ML_Lab_Week10_SVMLab

N RAKSHITHA
PES2UG23CS356
‘F’ section

1 **Inferences about the Linear Kernel's performance**
The linear kernel performed poorly on the Moons dataset because the data is non-linearly separable. A straight-line decision boundary cannot effectively separate the interleaving half-moon shapes, leading to lower accuracy and more misclassifications

2 **Comparison between RBF and Polynomial kernel decision boundaries**
Both RBF and Polynomial kernels captured the non-linear structure of the data better than the linear kernel. The RBF kernel typically produced smoother, more flexible boundaries, while the Polynomial kernel created more rigid, curved boundaries. RBF often generalizes better due to its localized influence

1. **Which kernel was most effective for this dataset?**
   The linear kernel was the most effective, achieving high accuracy and precision. This suggests that the banknote data is largely linearly separable in the variance-skewness feature space.
2. **Why might the Polynomial kernel have underperformed here?**
   The Polynomial kernel may have overcomplicated the decision boundary, introducing unnecessary complexity for a dataset that is already well-separated linearly. This can lead to overfitting and reduced generalization.

1. **Which margin (soft or hard) is wider?**
   The soft margin is wider. It allows a more relaxed boundary that tolerates some misclassifications to improve generalization.
2. **Why does the soft margin model allow "mistakes"?**
   The soft margin model (with small C) prioritizes a wider margin over perfect classification. It tolerates some misclassified points to avoid overfitting and handle noisy data better
3. **Which model is more likely to be overfitting and why?**
   The hard margin model (with large C) is more likely to overfit because it tries to classify every training point correctly, including outliers, which can lead to a very narrow and rigid decision boundary.
4. **Which model would you trust more for new data and why?**
   The soft margin model is more trustworthy for new data because it generalizes better by allowing flexibility in the decision boundary, making it more robust to noise and unseen variations.

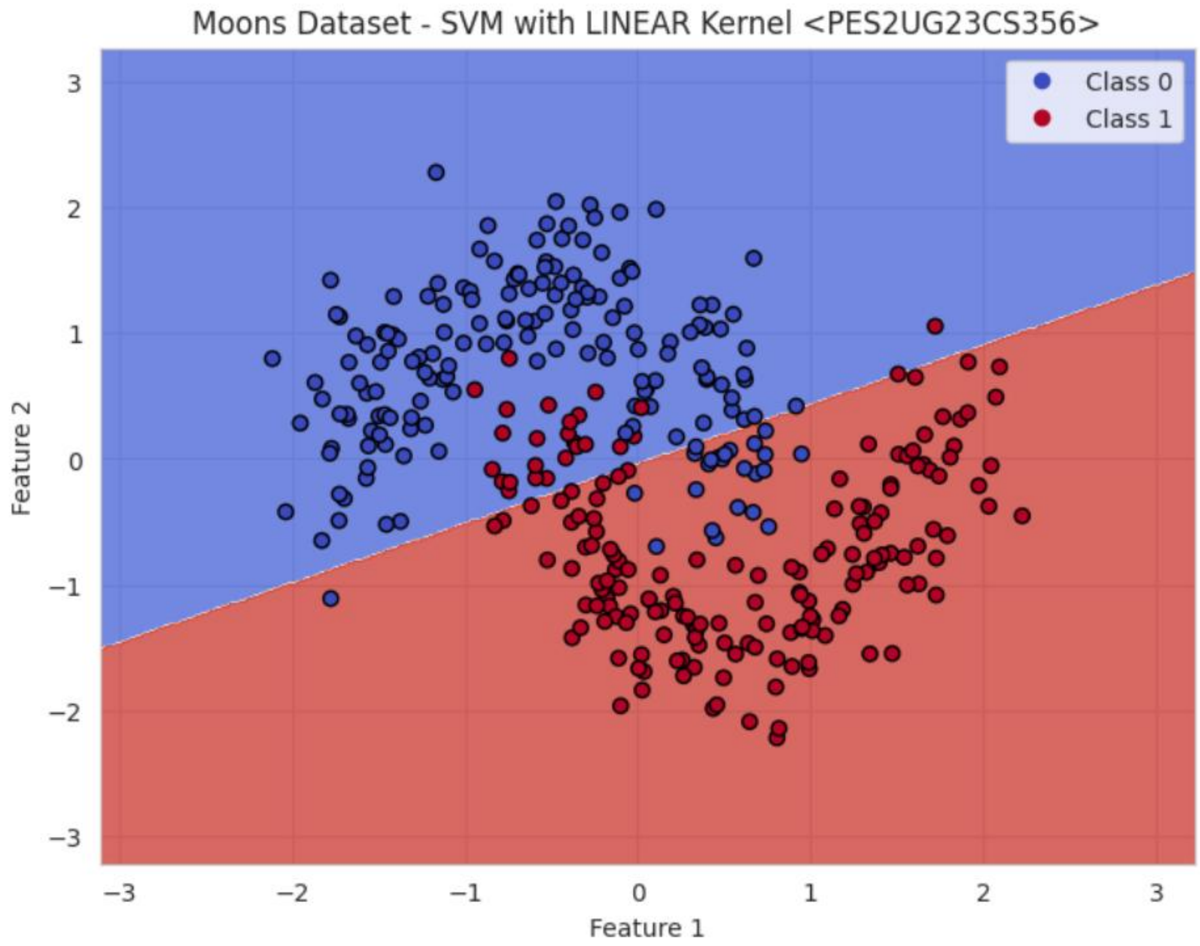
MOONS DATASET

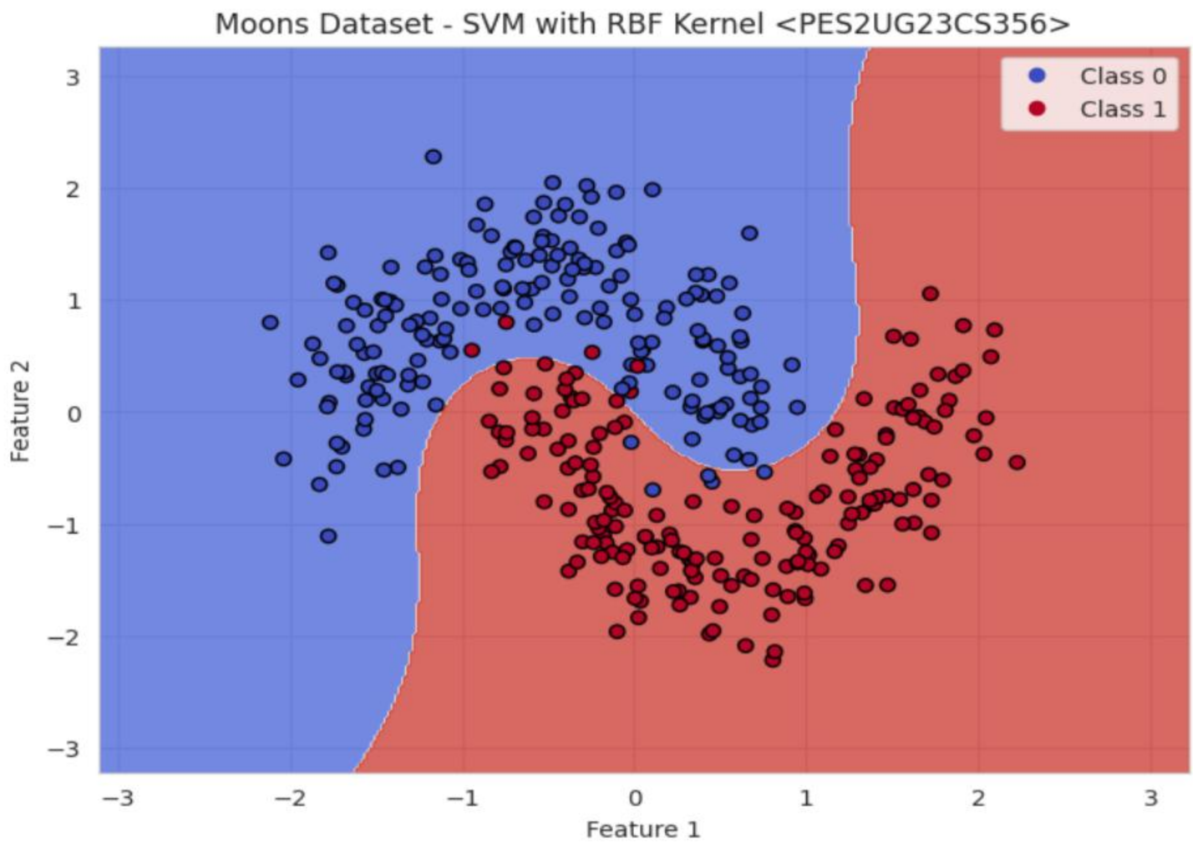
```
SVM with LINEAR Kernel <PES2UG23CS356>
              precision    recall  f1-score   support

           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 <PES2UG23CS356>
              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 <PES2UG23CS356>
              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
```

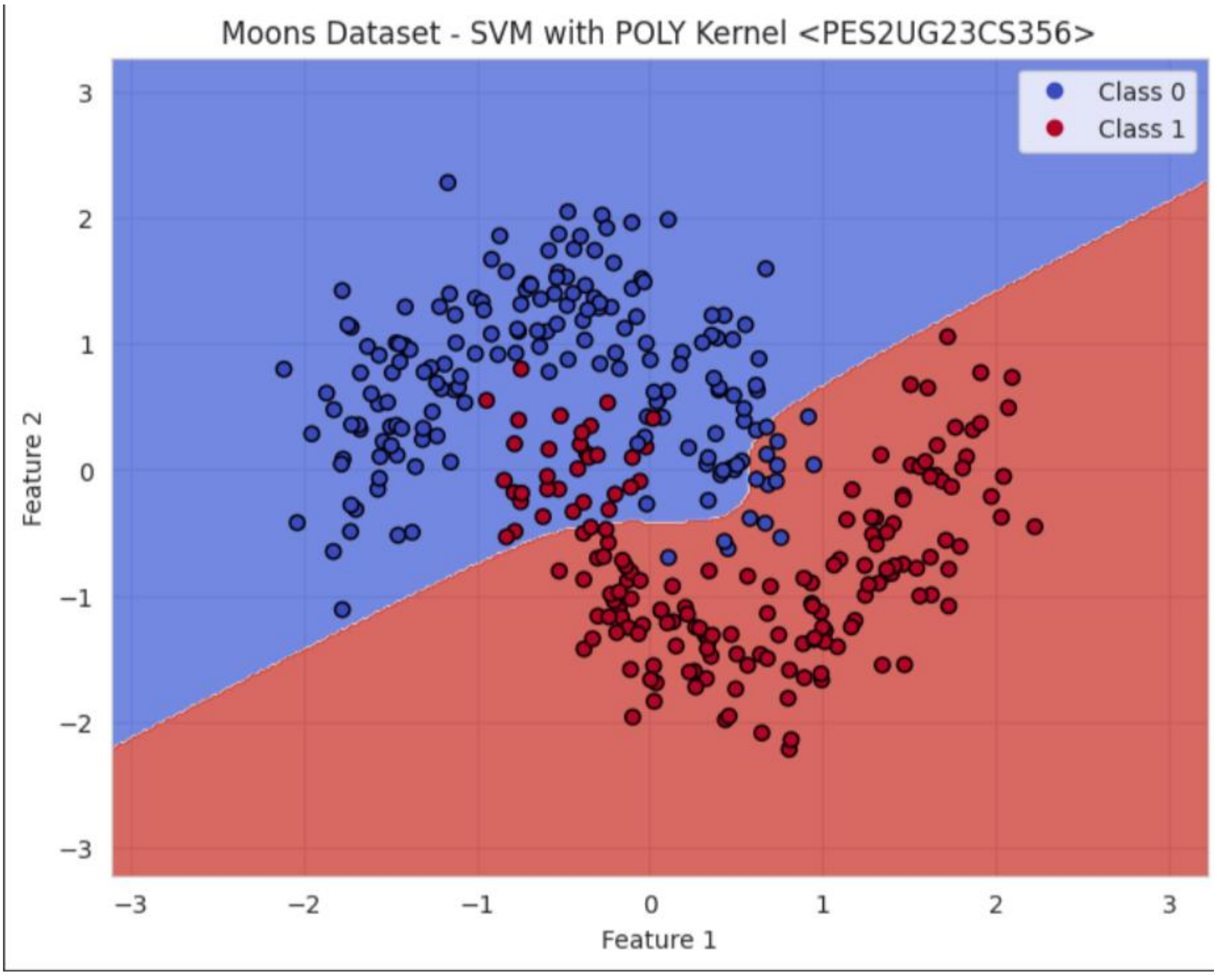
BANKNOTE DATASET

```
SVM with LINEAR Kernel <PES2UG23CS356>
              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 <PES2UG23CS356>
              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 <PES2UG23CS356>
              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
```

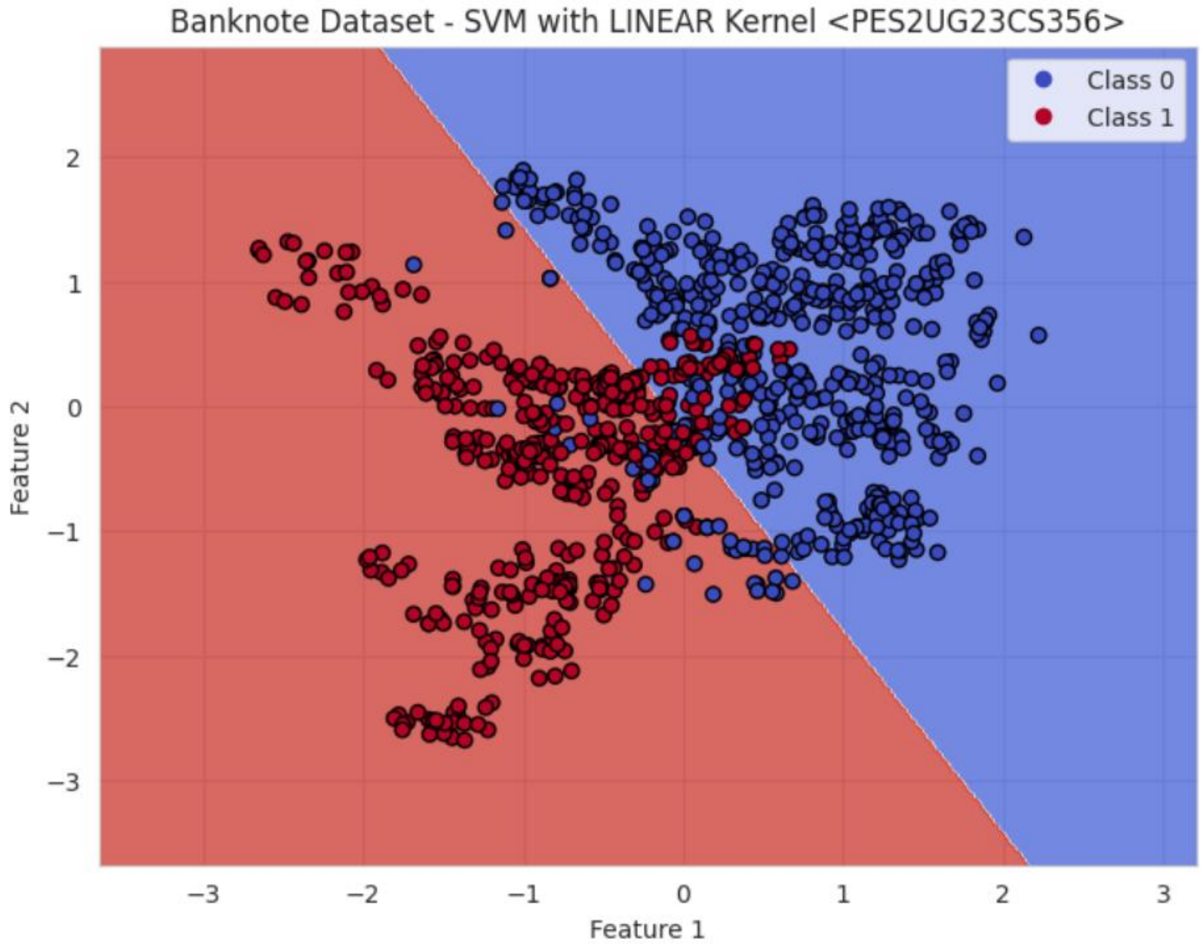
Decision Boundary Visualizations
MOONS DATASET

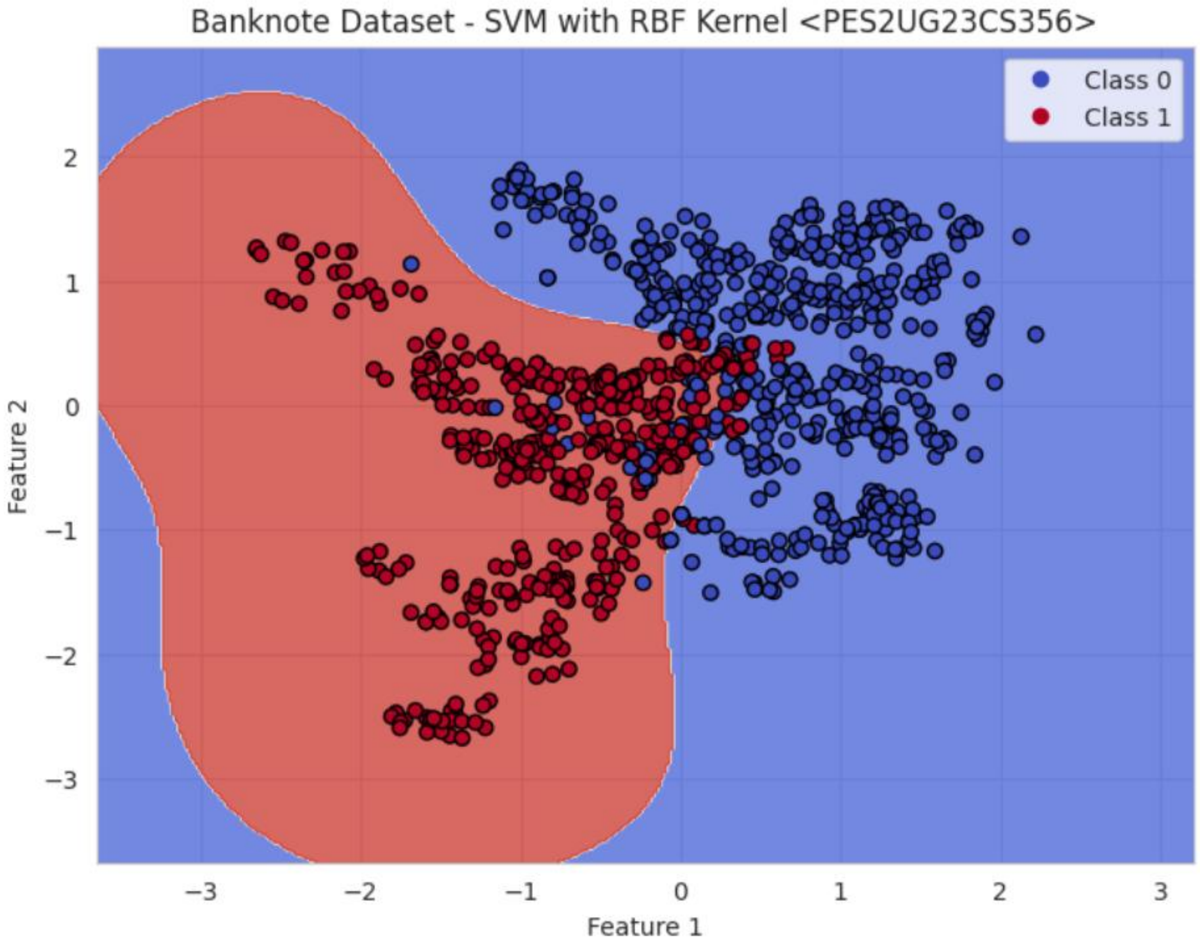


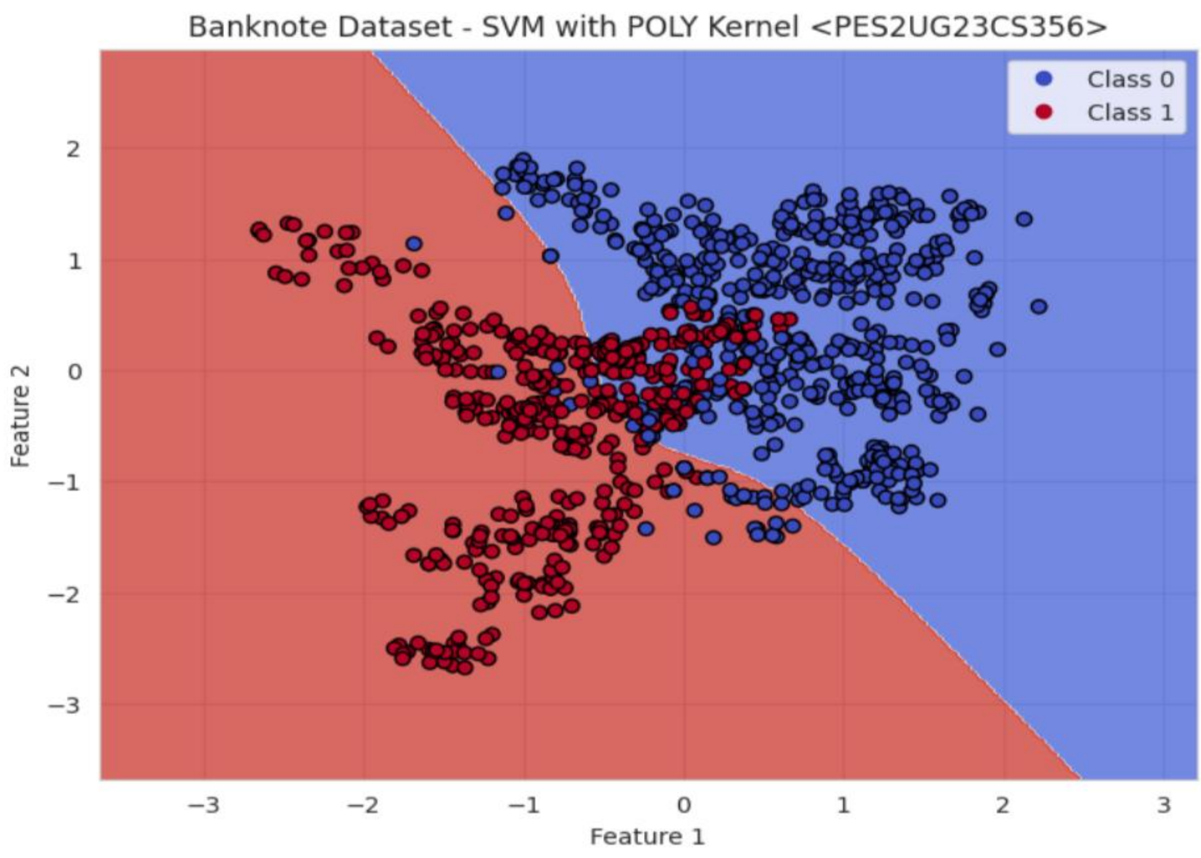

Moons Dataset - SVM with RBF Kernel <PES2UG23CS356>
Class 0
Class 1
Feature 2
Feature 1


Moons Dataset - SVM with POLY Kernel <PES2UG23CS356>
Class 0
Class 1
Feature 2
Feature 1

BANKNOTE DATASET




Banknote Dataset - SVM with RBF Kernel <PES2UG23CS356>
Class 0
Class 1
Feature 2
Feature 1
−3
−2
−1
0
1
2
3


Banknote Dataset - SVM with POLY Kernel <PES2UG23CS356>
Class 0
Class 1
Feature 2
Feature 1
−3
−2
−1
0
1
2
3

Margin Analysis