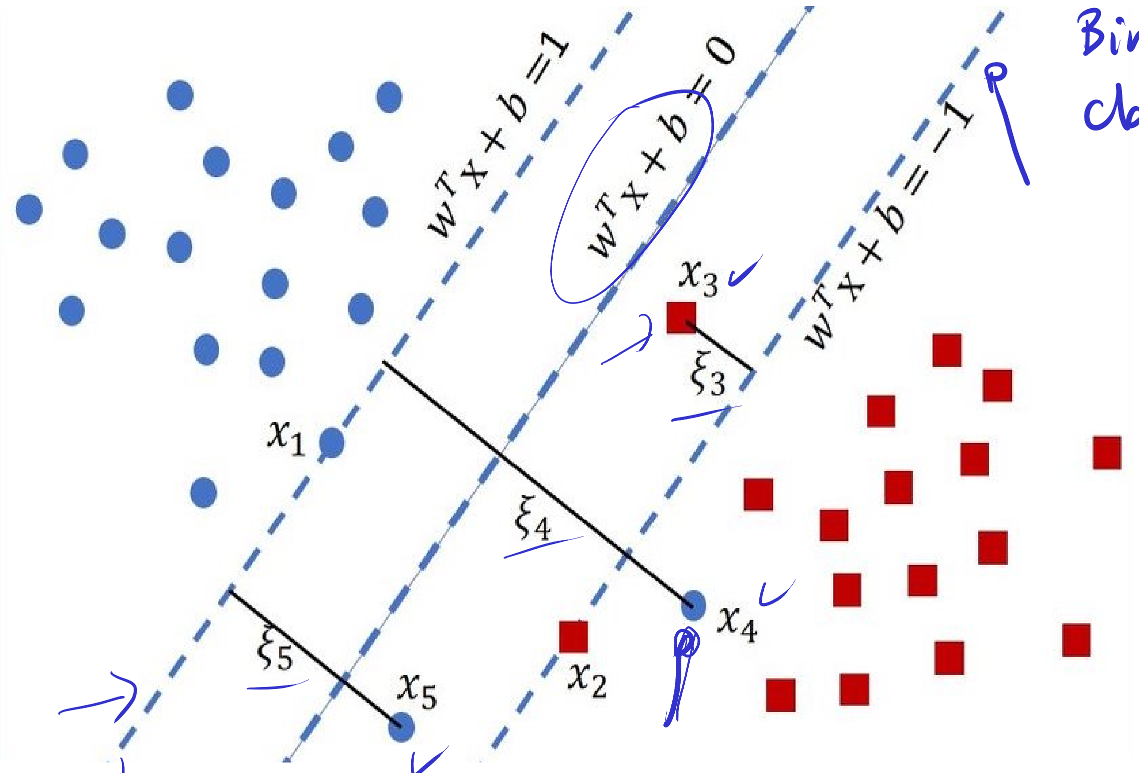
Introduction to Support Vector Machines (SVM)

1. Introduction

Support Vector Machines (SVM) are supervised learning models used for classification and regression tasks. They work by finding a hyperplane that best separates data into distinct classes, maximizing the margin between data points of different categories.

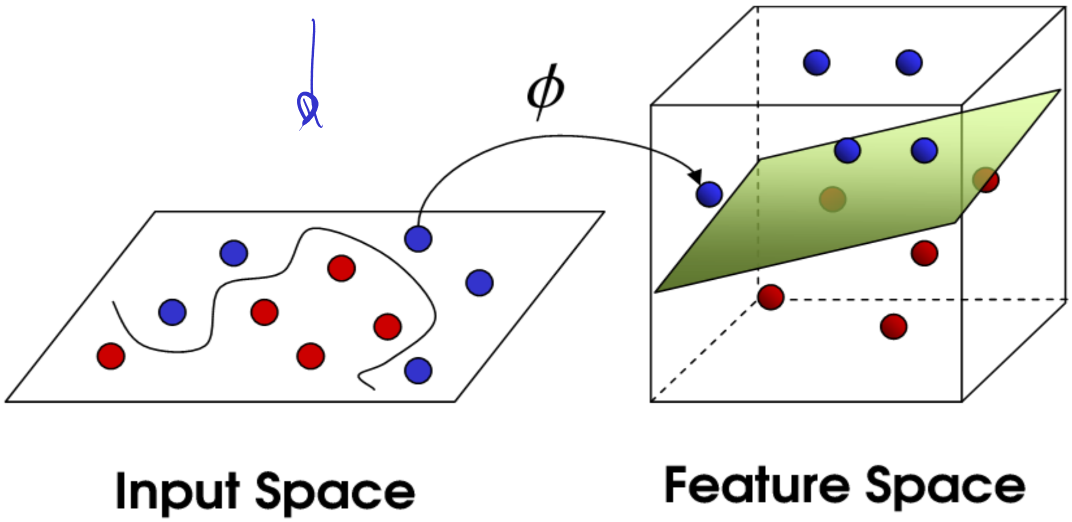
2. How SVM Works

Given training data, an SVM constructs a decision boundary (hyperplane) that maximizes the distance (margin) between data points of different classes. In cases where data is not linearly separable, the "kernel trick" allows mapping to higher-dimensional spaces.



Binary classification

margin



3. Mathematical Intuition

For a linear SVM, the decision boundary is:

$$\boxed{\underset{\text{weight}}{\mathbf{w}}^T \underset{\text{data}}{\mathbf{x}} + \overset{\text{bias}}{b} = 0}$$

Where:

- $\mathbf{w}$ is the weight vector,
- $\mathbf{x}$ is the feature vector,
- b is the bias.

The objective is to minimize:

$$\frac{1}{2}\|\mathbf{w}\|^2 + \overset{\text{regularization}}{C} \sum_{i=1}^{n} \overset{\text{slack variables}}{\xi_i}$$

subject to constraints on classification error ξ_i, with C controlling the trade-off between margin and misclassification.

4. Applications in Data Science

SVMs are widely used in:

- ✓ **Text classification**: spam detection, sentiment analysis
- ✓ **Image recognition**: face and object classification
- ✓ **Bioinformatics**: gene classification, protein categorization
- ✓ **Fraud detection**: binary classification of transaction legitimacy

5. Pros and Cons

Advantages

- Effective in high-dimensional spaces ✓
- Works well with clear margin of separation
- Flexible with different kernels (linear ✓, polynomial ✓, RBF) gaussian

Disadvantages

- Computationally expensive with large datasets
- Less interpretable than models like logistic regression
- Sensitive to feature scaling → standadScaler
  MinMax scaler

6. Conclusion

Support Vector Machines are a powerful classification tool for structured and semi-structured data. Although computationally heavy for large datasets, their accuracy and effectiveness in high-dimensional spaces make them a key algorithm in any data scientist's toolkit.