TEECE 2/2L – ECE Elective 2/2L

M4: Support Vector Machine

Lectured by:

Engr. Timothy M. Amado
Faculty, Electronics Engineering Department

Copyright Notice

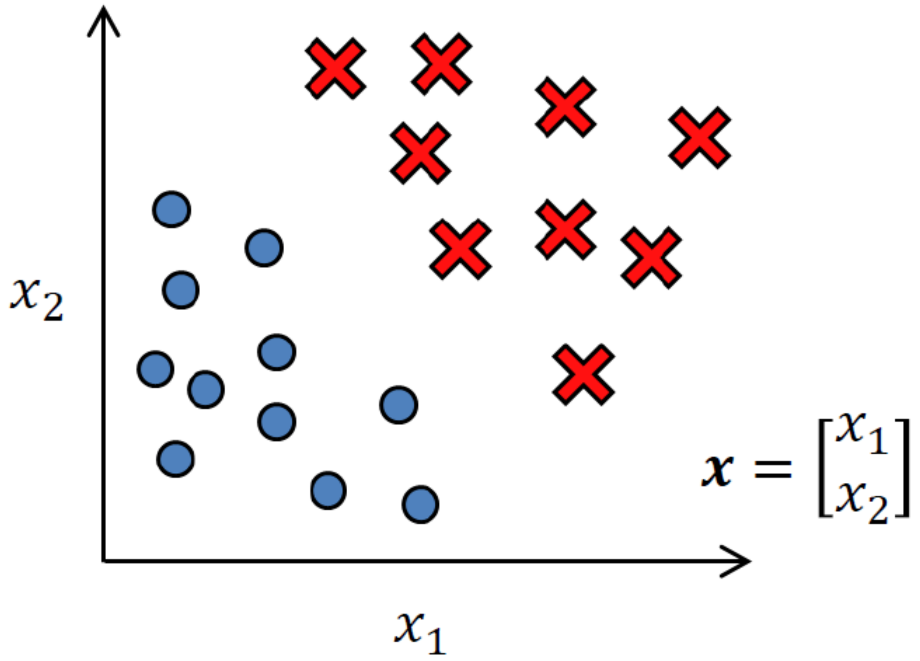
Outline

Support Vector Machine

- Large Margin Classifier
- Nonlinear SVM using Kernels
- Multiple Kernel Learning
- Hyperparameter Tuning

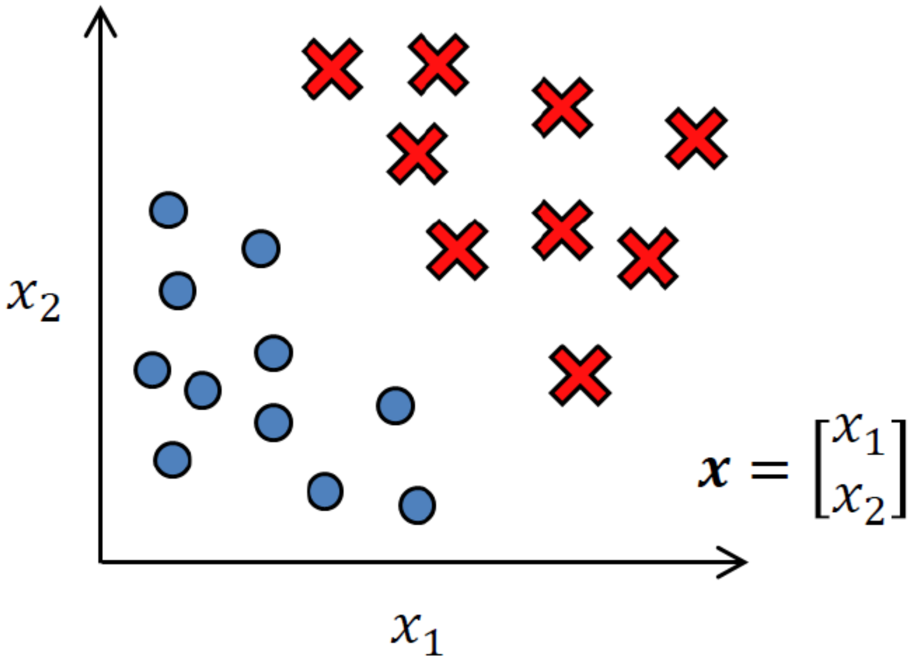
Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification tasks. It aims to find the optimal hyperplane in an N-dimensional space to separate data points into different classes. The algorithm maximizes the margin between the closest points of different classes.



Large Margin Classifier

Consider a classification data set whose samples from different classes can be separated by a line (linearly separable):

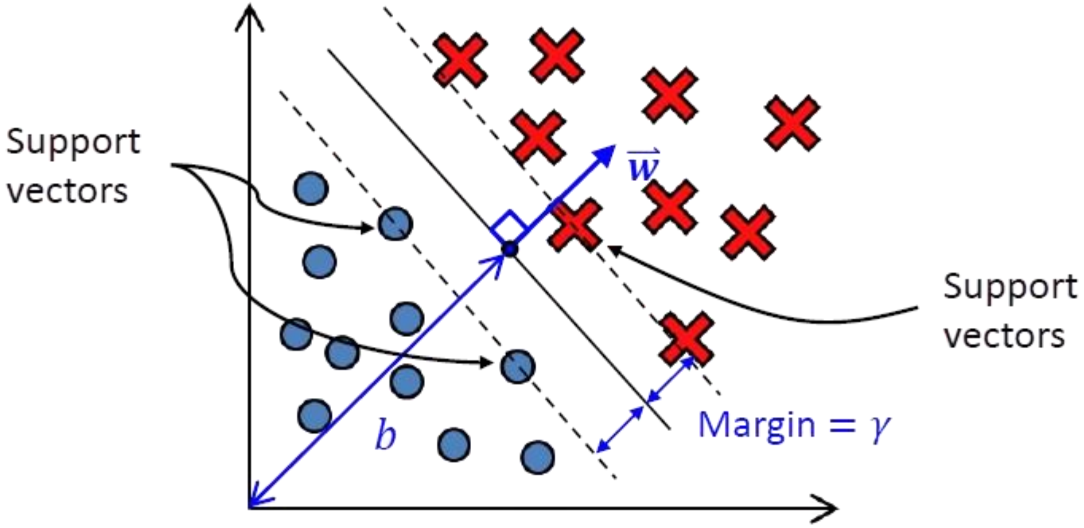


There exists many straight lines that can separate the positive from the negative samples.

What is the equation of the *best-separating line*?

Large Margin Classifier

Consider a classification data set whose samples from different classes can be separated by a line (linearly separable):

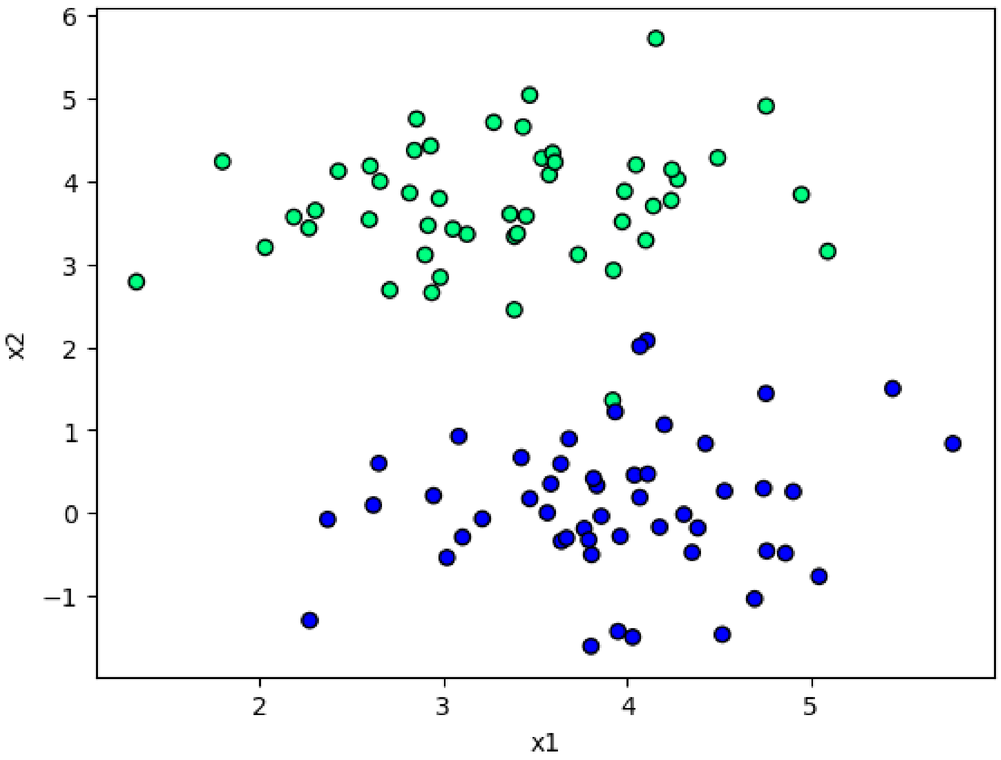


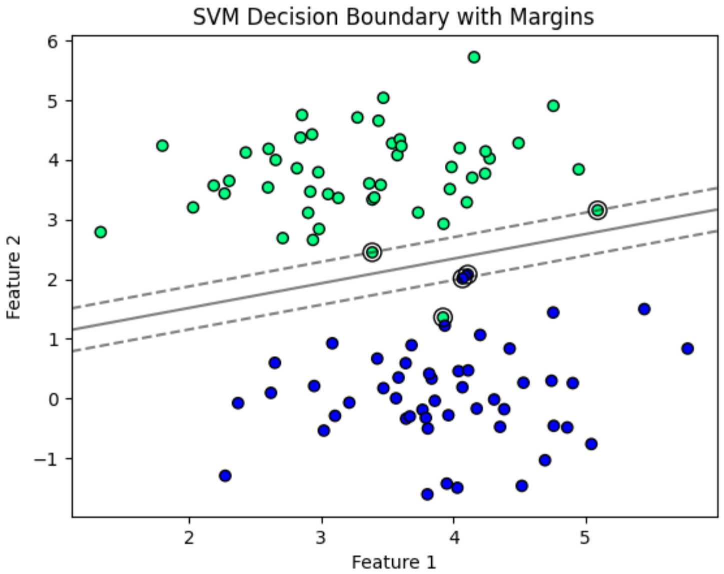
There exists many straight lines that can separate the positive from the negative samples.

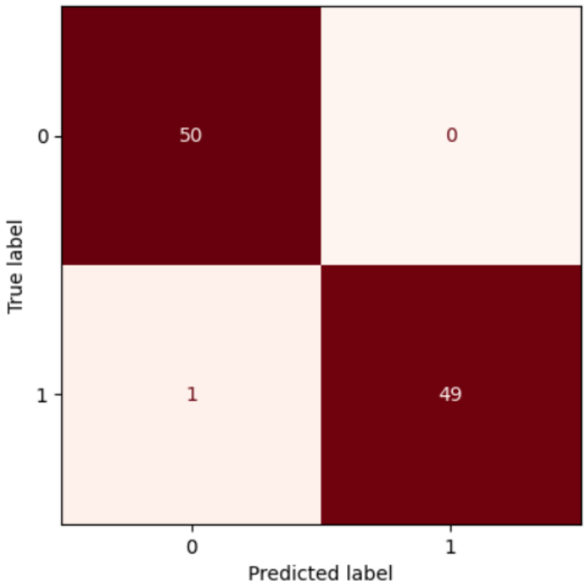
What is the equation of the *best-separating line*?

Large Margin Classifier

Programming Demonstration

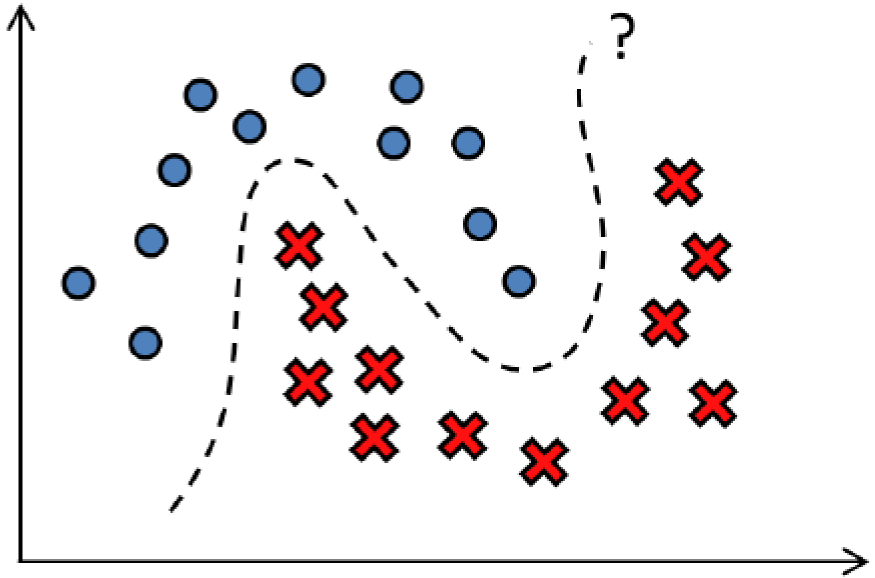






Nonlinear SVM Using Kernels

For samples that cannot be separated by a line, the classification problem now becomes a non-linear problem.



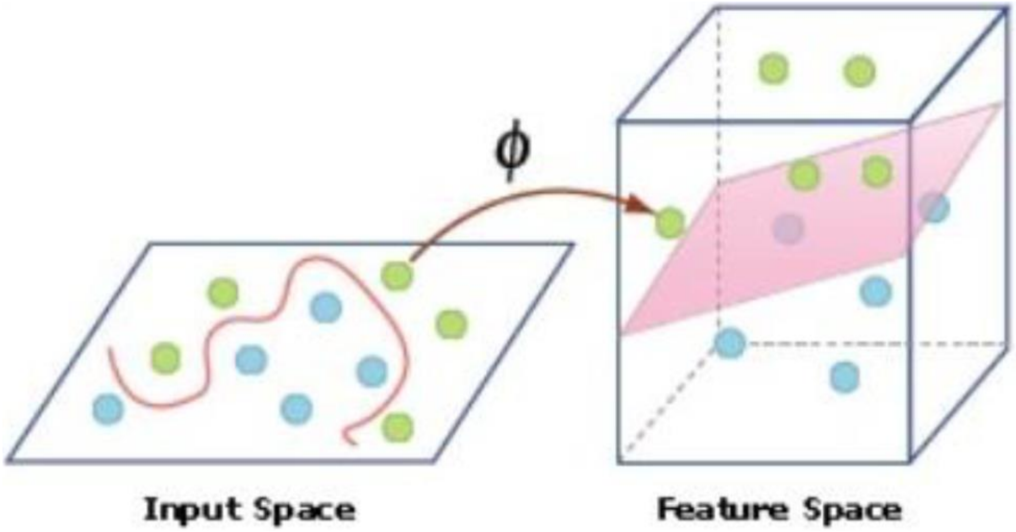
The “kernel trick” allows SVMs to handle non-linear separations by implicitly mapping data to higher-dimensional spaces.

$$K\left(x, x'\right) = x^T x'$$

Common kernels include the Polynomial kernel and the Gaussian Radial Basis Function (RBF) kernel.

Nonlinear SVM Using Kernels

For samples that cannot be separated by a line, the classification problem now becomes a non-linear problem.



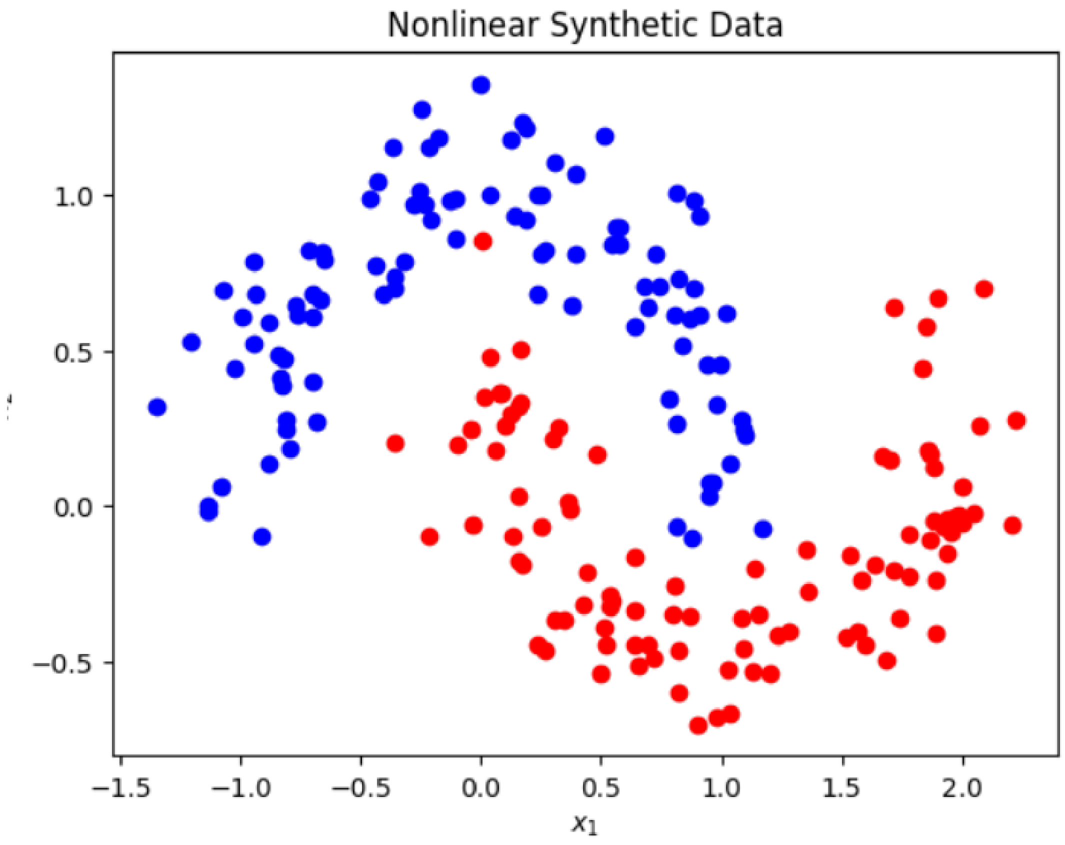
The "kernel trick" allows SVMs to handle non-linear separations by implicitly mapping data to higher-dimensional spaces.

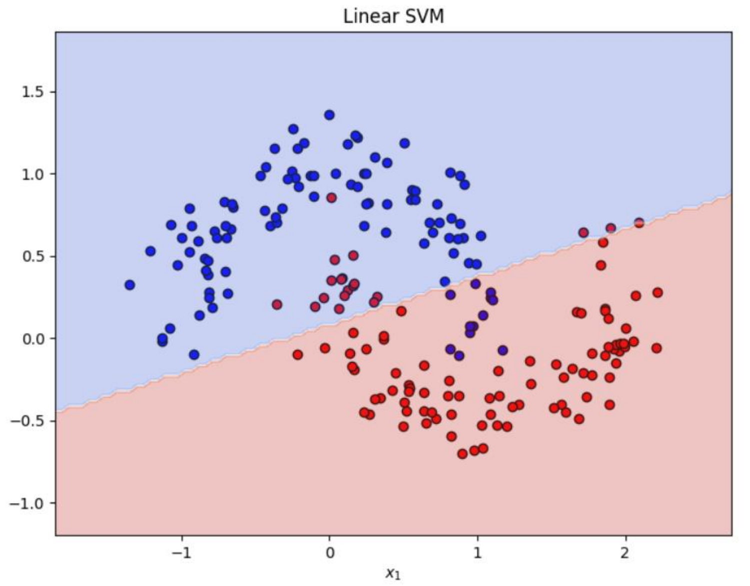
$$K\left(x, x'\right) = x^T x'$$

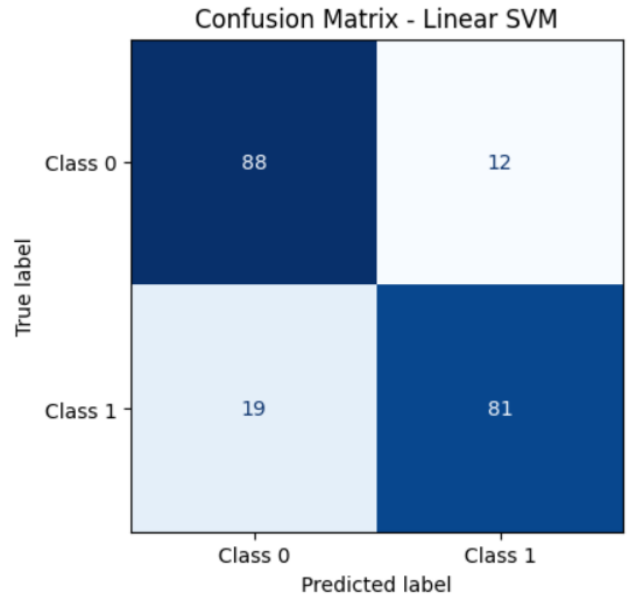
Common kernels include the Polynomial kernel and the Gaussian Radial Basis Function (RBF) kernel.

Nonlinear SVM Using Kernels

Programming Demonstration

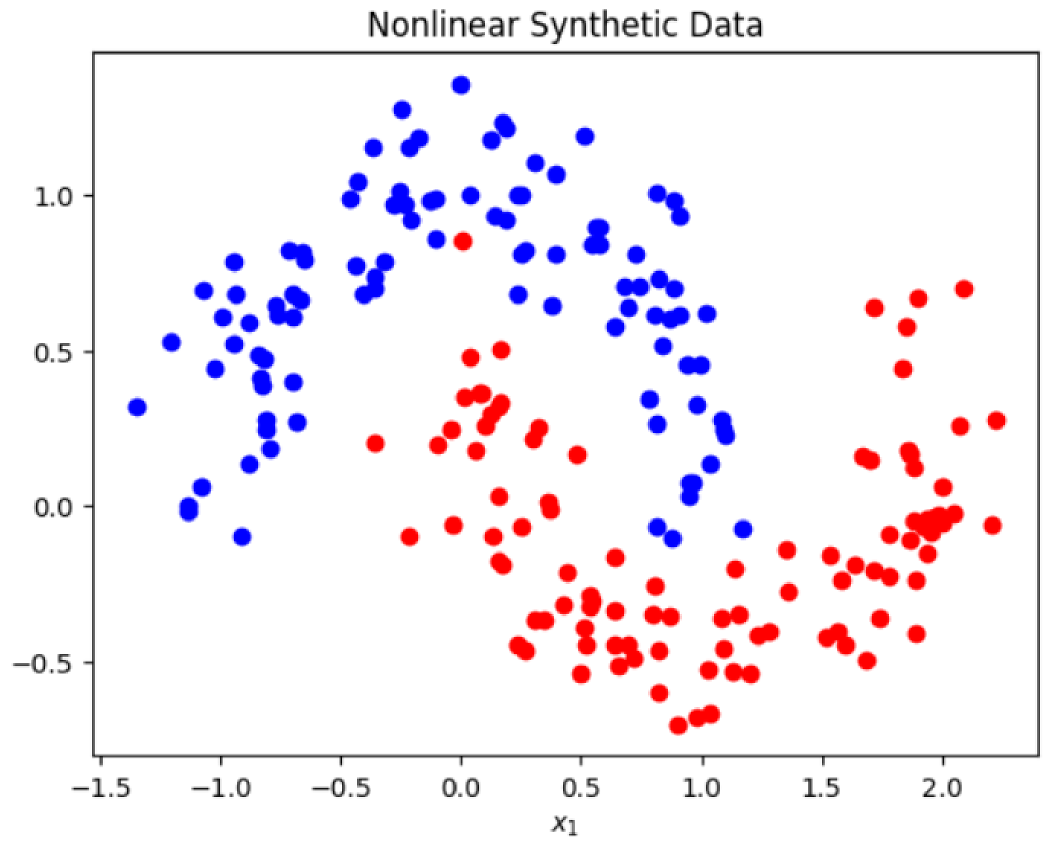


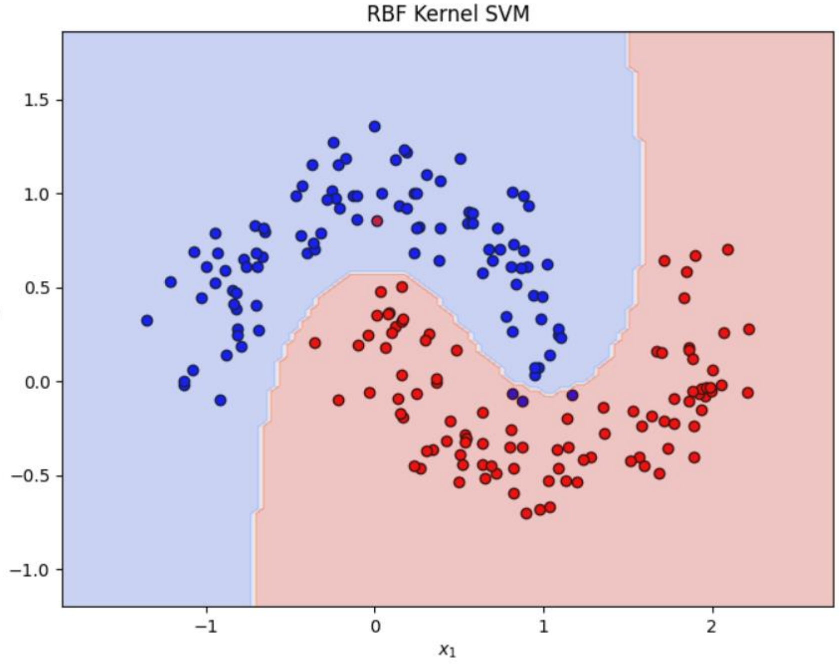


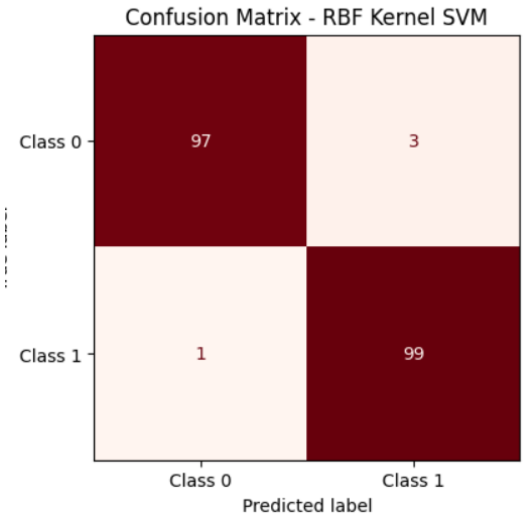


Nonlinear SVM Using Kernels

Programming Demonstration







Multiple Kernel Learning

There are many kernel functions, other than the Gaussian RBF. One can create new kernels by combining base kernels.

Common kernels:

Gaussian RBF

$$K(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|^2}{\sigma}\right)$$

Sigmoid kernel

$$K(\boldsymbol{x}, \boldsymbol{x}') = \tanh(a\langle \boldsymbol{x}, \boldsymbol{x}' \rangle + b)$$

Polynomial kernel

$$K(\boldsymbol{x}, \boldsymbol{x}') = (\langle \boldsymbol{x}, \boldsymbol{x}' \rangle + 1)^d$$

Linear kernel

$$K(\boldsymbol{x}, \boldsymbol{x}') = \langle \boldsymbol{x}, \boldsymbol{x}' \rangle$$

Hyperparameter Tuning

Hyperparameter tuning in Support Vector Machines (SVM) is the process of optimizing the parameters set prior to training to enhance model performance.

These hyperparameters are not learned from the data but significantly influence the model's behavior.

Hyperparameter Tuning

Key Hyperparameters in SVM

1. Kernel Type
    - Determines the transformation of input data (e.g., linear, polynomial, RBF, sigmoid).
    - Example: Use linear kernels for linearly separable data; RBF for non-linear patterns.
2. Regularization (C)
    - Controls the trade-off between maximizing margin (generalization) and minimizing training error.
    - High C: Prioritizes correct classification (risk of overfitting).
    - Low C: Allows more misclassifications for a larger margin (risk of underfitting).
3. Gamma (γ)
    - Defines the influence radius of a training example (for RBF, polynomial, sigmoid kernels).
    - High γ: Model captures fine details (risk of overfitting).
    - Low γ: Smoother decision boundaries.

Hyperparameter Tuning

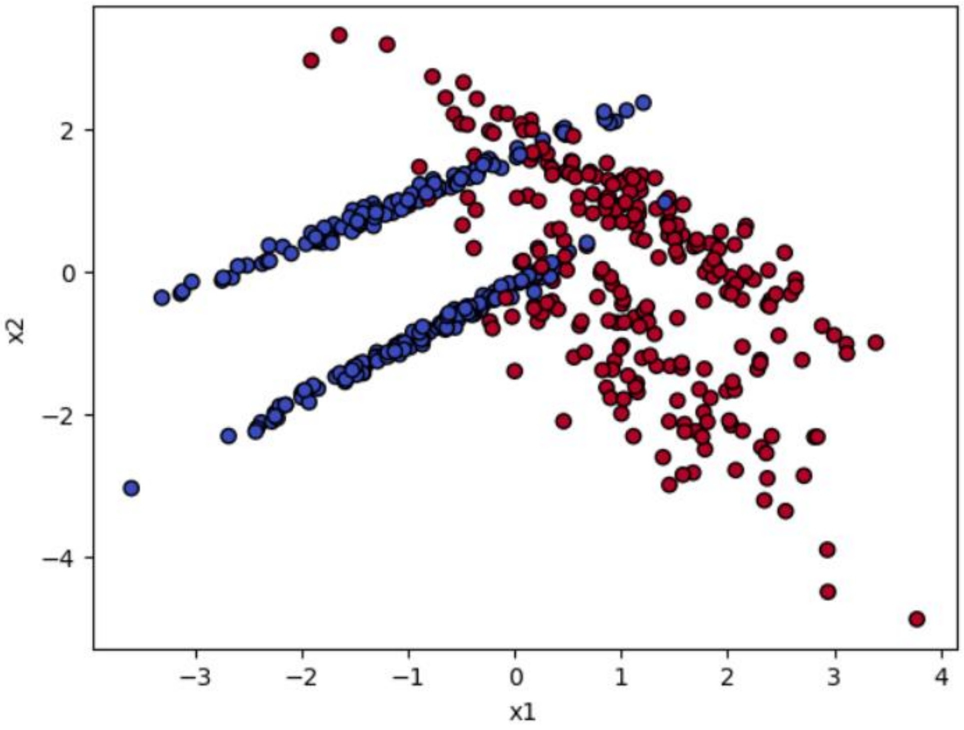
Key Hyperparameters in SVM

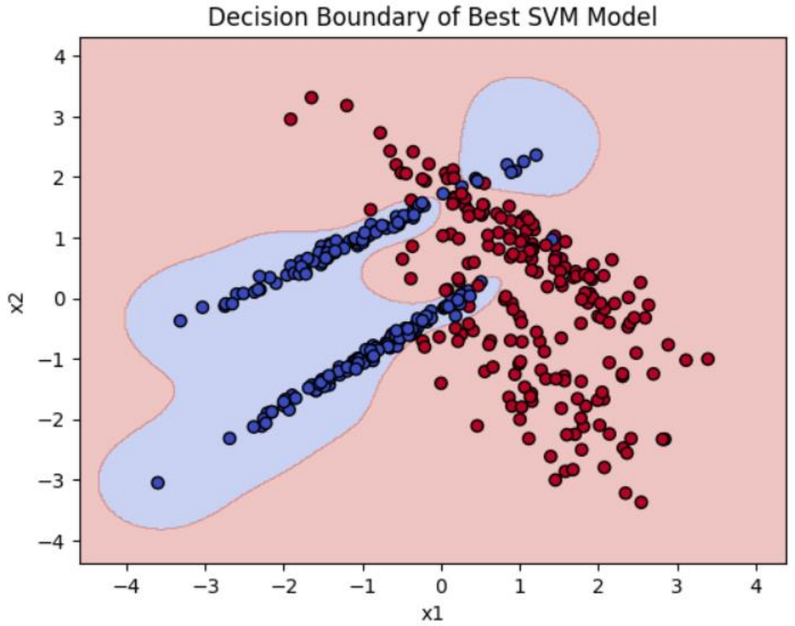
4. Degree (for polynomial kernels):
   - Sets the polynomial's degree. Higher degrees model complex relationships but may overfit.

5. Coef0 (for polynomial/sigmoid kernels):
   - Adjusts the impact of high-degree terms in polynomial kernels or the threshold in sigmoid kernels.

Nonlinear SVM Using Kernels

Programming Demonstration






```
Best Parameters: {'C': 100, 'gamma': 1, 'kernel': 'rbf'}
Best Model Accuracy on Test Set: 0.93
Classification Report:
              precision    recall  f1-score   support

           0       0.92      0.94      0.93        50
           1       0.94      0.92      0.93        50

    accuracy                           0.93       100
   macro avg       0.93      0.93      0.93       100
weighted avg       0.93      0.93      0.93       100
```

Thank You!