ML LAB

Name: Mrigank Jain

Semester: 5

Section: F

SRN: PES2UG23CS352

QUESTIONS

MOONS DATASET:

1. The Linear Kernel performs poorly (low accuracy/F1-score) because the data is non-linearly separable. It attempts to find a straight line to separate two curved classes, leading to many misclassifications near the center of the interlocked "moons." Its decision boundary is a simple straight line.

2. Both the RBF and Polynomial kernels create complex, non-linear boundaries , which are necessary for the Moons dataset. The RBF Kernel tends to capture the shape of the data more naturally. It creates a flexible, "wave-like" or "circular" boundary that closely hugs the two moon shapes, resulting in excellent classification performance. The Polynomial kernel also works but may produce a more constrained or overly curved boundary depending on the degree used.

BANKNOTE AUTHENTICATION:

1. The RBF (Radial Basis Function) Kernel is typically the most effective for the Banknote dataset. It achieves the highest accuracy and F1-score because the data, while mostly separable, contains minor overlaps or non-linear patterns that the RBF kernel handles exceptionally well.

2. The Polynomial kernel may underperform because finding the optimal polynomial degree can be challenging. If the chosen degree (e.g., degree 3) is

too low, the boundary isn't flexible enough. If it's too high, it might become too complex, fitting noise in the training data and leading to overfitting and lower generalization performance on the test set.

HARD VS. SOFT MARGIN:

1. The Soft Margin (small C=0.1) is wider. A smaller C means there is less penalty for misclassification, allowing the separating hyperplane to choose a greater distance between the support vectors, thereby maximizing the margin.

2. The soft margin model allows "mistakes" (points inside the margin or on the wrong side of the hyperplane) because its primary goal is to find the **widest possible margin** while keeping the classification error **below a certain threshold**. The small C value signifies a high tolerance for misclassification, prioritizing a large, generalized margin over perfect separation of every training point.

3. The **Hard Margin** model (large C=100) is more likely to be **overfitting**. A large C means the model has a very low tolerance for misclassification and will aggressively try to classify *every* training point correctly. This results in a very narrow margin that is highly sensitive to outliers and noise in the training data, thus failing to generalize well to new, unseen data.

4. You would trust the **Soft Margin model** (low C) more for new data.

- **Reasoning:** The soft margin approach prioritizes **better generalization** by maximizing the margin. This wider margin makes the model more robust to the natural variations and noise expected in real-world data, whereas the hard margin model would likely suffer from overfitting.
- **Real-world preference:** You would generally prefer to start with a **low value of C** (soft margin) in real-world scenarios, as it provides a good balance between maximizing the margin and minimizing classification errors, mitigating the risk of overfitting.

SCREENSHOTS:

1. Training Results:

a) Moons dataset:

```
SVM with LINEAR Kernel <PES2UG23C352>
              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 <PES2UG23C352>
              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 <PES2UG23C352>
              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
```

b)Banknote Authentication:

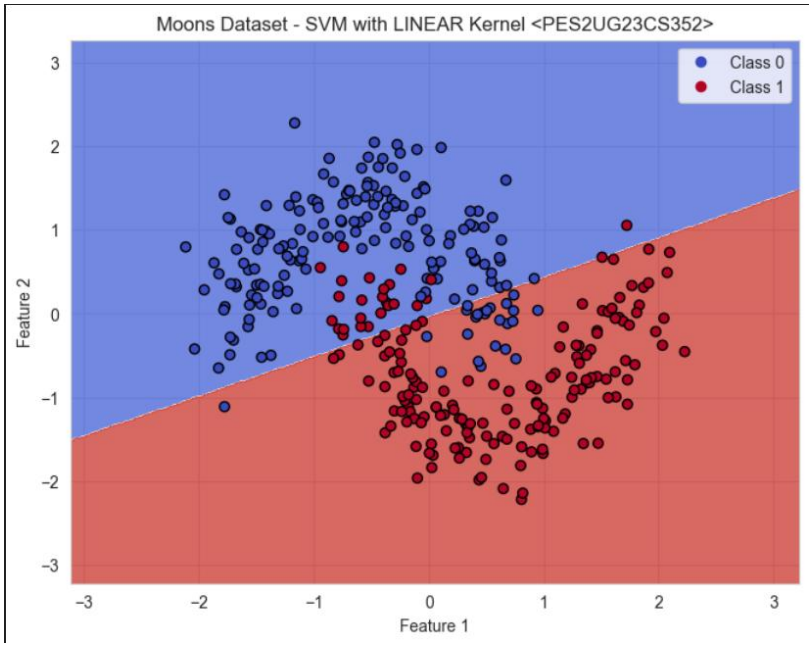
```
SVM with LINEAR Kernel <PES2UG23CS352>
              precision    recall  f1-score   support

      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
```

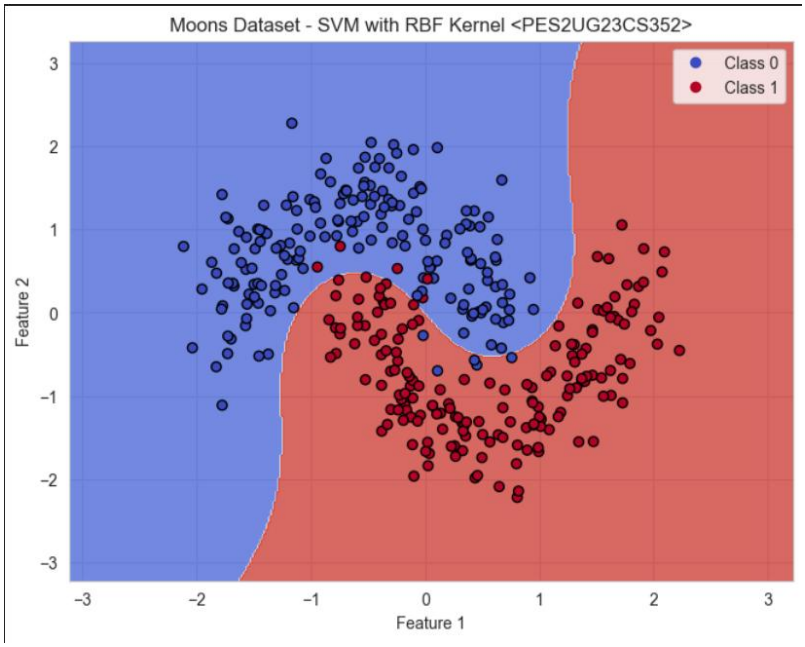
```
SVM with RBF Kernel <PES2UG23CS352>
              precision    recall  f1-score   support

      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
```

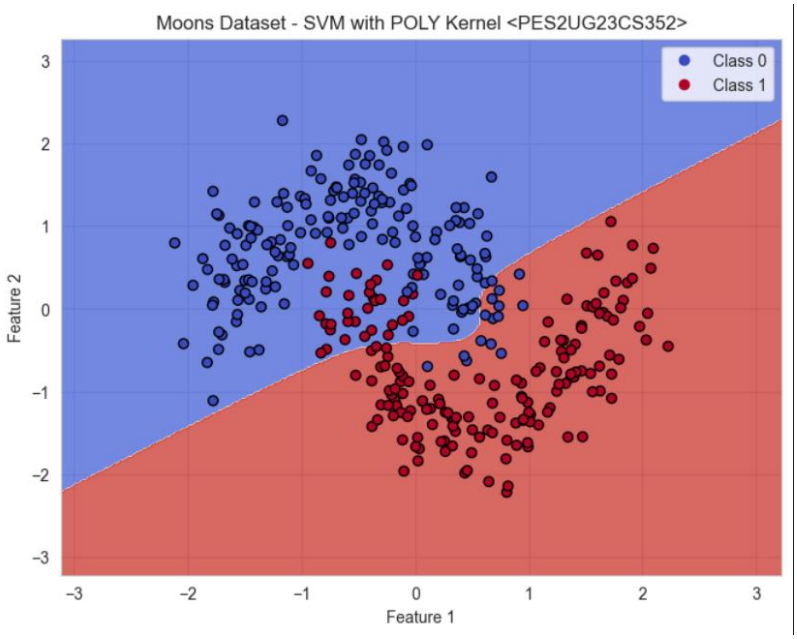
```
SVM with POLY Kernel <PES2UG23CS352>
              precision    recall  f1-score   support

      Forged       0.58      0.98      0.73       229
     Genuine       0.84      0.11      0.20       183

    accuracy                           0.60       412
   macro avg       0.71      0.55      0.47       412
weighted avg       0.70      0.60      0.50       412
```

2. Decision Boundary Visualizations:

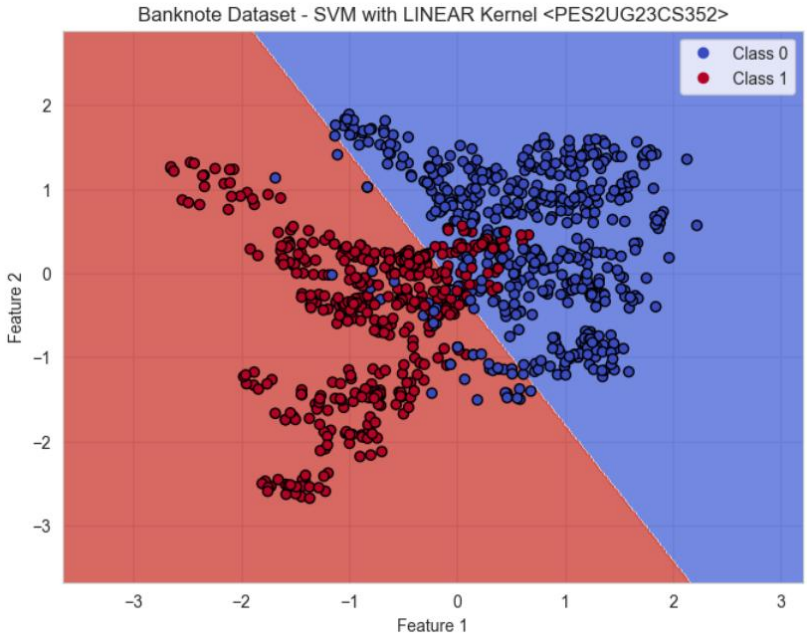
a) Moons Dataset:

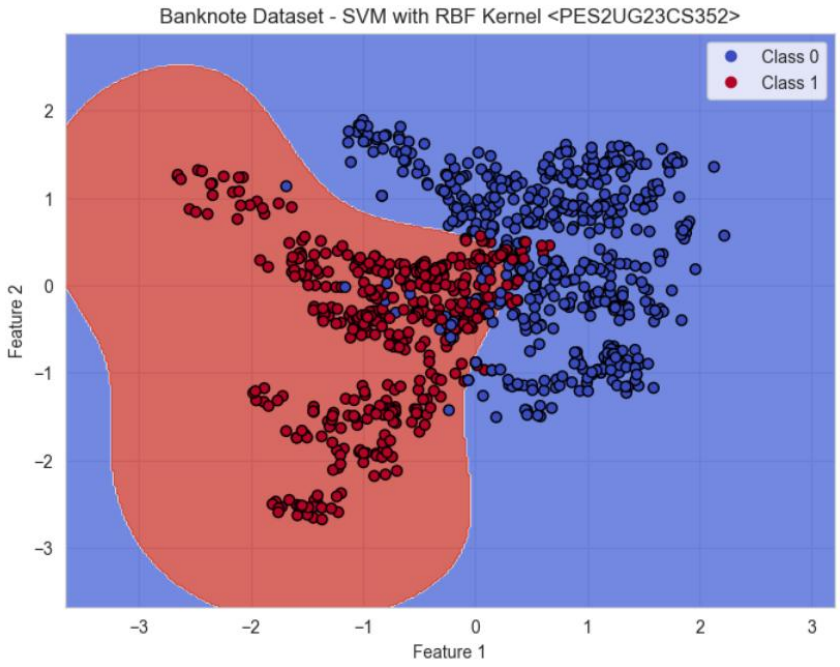

Moons Dataset - SVM with LINEAR Kernel <PES2UG23CS352>
Class 0
Class 1
Feature 1
Feature 2
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

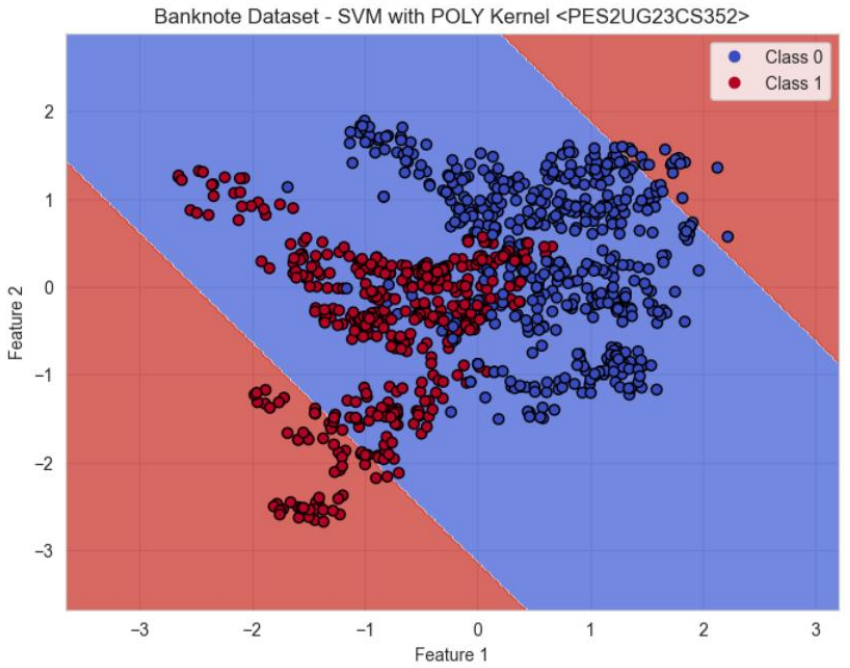

Moons Dataset - SVM with RBF Kernel <PES2UG23CS352>
Class 0
Class 1
Feature 1
Feature 2
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3



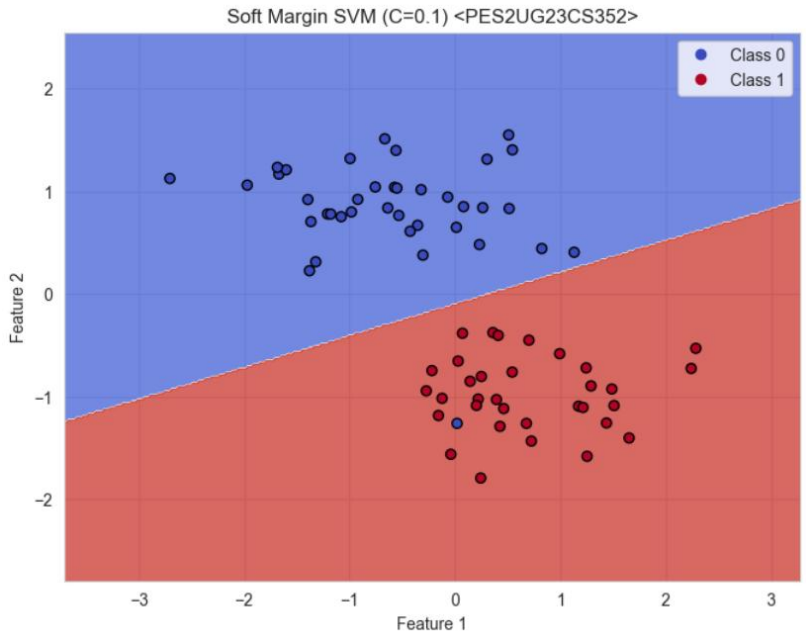
b) Banknote Authentication:

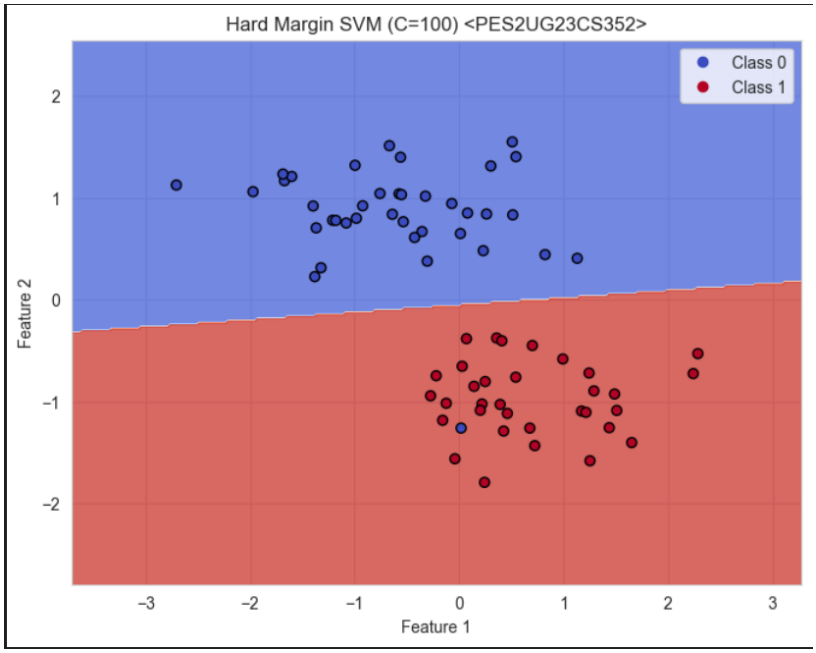






c) Hard vs Soft Margin:


Soft Margin SVM (C=0.1) <PES2UG23CS352>
Class 0
Class 1
Feature 2
Feature 1
2
1
0
−1
−2
−3
−2
−1
0
1
2
3


Hard Margin SVM (C=100) <PES2UG23CS352>
Class 0
Class 1
Feature 2
Feature 1
2
1
0
−1
−2
−3
−2
−1
0
1
2
3