Machine Learning Lab

Week-10

SVM lab

Name: Neha Harish

SRN: PES2UG23CS378

Section: F

Analysis Questions:

PART-1

1. The Linear Kernel achieved an accuracy of 0.87, which indicates reasonably good performance but not optimal for the non-linear moons dataset. Visually, its decision boundary appears as a straight line, failing to follow the curved separation of the two interleaving classes. This shows that the linear model cannot capture the inherent non-linearity in this dataset, resulting in some misclassified points near the overlap.
2. The RBF Kernel (accuracy = 0.97) captures the curved pattern of the moons dataset much more naturally. Its boundary wraps closely around each cluster, providing smooth and flexible separation. In contrast, the Polynomial Kernel (accuracy = 0.89) forms a more rigid, global curve that doesn't perfectly adapt to local bends in the data, it sometimes overfits in one region and underfits in another. Hence, RBF clearly models the moon-shaped clusters more effectively.

PART-2

1. The RBF kernel is clearly the most effective for the Banknote dataset, achieving the highest accuracy of 93%.
It forms flexible, non-linear decision boundaries that adapt to the subtle variations in the feature space. This allows it to correctly classify both Forged and Genuine notes even when the data is not perfectly linearly separable. The Linear kernel performs moderately well (88%), while the Polynomial kernel underperforms at 84%.

2. The Banknote dataset is largely linearly or mildly non-linear, unlike the Moons dataset which has strongly curved, circular class boundaries. The Polynomial kernel introduces complex, high-degree interactions that can overfit in such nearly-linear data, producing unnecessarily intricate boundaries that do not generalize well. In contrast, in the Moons dataset those polynomial curves helped fit the curved pattern slightly better. Hence, the polynomial mapping becomes too rigid and high-variance for the Banknote dataset, leading to reduced performance.

PART-3

1. When comparing the two plots, the Soft Margin model (C=0.1) produces a wider margin than the Hard Margin model (C=100). This is because a lower C value allows the SVM to focus more on maximizing the margin rather than perfectly classifying every point. As a result, the decision boundary in the Soft Margin plot appears smoother and more general, while the Hard Margin model forces the boundary to tightly fit around all training points, resulting in a narrower margin.

2. In the Soft Margin (C=0.1) plot, some points lie inside the margin or on the wrong side of the boundary because the model is designed to tolerate a few misclassifications. These "mistakes" are intentional, as the SVM uses slack variables to allow flexibility in order to create a simpler, more general boundary. The primary goal of this model is not perfect accuracy on the training set but achieving a balance that improves performance on unseen data.

3. Between the two, the Hard Margin model (C=100) is more likely to overfit the training data. By trying to classify every training point correctly, including noise or

outliers, it creates a more complex boundary that doesn't generalize well. The Soft Margin model, in contrast, avoids overfitting by allowing minor errors, which leads to better stability when new or slightly different data is introduced.

4. If a new, unseen data point were introduced, the Soft Margin model (C=0.1) would be more trustworthy for classification. It is less sensitive to noise and small variations in the data, which makes it better suited for real-world conditions. In most practical scenarios where datasets contain noise or overlapping classes, it is generally advisable to start with a low C value to ensure the model generalizes well rather than memorizing the training set.

Screenshots:

Moon Dataset-

1. Classification Report for SVM with LINEAR Kernel with SRN

```
SVM with LINEAR Kernel PES2UG23CS378
              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 PES2UG23CS378
              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 PES2UG23CS378
              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

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

Banknote Dataset-

1. Classification Report for SVM with LINEAR Kernel

```
SVM with LINEAR Kernel PES2UG23CS378
              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 PES2UG23CS378
              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

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

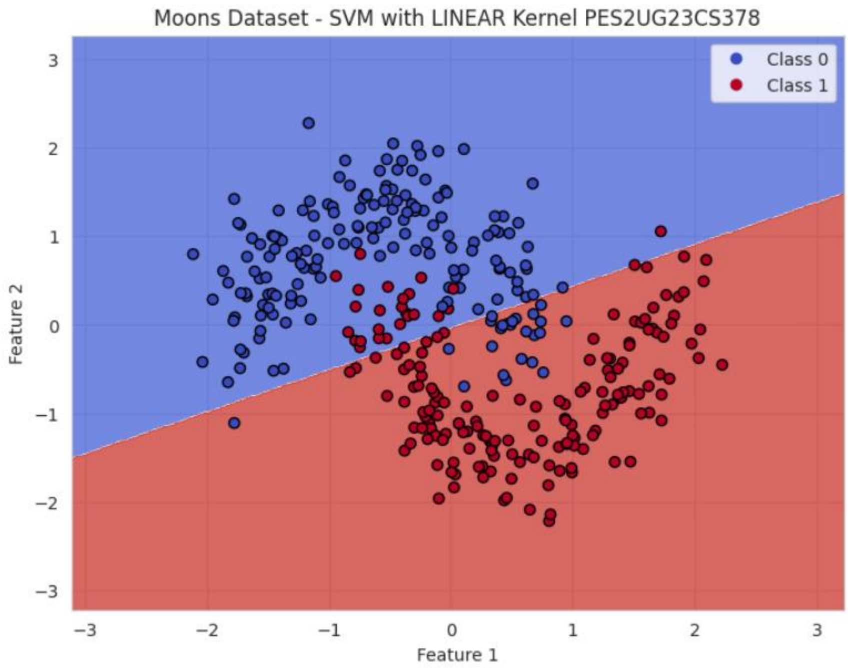
3. Classification Report for SVM with POLY Kernel

```
SVM with POLY Kernel PES2UG23CS378
              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

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

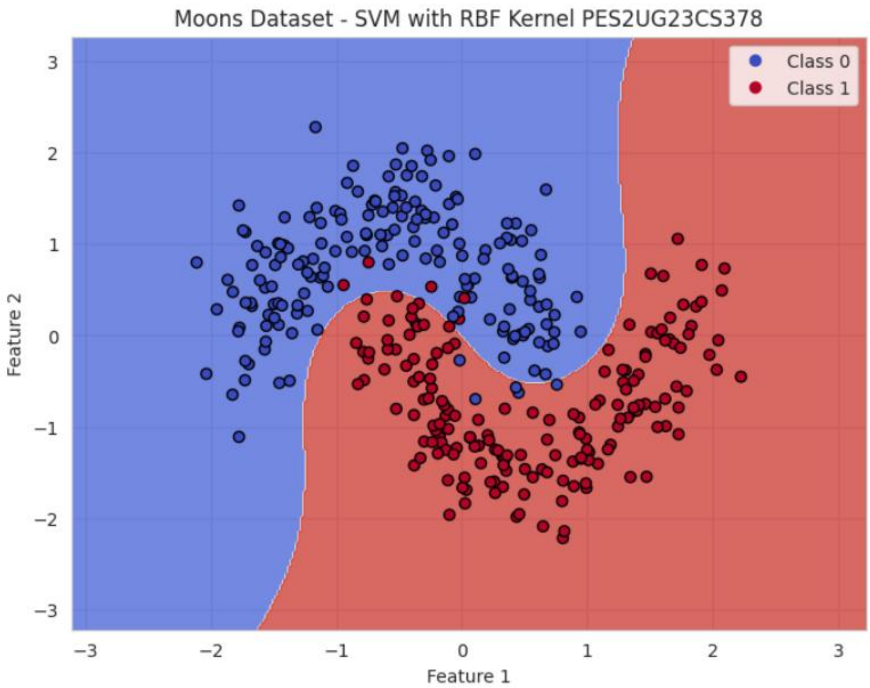
Decision Boundary Visualizations:

Moons Dataset:

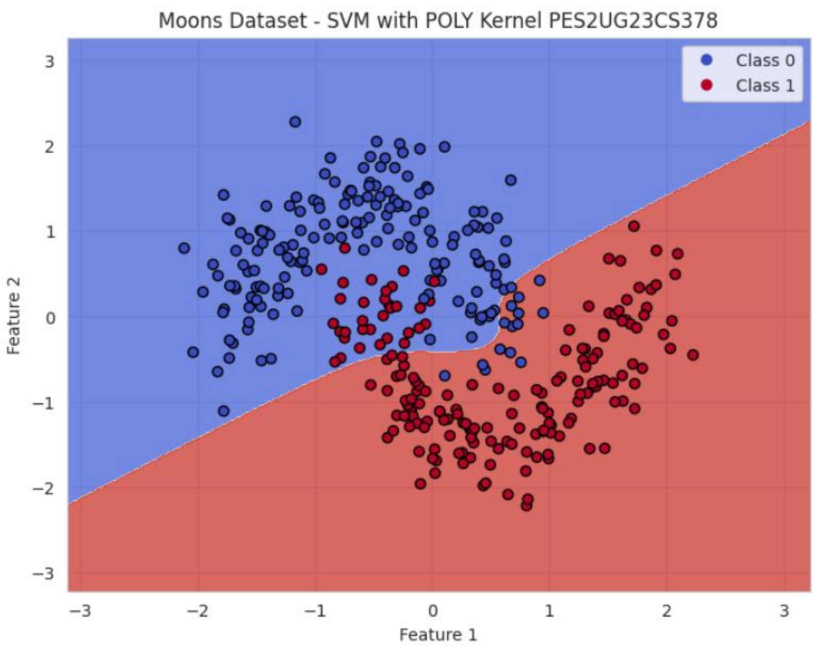
1. Moons Dataset - SVM with LINEAR Kernel



2. Moons Dataset - SVM with RBF Kernel

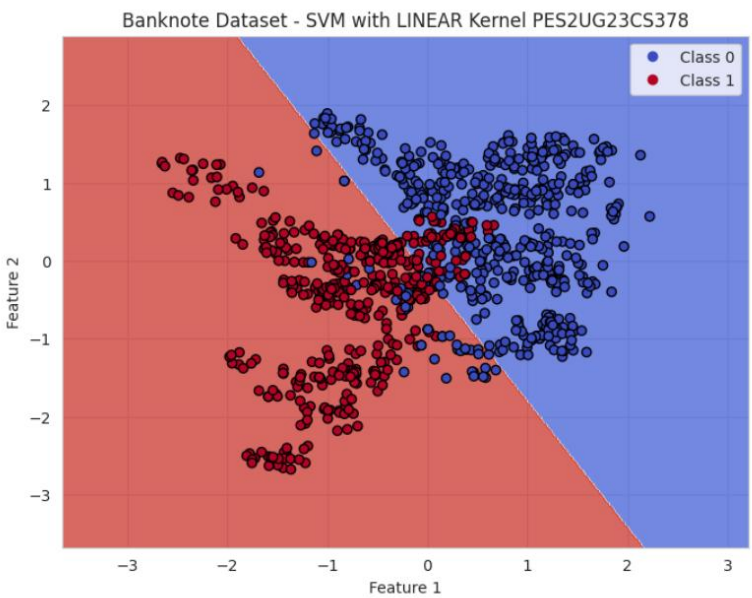


3. Moons Dataset - SVM with POLY Kernel

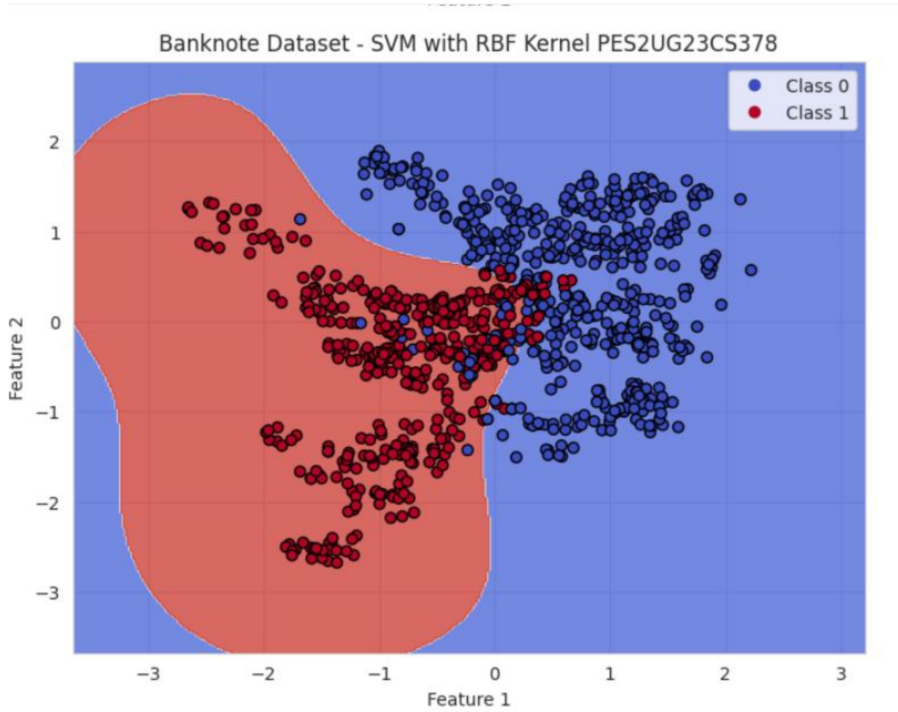


Banknote Dataset:

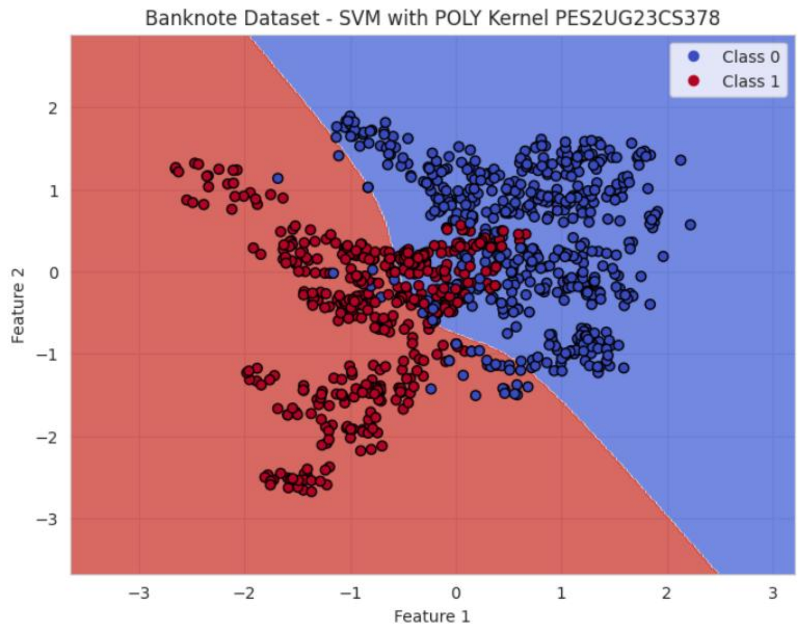
1. Banknote Dataset - SVM with LINEAR Kernel



2. Banknote Dataset - SVM with RBF Kernel

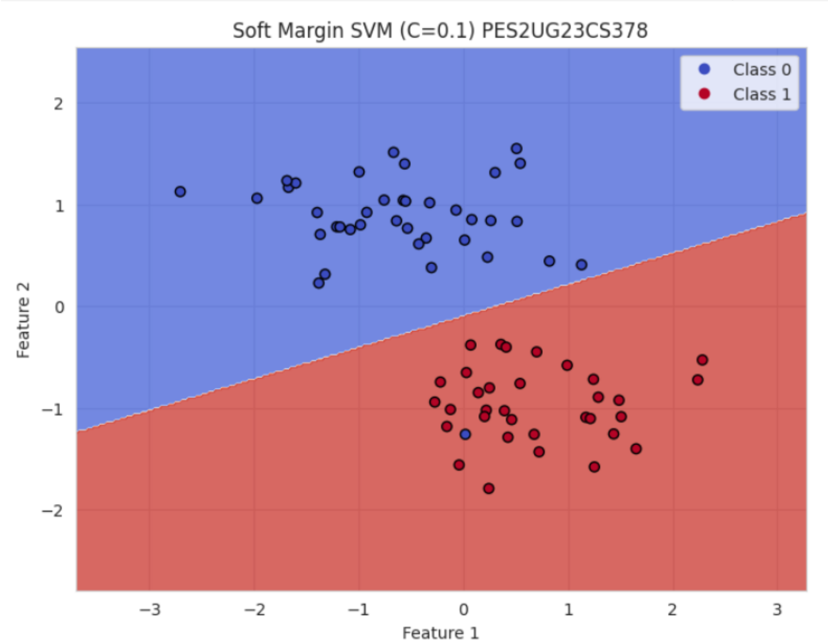


3. Banknote Dataset - SVM with POLY Kernel



Margin Analysis:

1. Soft Margin SVM (C=0.1)



2. Hard Margin SVM (C=100)