Machine Learning Lab Week 10
Support Vector Machines (SVM)

NAME: DAKSH YADAV
SRN: PES2UG23CS926
SECTION: C

Analysis Questions

Moons Dataset Questions

1.Based on the metrics and the visualizations, what inferences about the performance of the Linear Kernel can you draw?

The Linear kernel achieved an accuracy of 0.87, with balanced precision and recall (≈0.85 - 0.89) for both classes. This indicates that while the model separates the data reasonably well, it struggles slightly around the overlapping boundary regions where the classes are not linearly separable.

From the visualization, the linear decision boundary appears as a straight line, which cannot fully adapt to the curved “moon-shaped” distribution of points. Consequently, a few samples near the boundary are misclassified. Overall, the linear SVM provides a decent baseline but is limited for non-linear data like Moons.

2.Compare the decision boundaries of the RBF and Polynomial kernels. Which one seems to capture the shape of the data more naturally?

The RBF kernel achieved the highest accuracy (0.97) with excellent precision and recall for both classes, indicating nearly perfect separation. The decision boundary from the RBF kernel closely follows the natural curved contours of the Moons dataset, wrapping around each cluster with a smooth, flexible boundary.

In contrast, the Polynomial kernel performed slightly better than linear (0.89 accuracy) but worse than RBF. Its decision boundary is curved but less smooth and can form unnecessary bends or wiggles, indicating mild overfitting or suboptimal alignment with the data's true structure.

Thus, the RBF kernel captures the shape of the Moons dataset more naturally, providing both high accuracy and visually accurate decision regions.

Banknote Dataset Questions

1.In this case, which kernel appears to be the most effective?

The RBF kernel is the most effective for the Banknote dataset, achieving an accuracy of 0.93, outperforming both the Linear (0.88) and Polynomial (0.84) kernels.

Its high precision and recall (around 0.90 - 0.96 for both classes) indicate strong generalization and separation between "Forged" and "Genuine" notes.
Although the Linear kernel already performs well suggesting partial linear separability - the RBF kernel's ability to model slightly non-linear patterns allows it to capture subtle feature interactions, improving overall accuracy and reducing misclassification near boundary regions.

2.The Polynomial kernel shows lower performance here compared to the Moons dataset. What might be the reason for this?

The Polynomial kernel performs worse here (accuracy 0.84) because the Banknote dataset is mostly linearly separable and does not require a complex, high-degree polynomial transformation.

When applied to already well-separated data, polynomial kernels can introduce unnecessary complexity, causing the model to overfit training variations rather than capture useful distinctions.
In contrast, the Moons dataset benefits from curvature modeling (non-linear structure), while the Banknote dataset penalizes it, leading to reduced generalization performance for the Polynomial kernel.

Hard vs. Soft Margin Questions

1. Compare the two plots. Which model, the "Soft Margin" (C=0.1) or the "Hard Margin" (C=100), produces a wider margin?

The Soft Margin model (C=0.1) produces a noticeably wider margin.
This is because a smaller C value relaxes the penalty on misclassification, allowing the optimizer to maximize the margin even if it means some training points fall within or across the boundary.

The Hard Margin model (C=100) enforces stricter separation, reducing margin width in an attempt to perfectly classify all training points.

2.Look closely at the "Soft Margin" (C=0.1) plot. You'll notice some points are either inside the margin or on the wrong side of the decision boundary. Why does the SVM allow these "mistakes"? What is the primary goal of this model?

The Soft Margin SVM allows such “mistakes” because its goal is to generalize better, not to perfectly classify every training sample. By permitting a few violations (slack variables), the model avoids overfitting to noise or outliers. The parameter C controls this trade-off: a smaller C reduces the penalty for errors, letting the model prioritize a smoother, more robust decision boundary that performs well on unseen data.

3.Which of these two models do you think is more likely to be overfitting to the training data? Explain your reasoning.

The Hard Margin model (C=100) is more prone to overfitting.It strictly enforces correct classification of all training points, including outliers or noisy samples, leading to a very tight boundary that may not generalize well to new data.

The Soft Margin model (C=0.1) trades a few training errors for smoother separation, which typically improves generalization.

4. Imagine you receive a new, unseen data point. Which model do you trust more to classify it correctly? Why? In a real-world scenario where data is often noisy, which value of C (low or high) would you generally prefer to start with?

The Soft Margin model (with a lower C value) is generally more trustworthy for classifying new, unseen data.
Its broader margin and tolerance for minor training errors makes it more flexible to data variability and noise.
In real-world scenarios, where data imperfections are common, starting with a smaller C (e.g., 0.1 or 1.0) helps the model avoid overfitting and often yields better real-world accuracy.

SCREENSHOTS

Training Results

1. Moon Dataset

- Classification Report for SVM with LINEAR Kernel

SVM with LINEAR Kernel PES2UG23CS926_C

	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

- Classification Report for SVM with RBF Kernel

```
SVM with RBF Kernel PES2UG23CS926_C
              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
```

- Classification Report for SVM with POLY Kernel

```
SVM with POLY Kernel PES2UG23CS926_C
              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
```

2. Banknote Dataset

- Classification Report for SVM with LINEAR Kernel

SVM with LINEAR Kernel PES2UG23CS926

	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

- Classification Report for SVM with RBF Kernel

SVM with RBF Kernel PES2UG23CS926

	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

- Classification Report for SVM with POLY Kernel

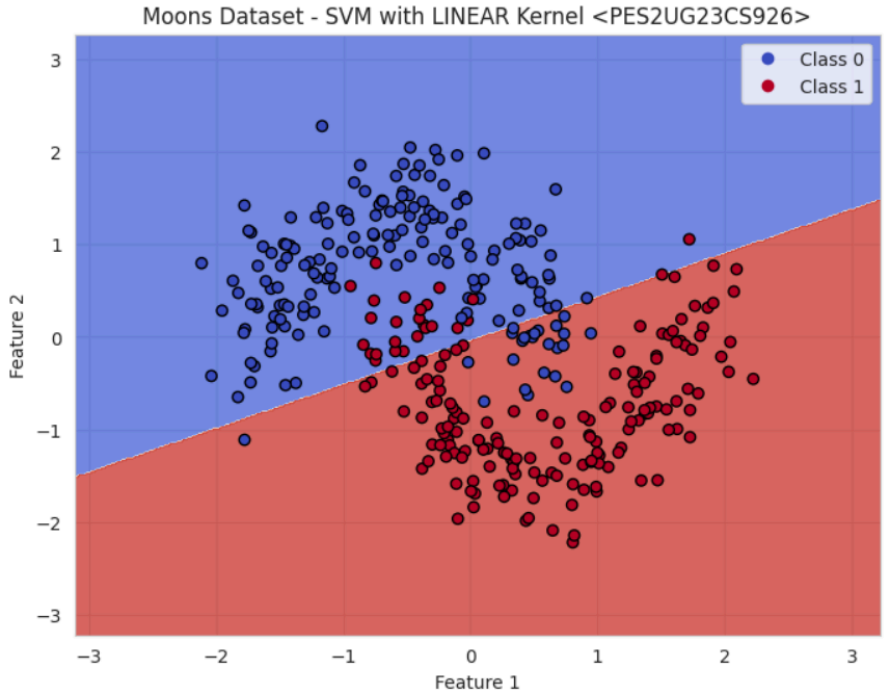
SVM with POLY Kernel PES2UG23CS926

	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

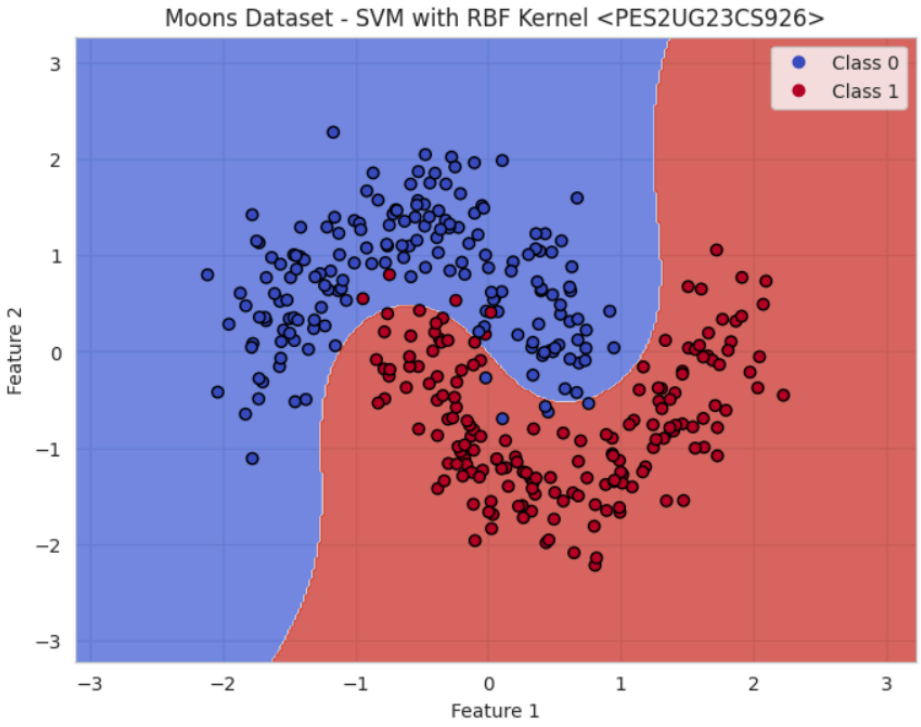
Decision Boundary Visualizations

1.Moons Dataset

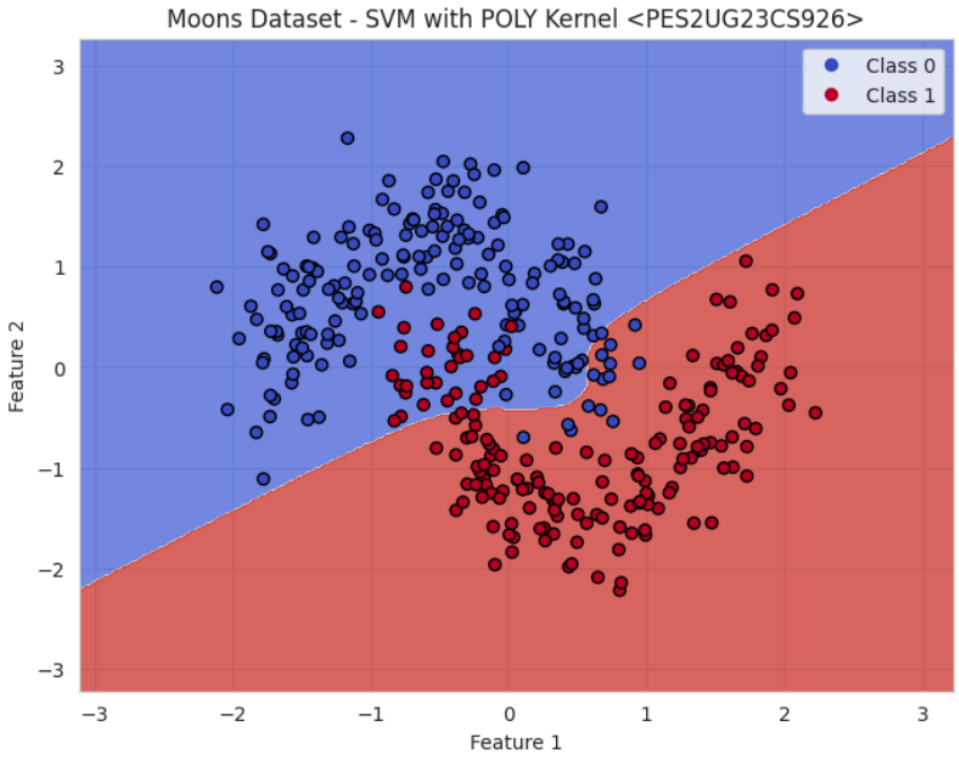
- Moons Dataset - SVM with LINEAR Kernel



- Moons Dataset - SVM with RBF Kernel

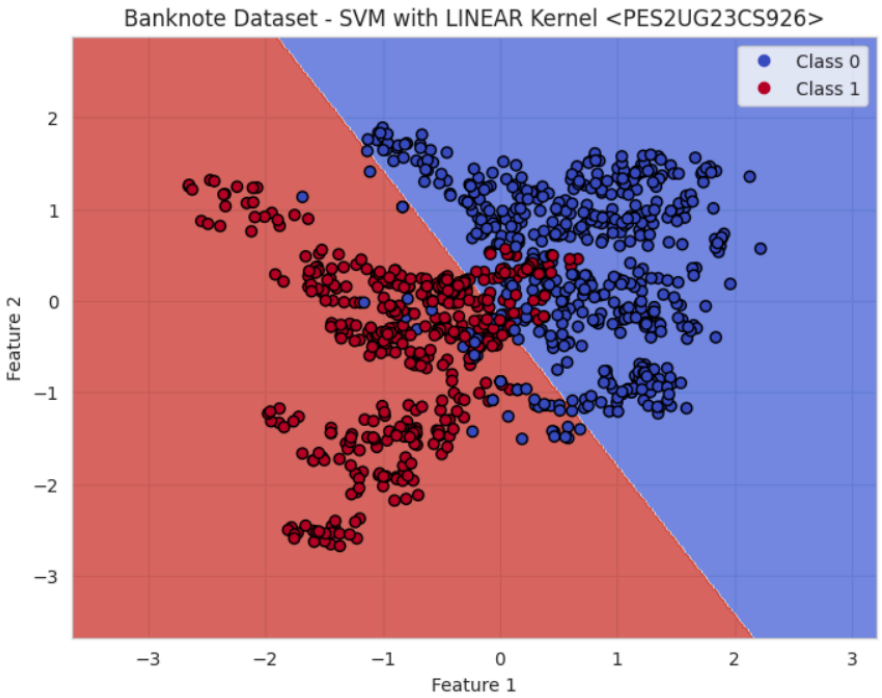


- Moons Dataset - SVM with POLY Kernel

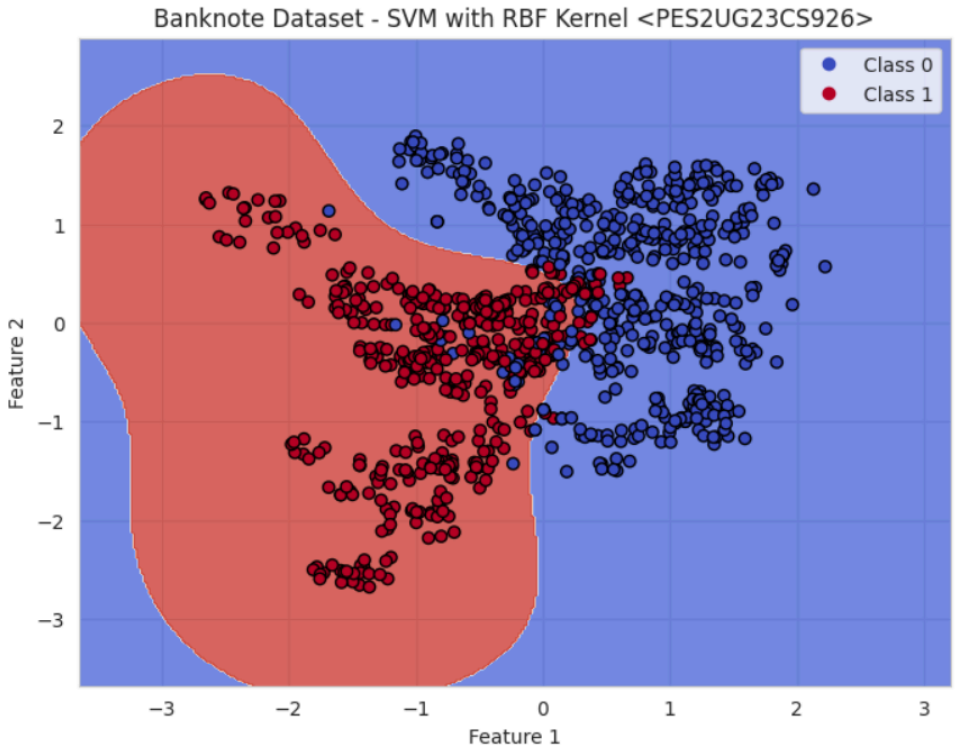


2.Banknote Dataset

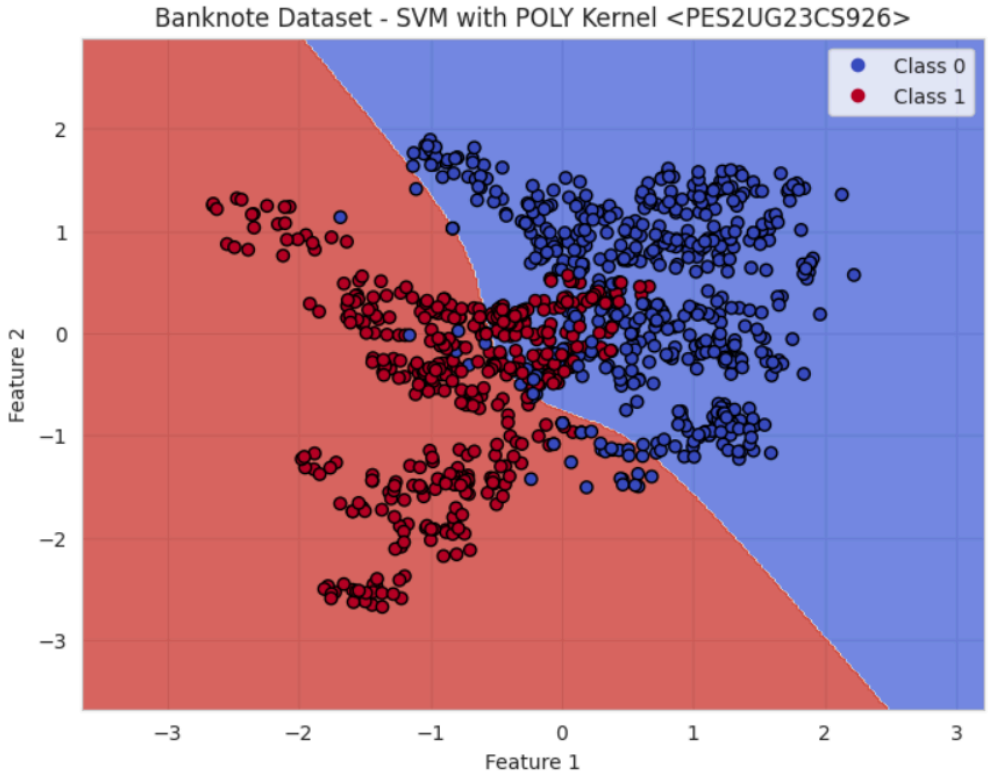
- Banknote Dataset - SVM with LINEAR Kernel



- Banknote Dataset - SVM with RBF Kernel

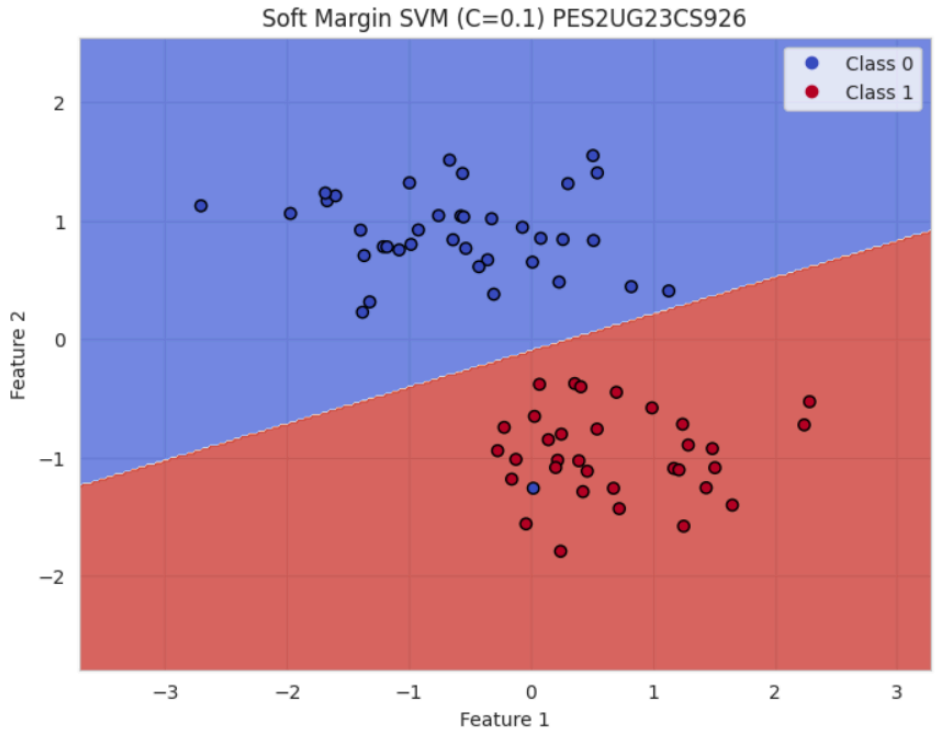


- Banknote Dataset - SVM with POLY Kernel



3. Margin Analysis

- Soft Margin SVM (C=0.1)



- Hard Margin SVM (C=100)