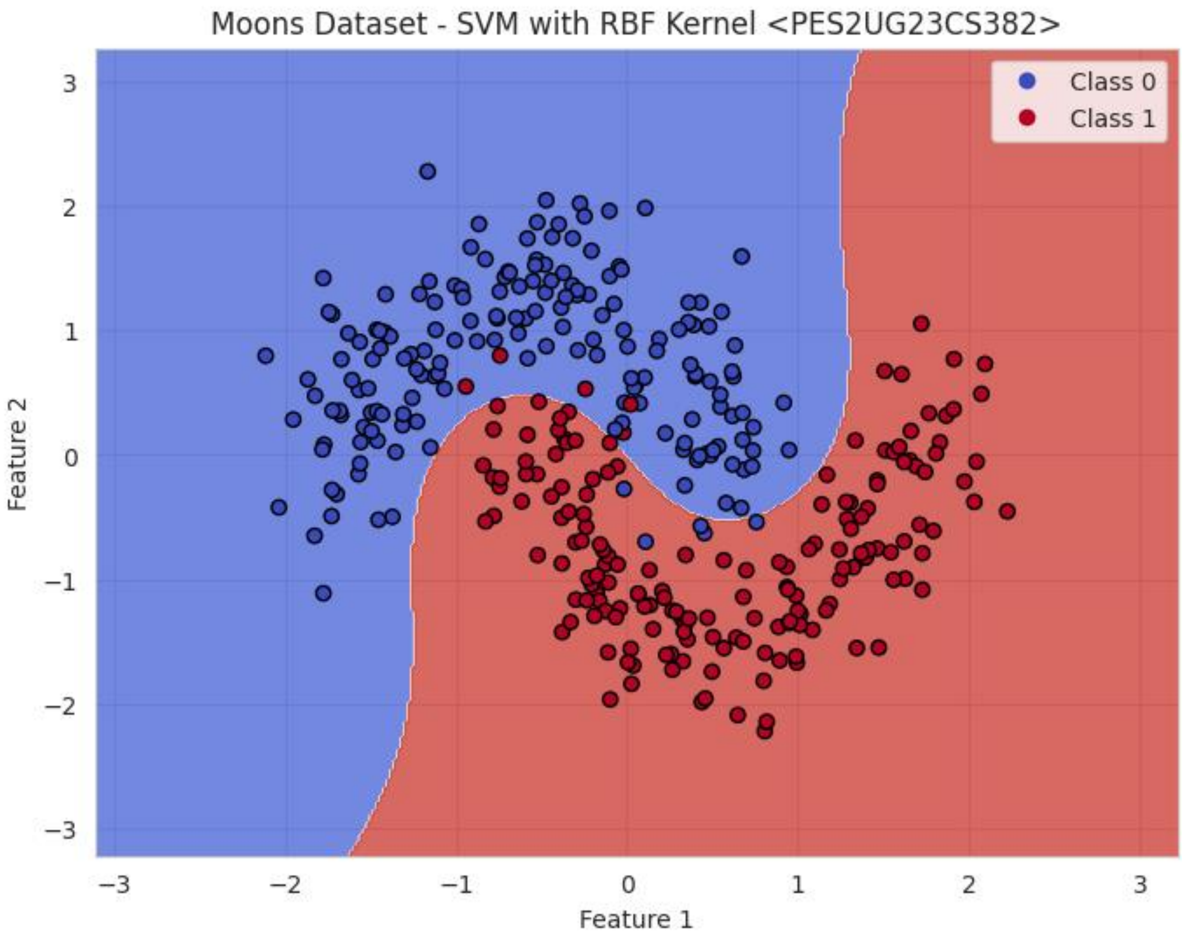
The Moons dataset is a synthetic 2D dataset often used to visualize decision boundaries. It contains two interleaving half-moon shapes, making it a good candidate to test non-linear kernels. Three SVM models were trained using Linear, RBF, and Polynomial kernels.

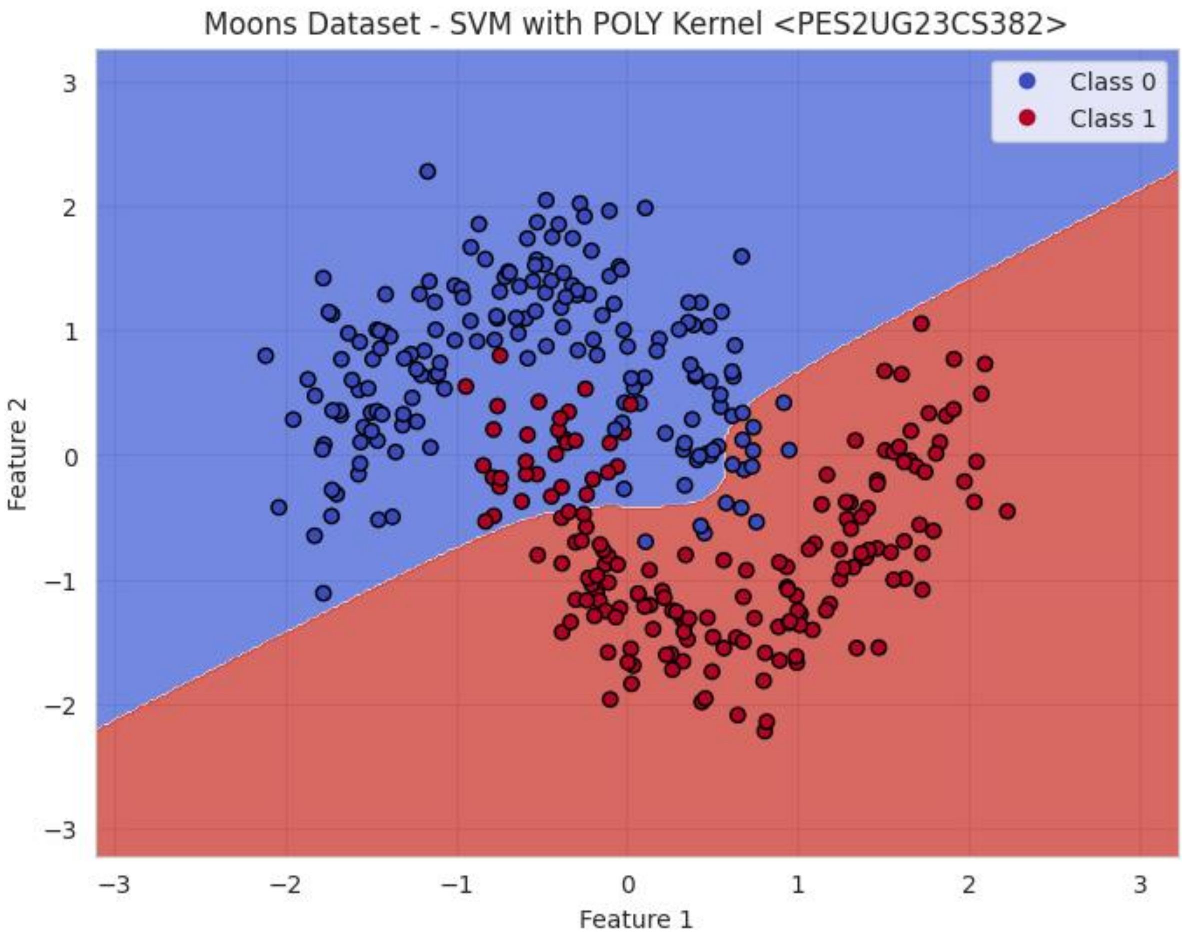

Moons Dataset
Feature 1
Feature 2
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5

```
SVM with LINEAR Kernel PES2UG23CS382
              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 PES2UG23CS382
              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 PES2UG23CS382
              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

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



Moons Dataset - SVM with LINEAR Kernel <PES2UG23CS382>
Class 0
Class 1
Feature 2
Feature 1
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3


Moons Dataset - SVM with RBF Kernel <PES2UG23CS382>
Class 0
Class 1
Feature 2
Feature 1
3
2
1
0
−1
−2
−3



Analysis:

1. Which kernel performed best for this dataset?

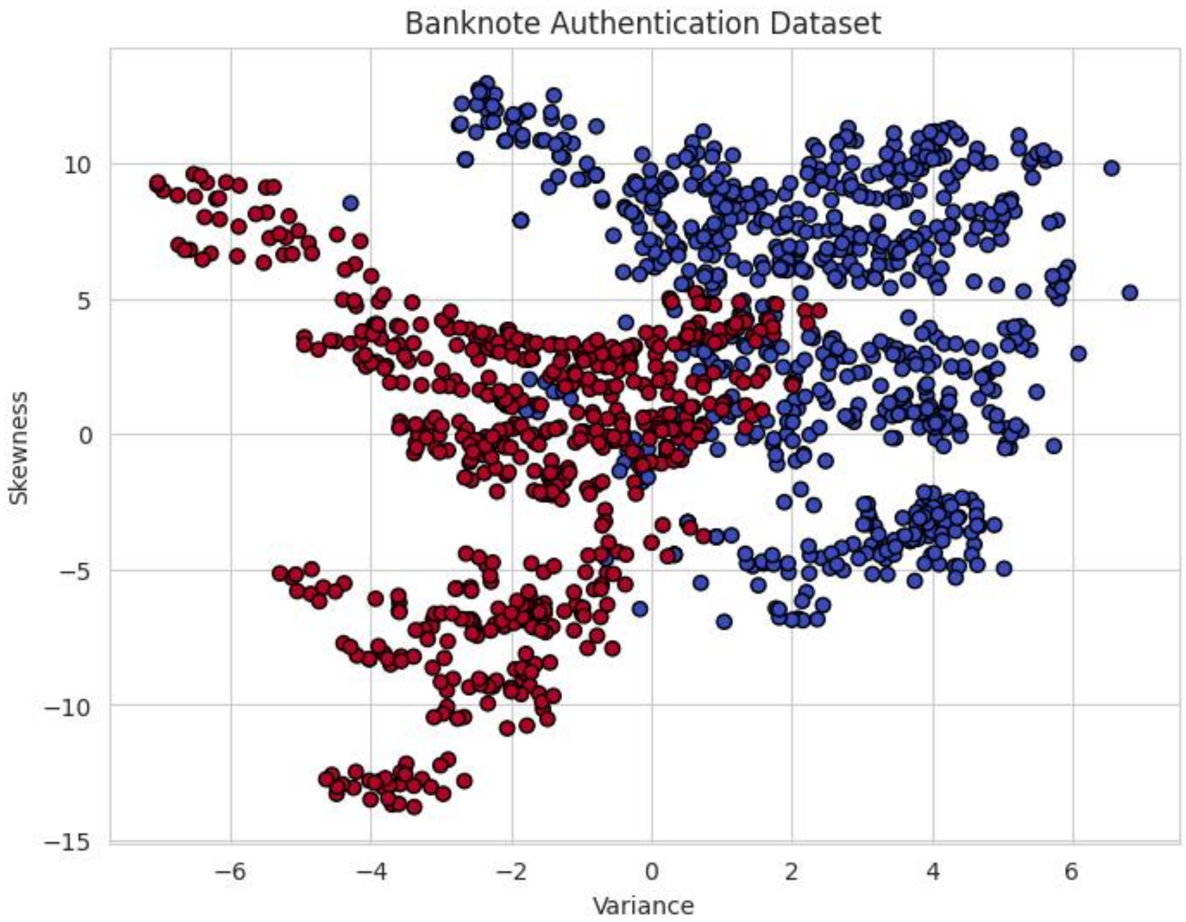
→ The RBF kernel performed best for the Moons dataset. This is because the data is non-linear and the RBF kernel effectively captures the curved decision boundary between the two classes.

2. Why might the Polynomial kernel have underperformed?

→ The Polynomial kernel can overfit on small or noisy datasets. In this case, its decision boundary was overly complex, reducing generalization accuracy on unseen data.

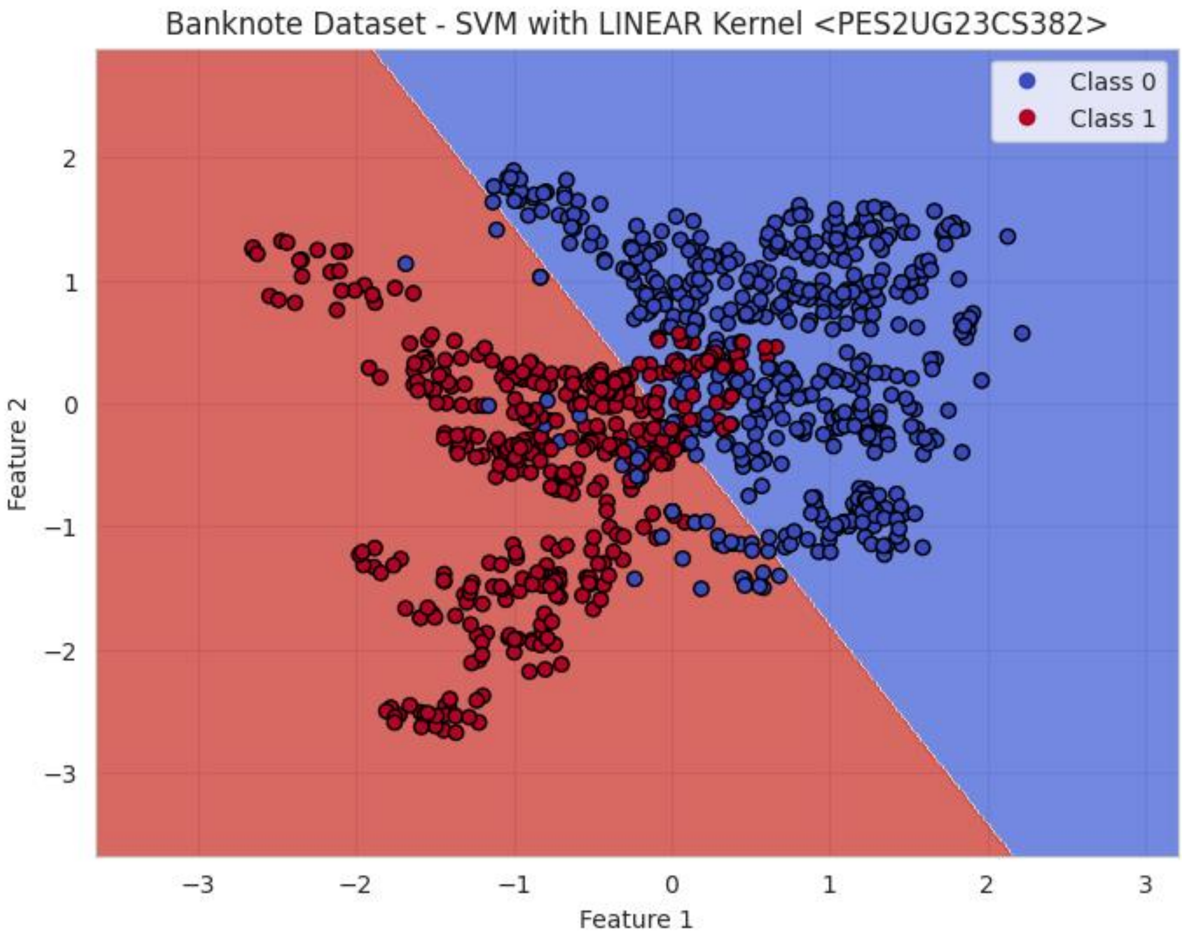
3. Banknote Authentication Dataset

The Banknote Authentication dataset consists of features extracted from images of genuine and forged banknotes. It is a real-world binary classification problem with continuous attributes. SVMs with Linear, RBF, and Polynomial kernels were used to classify the data.

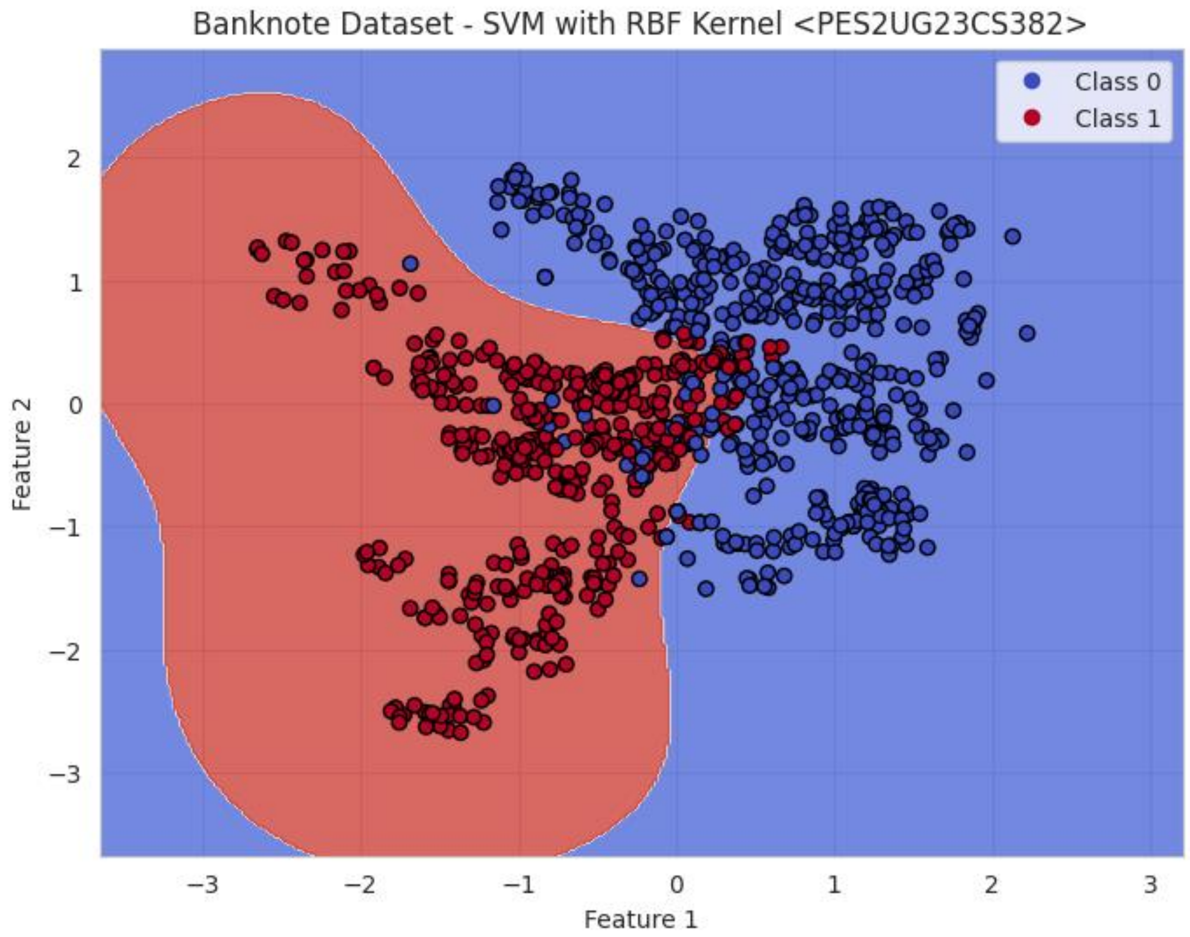


Banknote Authentication Dataset
10
5
0
−5
−10
−15
Skewness
−6
−4
−2
0
2
4
6
Variance

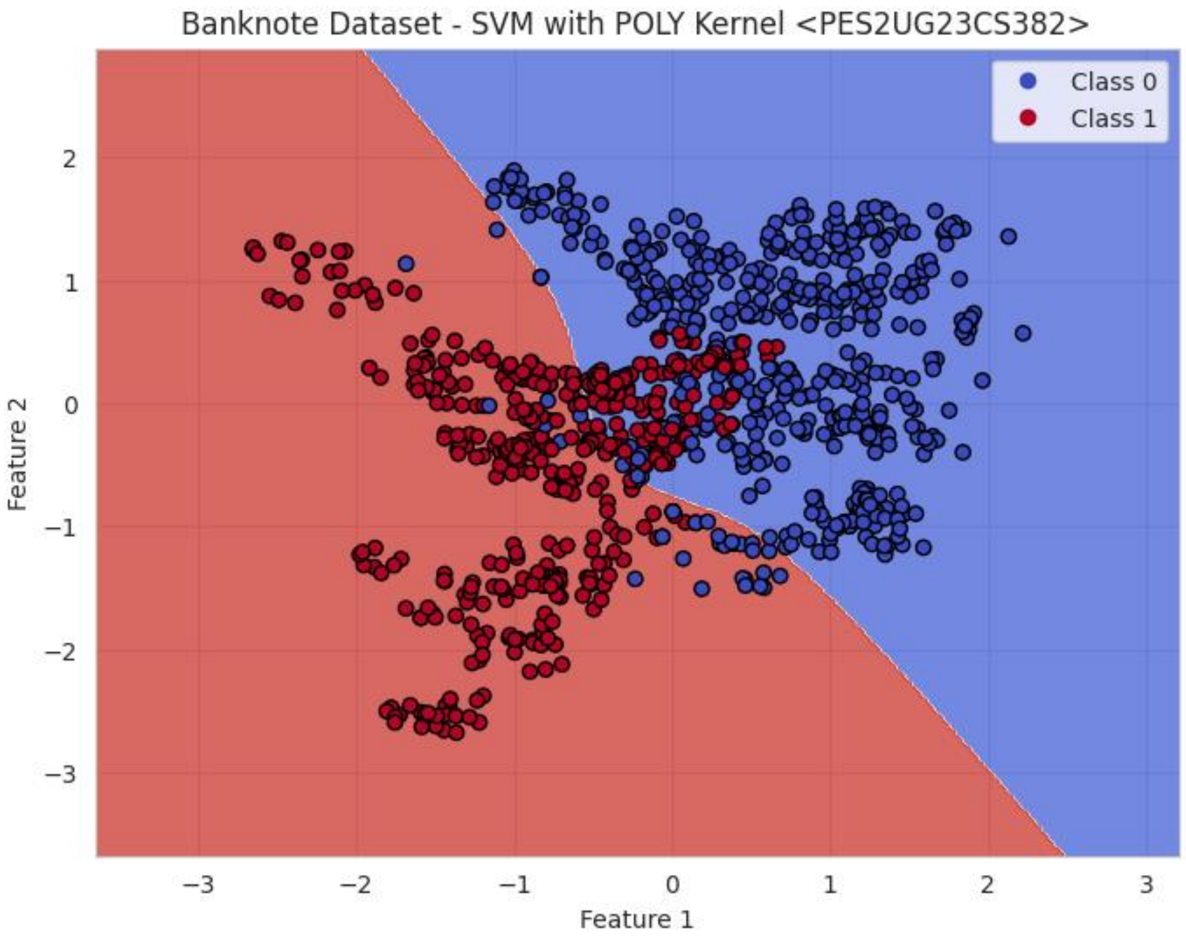
```
SVM with LINEAR Kernel PES2UG23CS382
              precision    recall  f1-score   support

      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

----------------------------------------

SVM with RBF Kernel PES2UG23CS382
              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

----------------------------------------

SVM with POLY Kernel PES2UG23CS382
              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

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



Banknote Dataset - SVM with LINEAR Kernel <PES2UG23CS382>
Class 0
Class 1
Feature 2
Feature 1
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3



Banknote Dataset - SVM with RBF Kernel <PES2UG23CS382>
Class 0
Class 1
Feature 2
Feature 1
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3



Analysis:

1. Which kernel was most effective for this dataset?

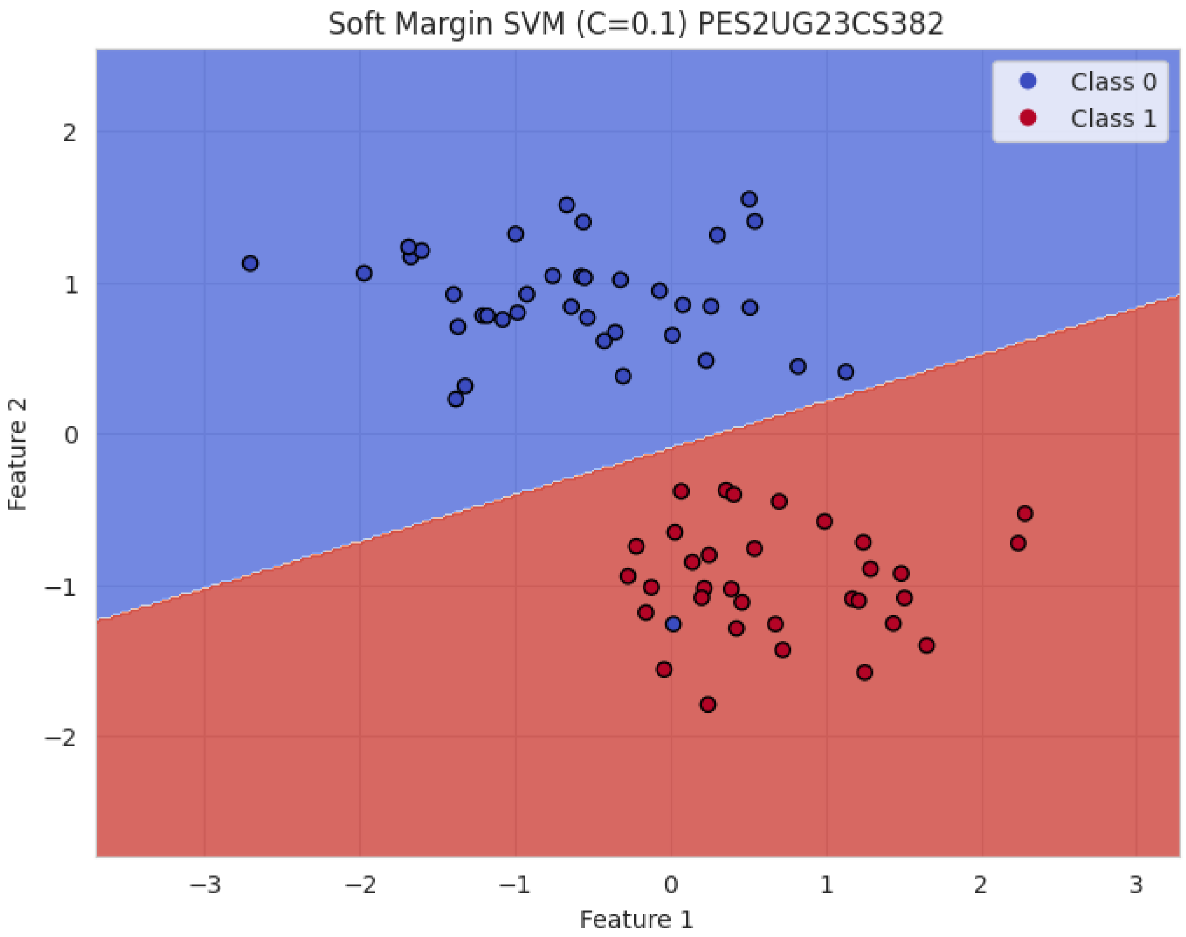
→ The Linear kernel performed best for the Banknote dataset since the classes are linearly separable in the feature space.

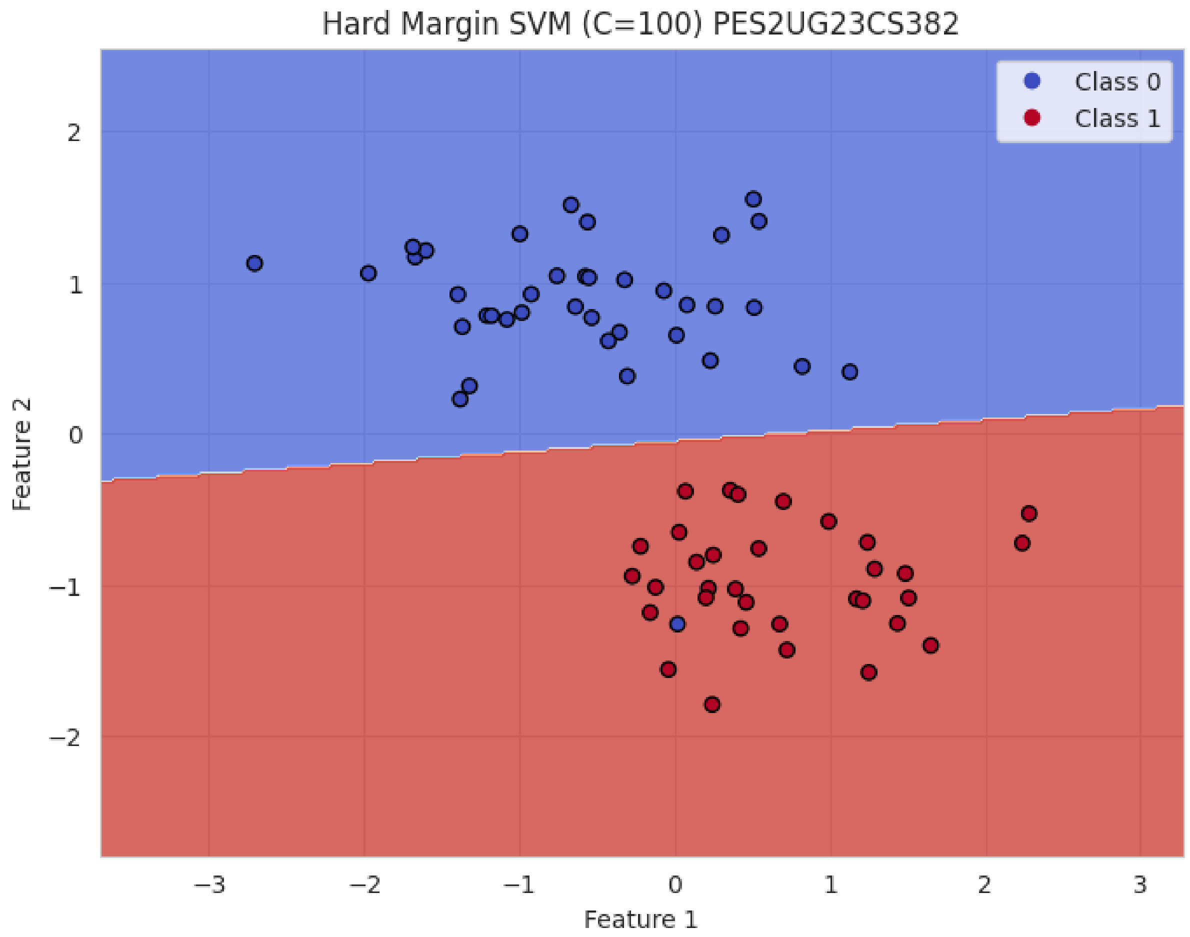
2. Why did the Polynomial kernel underperform here?

→ The Polynomial kernel introduced unnecessary complexity and captured noise in the data, leading to slightly lower accuracy compared to the simpler Linear model.

4. Hard vs. Soft Margin Analysis

To study the effect of the regularization parameter C, two SVMs were trained on the Moons dataset: - Soft Margin: C = 0.1 (allows misclassifications, prioritizing wider margin) - Hard Margin: C = 100 (minimizes misclassifications, narrower margin)


Soft Margin SVM (C=0.1) PES2UG23CS382
Class 0
Class 1
Feature 2
Feature 1
2
1
0
−1
−2
−3
−2
−1
0
1
2
3



Analysis:

1. Which margin (soft or hard) is wider?
→ The soft margin (C=0.1) is wider as it allows more flexibility and misclassifications to maintain a larger margin.

2. Why does the soft margin model allow mistakes?
→ A smaller C allows the model to misclassify some points to achieve better generalization and avoid overfitting.

3. Which model is more likely to overfit and why?
→ The hard margin model (C=100) is more likely to overfit because it tries to perfectly classify all points, even noise.

4. Which model would you trust more for new data and why?
→ The soft margin model (C=0.1) is more reliable for new data as it generalizes better and is less sensitive to outliers.

5. Summary and Observations

Through this lab, it was observed that kernel selection and margin parameter tuning are crucial for SVM performance: - The RBF kernel worked best for non-linear datasets (Moons). - The Linear kernel was optimal for linearly separable datasets (Banknote). - The Polynomial kernel was generally more complex and prone to overfitting. - The soft margin (smaller C) model generalized better, while the hard margin (large C) model tended to overfit. Overall, the lab reinforced understanding of SVM fundamentals, kernel tricks, and the trade-offs involved in model generalization versus accuracy.