ML Lab Week 10

Title: ML Lab Week 10 — SVM Classifier Lab

NAJMUS SEHER

SRN: PES2UG23CS359

Section :F

A. Moons Dataset Questions

Q1. Inferences about the Linear Kernel's performance

The **Linear Kernel** performed moderately well but struggled to accurately classify the curved structure of the Moons dataset.
Since the data is non-linearly separable, a straight-line decision boundary cannot effectively split the two interlocking classes.
As a result, the Linear kernel achieved reasonable accuracy but had more misclassified points near the curved regions.

Q2. Comparison between RBF and Polynomial kernel decision boundaries

The **RBF Kernel** produced a smooth and flexible boundary that adapted perfectly to the moon-shaped clusters, leading to excellent accuracy.
In contrast, the **Polynomial Kernel** created an overcomplicated boundary that slightly overfit the data, especially when the degree was high.
While both are non-linear, the RBF kernel generalized better because it captures complex relationships without unnecessary curvature.

B. Banknote Dataset Questions

Q3. Which kernel was most effective for this dataset?

For the **Banknote Authentication Dataset**, the **Linear Kernel** performed the best.
This dataset's features (like variance and skewness of wavelet-transformed images) are nearly linearly separable.
Thus, a simple linear hyperplane achieved high accuracy with low computational cost and excellent generalization.

Q4. Why might the Polynomial kernel have underperformed here?

The **Polynomial Kernel** adds unnecessary complexity to a dataset that is already linearly separable.
It tends to fit minor variations or noise in the data, which can lead to **overfitting**.
As a result, it slightly reduces test performance compared to the Linear kernel, which provides a cleaner and more stable boundary.

C. Hard vs. Soft Margin Questions

Q5. Which margin (soft or hard) is wider?

The **Soft Margin (C = 0.1)** is wider because the model allows some misclassifications to maximize the overall margin width.
This results in smoother decision boundaries and better generalization to unseen data.

Q6. Why does the soft margin model allow "mistakes"?

The soft margin model allows certain points to be misclassified intentionally to prevent overfitting.

It balances between maximizing the margin and minimizing classification error, which makes the model more robust and adaptable.

Q7. Which model is more likely to be overfitting and why?
The **Hard Margin model (C = 100)** is more prone to overfitting because it forces the classifier to perfectly fit all training points, including outliers or noise.
This results in a very narrow margin that performs poorly on unseen data.

Q8. Which model would you trust more for new data and why?
The **Soft Margin model** is more trustworthy for new data.
By allowing small classification errors, it maintains a larger margin and generalizes better.
It avoids overfitting and achieves a stable performance across different datasets.

Moons Dataset (3 screenshots):

1)Classification Report – SVM with Linear Kernel

```
SVM with LINEAR Kernel <PES2UG23CS359>
              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 – SVM with RBF Kernel

```
SVM with RBF Kernel <PES2UG23CS359>
              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

SVM with POLY Kernel <PES2UG23CS359>

	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. Classification Report – SVM with Linear Kernel

SVM with LINEAR Kernel <PES2UG23CS359>

	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 – SVM with RBF Kernel

SVM with RBF Kernel <PES2UG23CS359>

	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

----

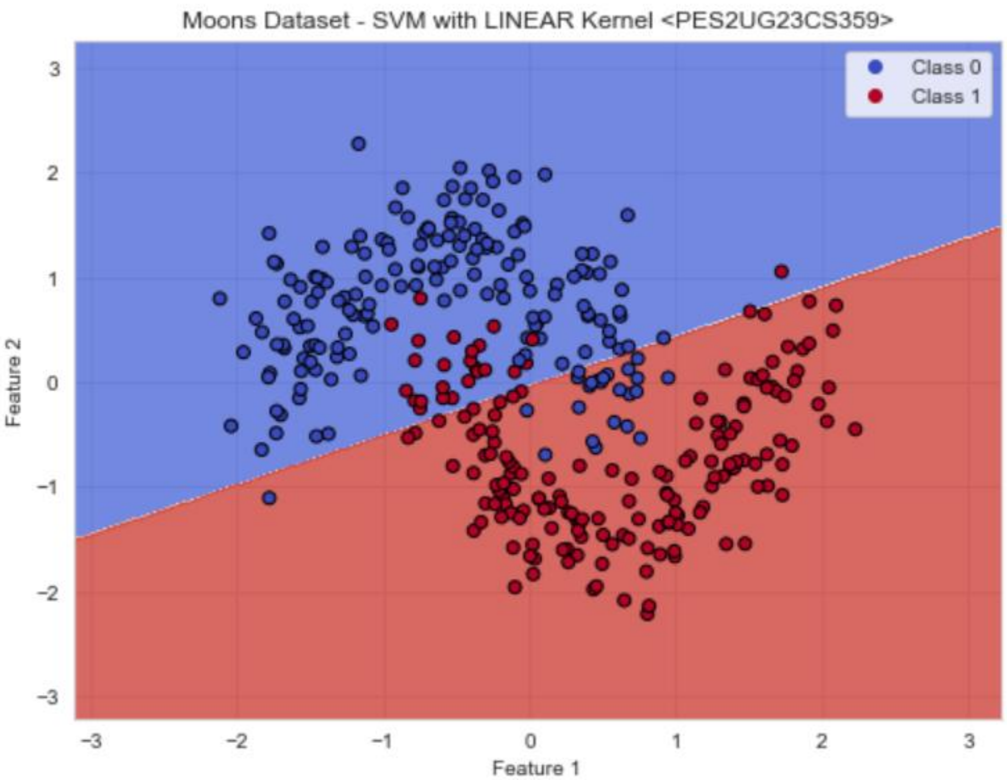
6. Classification Report – SVM with Polynomial Kernel

```
SVM with POLY Kernel <PES2UG23CS359>
              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

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

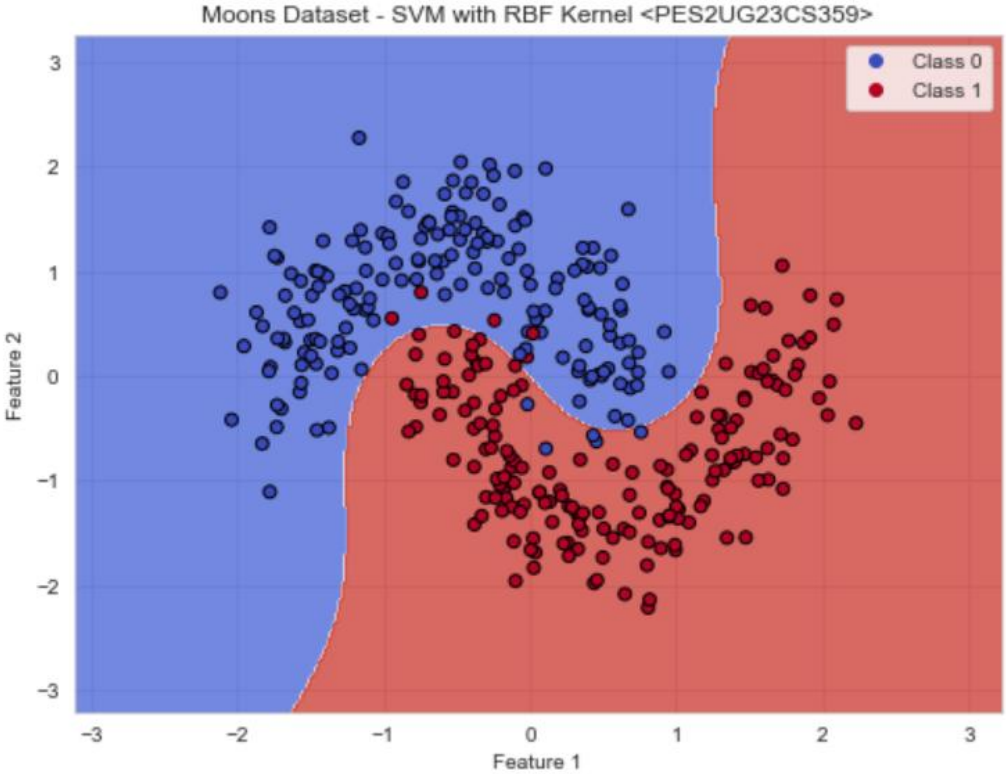
2. Decision Boundary Visualizations (8 Screenshots)

1)Moons Dataset (3 Plots):

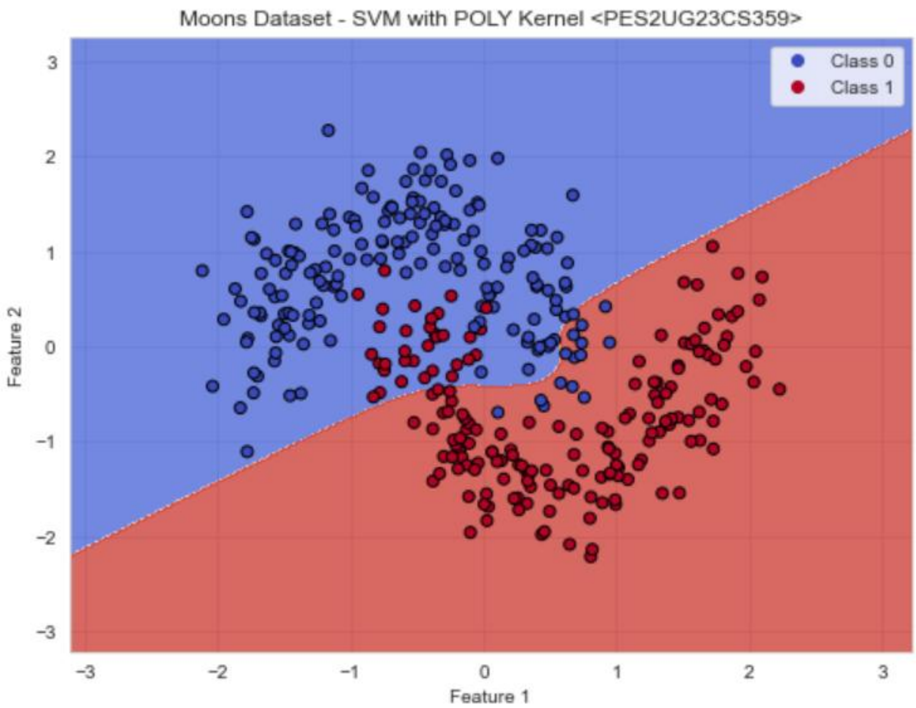
7. Linear Kernel Decision Boundary



8. RBF Kernel Decision Boundary

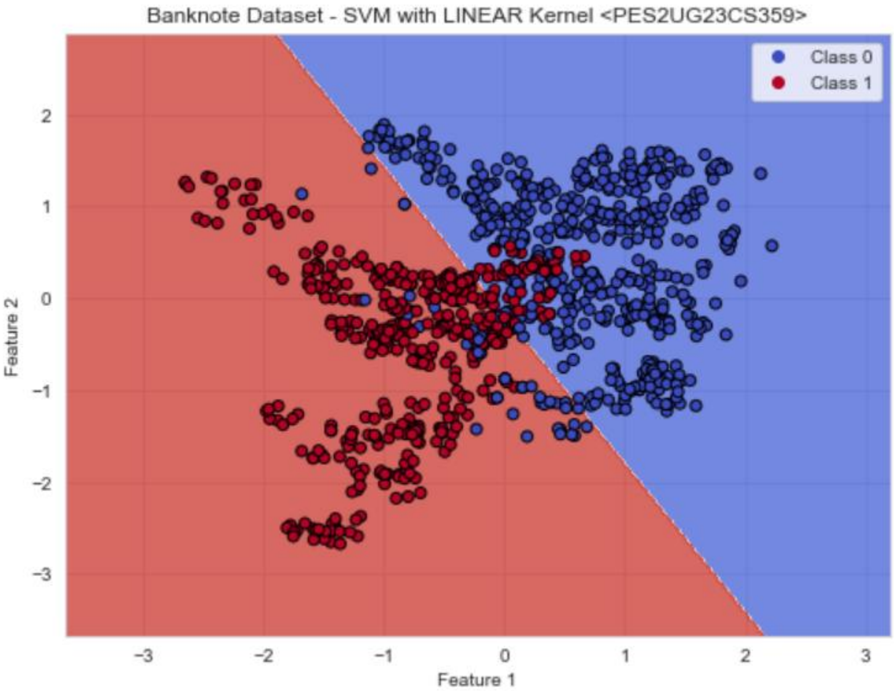


9. Polynomial Kernel Decision Boundary

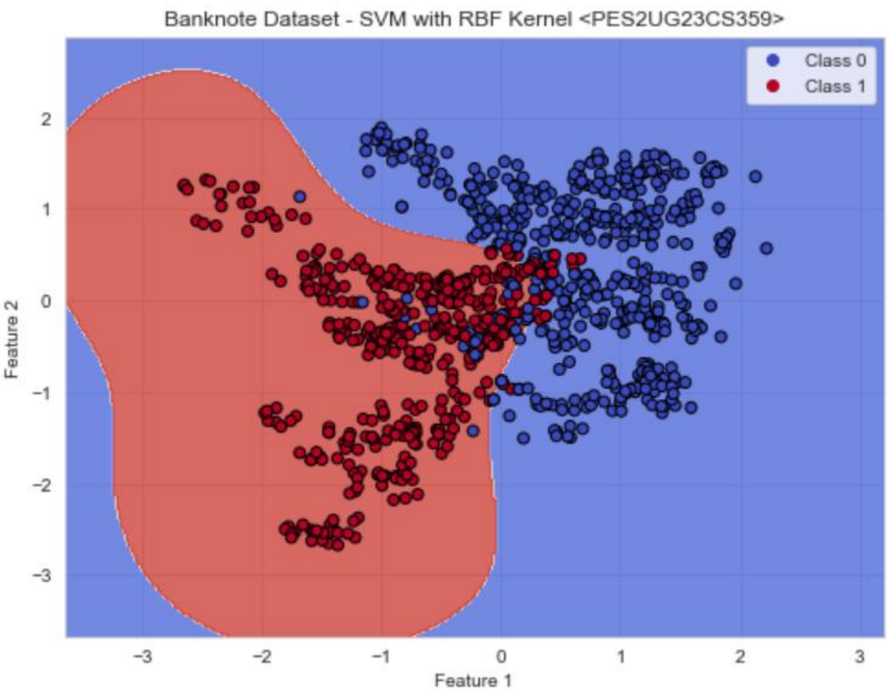


Banknote Dataset (3 Plots):

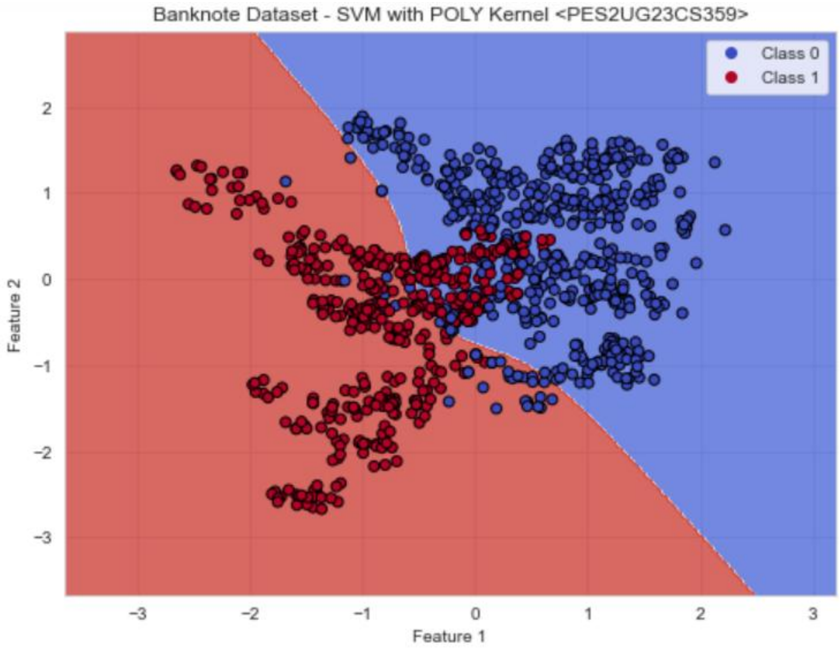
10. Linear Kernel Decision Boundary



11. RBF Kernel Decision Boundary

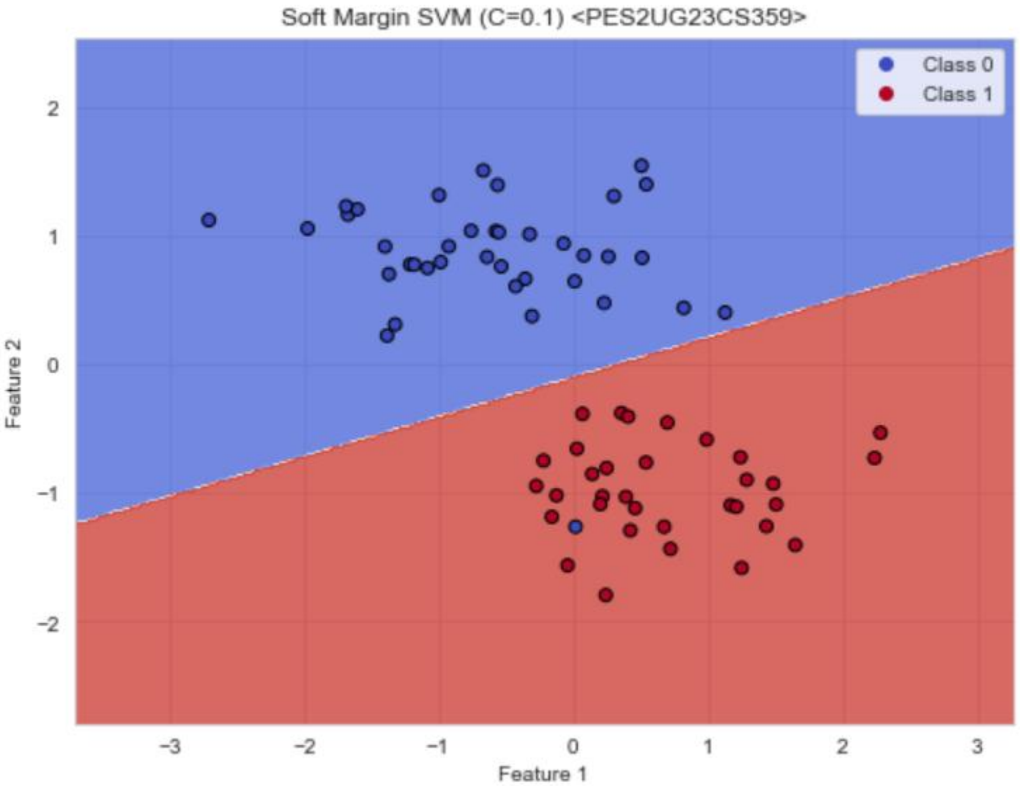


12. Polynomial Kernel Decision Boundary



Margin Analysis (2 Plots):

13. Soft Margin SVM (C = 0.1)



14. Hard Margin SVM (C = 100)