MACHINE LEARNING LAB - WEEK 10

Support Vector Machine (SVM) Classifier Lab

Name: NANDANI SONALE

SRN: PES2UG23CS364

Section: F

Course: MACHINE LEARNING LAB

1. Introduction

The purpose of this lab is to understand and implement Support Vector Machine (SVM) classifiers. We explored how different kernel functions—Linear, Radial Basis Function (RBF), and Polynomial—affect classification performance. We trained models on three datasets: a synthetic non-linear dataset (Moons), a real-world binary dataset (Banknote Authentication), and a linearly separable dataset with noise to study Soft and Hard margins. Each model's performance was evaluated using standard metrics and visualized through decision boundaries.

2. Dataset Description

• Moons Dataset – A synthetic, non-linearly separable dataset with interlocking half-moons used to demonstrate kernel effects. • Banknote Authentication Dataset – A real-world dataset used to classify whether a banknote is genuine or forged using image-based statistical features. • Linearly Separable Dataset with Noise – A simple two-class dataset with added outliers used to compare Soft and Hard margin SVMs.

3. Results and Analysis

For the Moons dataset, the Linear kernel showed underfitting with several misclassifications, while the RBF kernel achieved the best accuracy by capturing the curved data boundary. The Polynomial kernel performed moderately but introduced slight overfitting. In the Banknote dataset, the Linear kernel was the best performer, producing high accuracy and clean separation, as the data is almost linearly separable. The Polynomial kernel overfit slightly, and the RBF kernel performed comparably well but without added benefit. In the Hard vs Soft Margin experiment, the Soft Margin SVM (C=0.1) had a wider margin, tolerated minor misclassifications, and generalized better. The Hard Margin SVM (C=100) fit tightly to all points, resulting in overfitting, especially around noisy data points.

4. Key Learnings

• Feature scaling is essential for optimal SVM performance. • The choice of kernel should align with the data’s separability characteristics. • Soft Margin SVMs (small C) are preferred for noisy, real-world data due to better generalization. • Hard Margin SVMs (large C) can overfit when data contains outliers. • Visualization of decision boundaries is crucial for interpreting kernel effects.

5. Conclusion

In this lab, Support Vector Machines were effectively applied to understand the influence of kernel choice and margin control. The experiments showed that: • The RBF kernel performs best on non-linear data like Moons. • The Linear kernel is ideal for linearly separable data such as Banknote Authentication. • The Polynomial kernel can capture complex shapes but risks overfitting. • Soft margins allow flexibility and improve model generalization. This lab reinforced the importance of kernel selection and hyperparameter tuning (C) for achieving balance between bias and variance.

Prepared By: NANDANI SONALE

SRN: PES2UG23CS364

Section: F