Lab Title: ML Lab Week 10 – SVM Classifier Lab

- Full Name: Neema Shrivastava
- SRN: PES2UG23CS377
- Section: 5F
- Date of Submission: 12/10/2025

1. Moons Dataset Questions

Q1: Inferences about the Linear Kernel's performance
The Linear Kernel struggles to perform well on the Moons dataset because the dataset is inherently non-linear, with interlocking half-moon shapes. The linear decision boundary is unable to separate the two classes effectively, resulting in lower accuracy and poorer classification metrics compared to non-linear kernels. The straight-line separation imposed by the linear kernel does not capture the curvature of the data, leading to misclassifications along the overlapping regions of the two half-moons.

Q2: Comparison between RBF and Polynomial kernel decision boundaries
The RBF kernel is highly effective in capturing the natural shape of the Moons dataset. Its non-linear transformation allows it to create smooth, curved decision boundaries that follow the contours of the half-moon shapes, leading to higher accuracy and fewer misclassifications. The Polynomial kernel also generates a curved boundary, but depending on the degree, it may overfit certain regions of the data or fail to generalize smoothly. Overall, the RBF kernel provides a more natural and generalizable separation of the classes in this dataset.

2. Banknote Dataset Questions

Q1: Which kernel was most effective for this dataset?
For the Banknote Authentication dataset, the Linear and RBF kernels are the most effective. The features used for visualization, variance and skewness, are close to linearly separable, and both kernels produce a clear separation between genuine and forged notes. The Polynomial kernel, however, underperforms as it attempts to fit a more complex boundary than necessary, which does not provide additional classification benefits for this relatively simple feature space.

Q2: Why might the Polynomial kernel have underperformed here?
The Polynomial kernel underperforms on this dataset because the real-world data contains noise and simpler patterns that do not require a high-degree polynomial for separation. By attempting to fit a complex curve, the model can overfit small variations and noise in the data, leading to poorer

generalization on the test set. In this case, the simpler decision boundary produced by the Linear or RBF kernel is sufficient for effective classification.

3. Hard vs. Soft Margin Questions

Q1: Which margin (soft or hard) is wider?
The soft margin SVM with a small C value (C=0.1) produces a wider margin compared to the hard margin SVM with a large C value (C=100). The soft margin allows some points to be misclassified or lie within the margin, which increases the distance between the classes while still maintaining reasonable classification performance.

Q2: Why does the soft margin model allow "mistakes"?
The soft margin model allows certain misclassifications to occur in order to achieve better generalization. Its primary goal is not to classify every training point correctly but to find a balance between maximizing the margin and minimizing classification errors. This tolerance for mistakes helps the model avoid overfitting to outliers or noisy data points.

Q3: Which model is more likely to be overfitting and why?
The hard margin SVM with a large C value is more likely to overfit the training data. Since it tries to classify every point correctly, including outliers, it produces a narrower margin that is highly sensitive to noise. While the training accuracy may be high, this approach does not generalize well to new, unseen data.

Q4: Which model would you trust more for new data and why?
For new and unseen data, the soft margin SVM is more trustworthy. Its wider margin and tolerance for small errors allow it to generalize better, especially in real-world scenarios where noise and overlapping classes are common. In practice, starting with a smaller C value provides a safer approach for robust classification, reducing the risk of overfitting while maintaining reasonable accuracy.

4. Screenshots (14 Total)

Training Results (6 Screenshots)

1,2,3 - Moons – SVM with LINEAR Kernel, SVM with RBF Kernel, SVM with POLY Kernel

```
SVM with LINEAR Kernel - PES2UG23CS377
              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 - PES2UG23CS377
              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 - PES2UG23CS377
              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

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

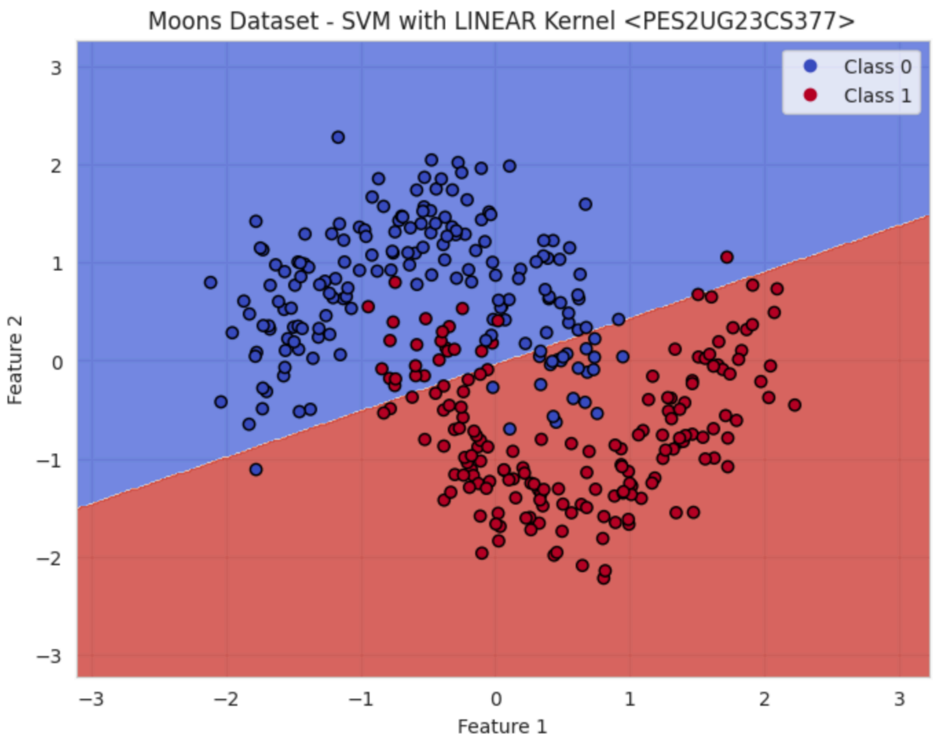
4,5,6. Banknote – SVM with LINEAR Kernel, SVM with RBF Kernel, SVM with POLY Kernel

```
SVM with LINEAR Kernel - PES2UG23CS377
              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 - PES2UG23CS377
              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 - PES2UG23CS377
              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

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

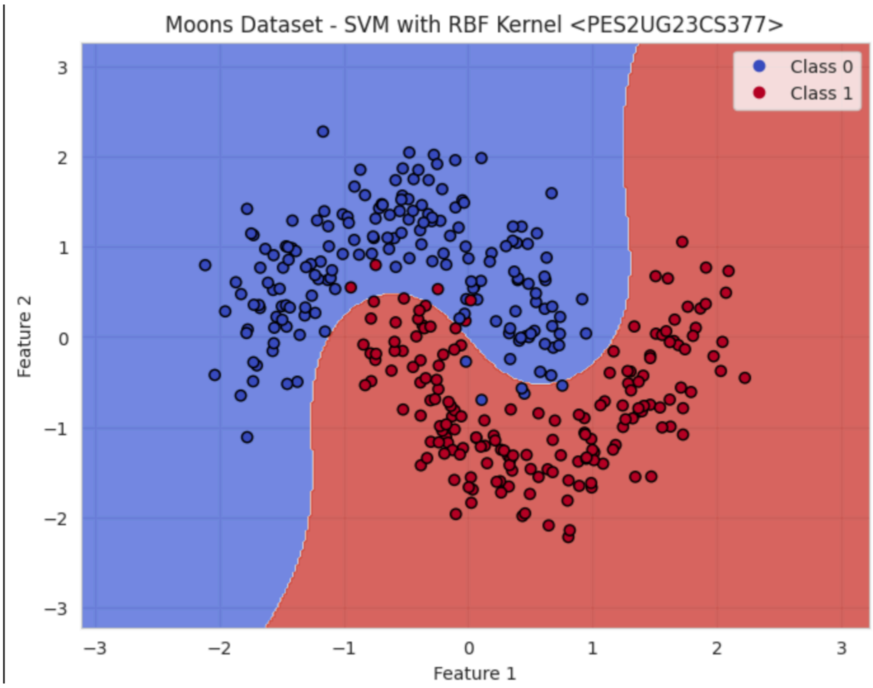
Decision Boundary Visualizations (8 Screenshots)

- Moons Dataset (3 plots)

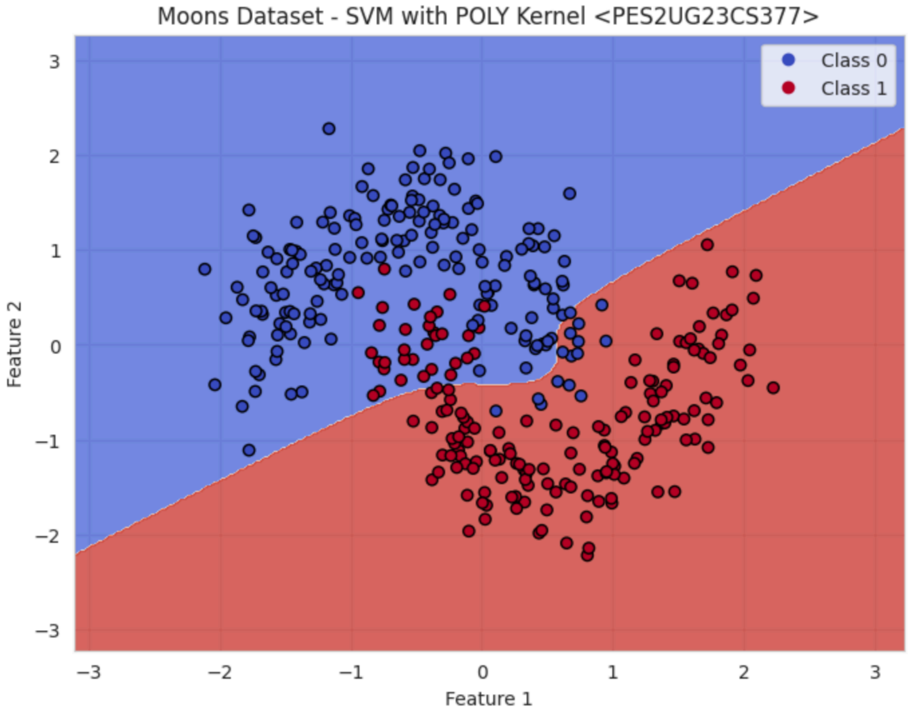
7. SVM with LINEAR Kernel



8. SVM with RBF Kernel

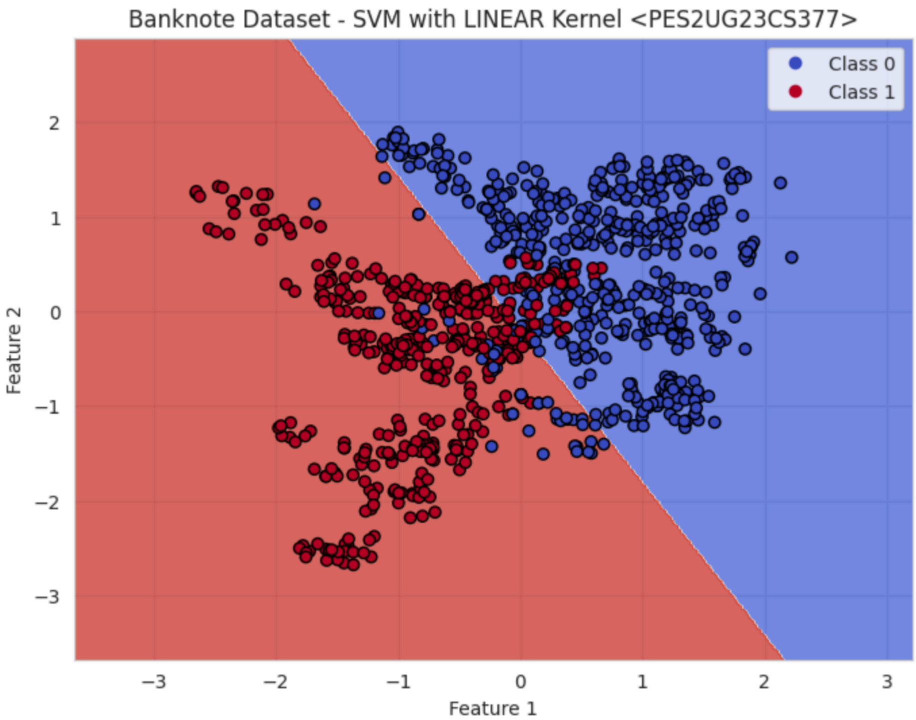


9. SVM with POLY Kernel

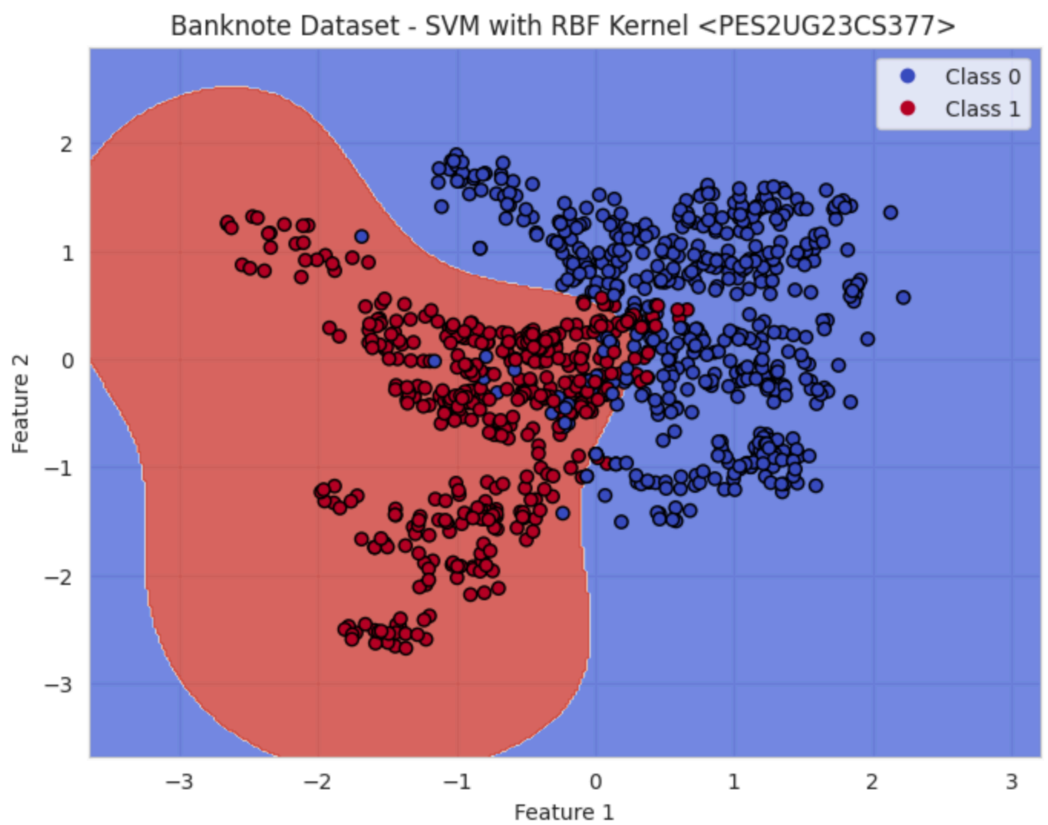


- Banknote Dataset (3 plots)

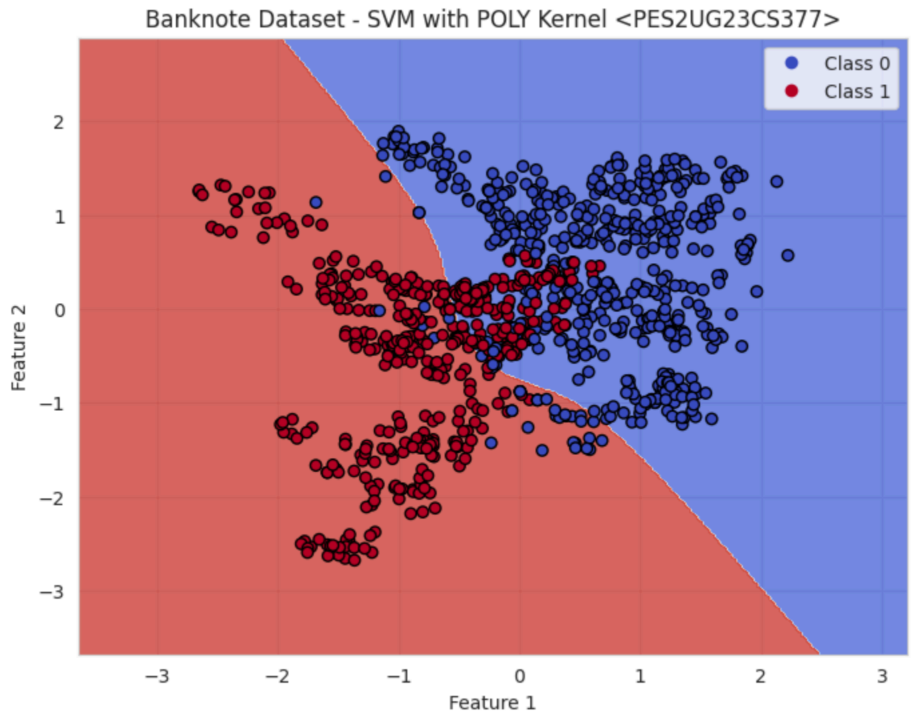
10. SVM with LINEAR Kernel



11. SVM with RBF Kernel

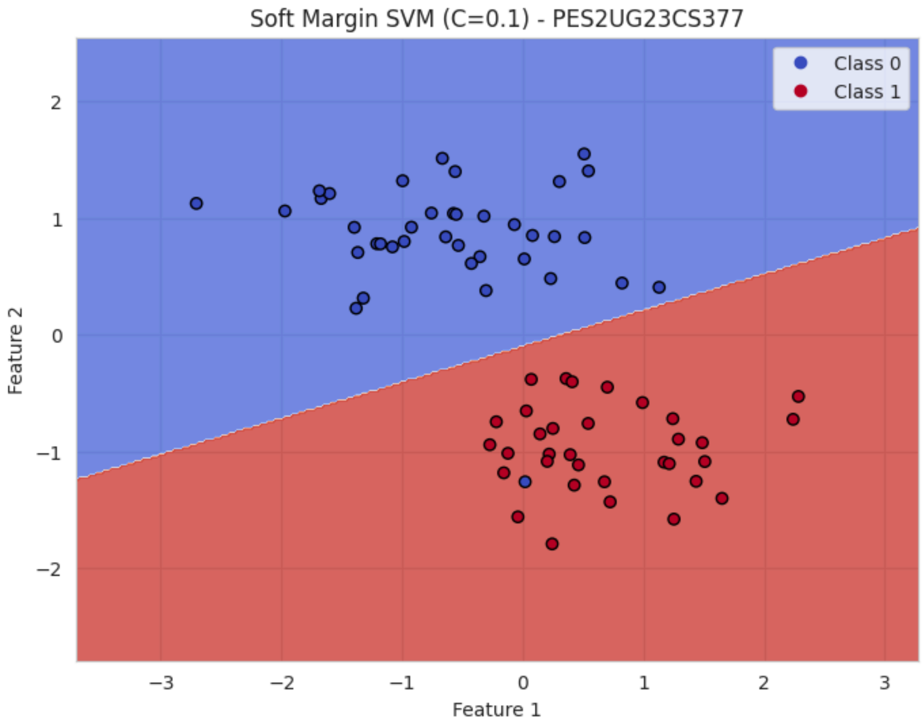


12. SVM with POLY Kernel



- Margin Analysis (2 plots)

13. Soft Margin SVM (C=0.1)



14. Hard Margin SVM (C=100)