Support Vector Machines: Hyperplane, Margin, Hard Margin Classifier, and Soft Margin Classifier

Support Vector Machines (SVMs) are a set of supervised learning methods used for classification, regression, and outliers detection. This document explores the concepts of hyperplanes, margins, and the distinction between hard margin and soft margin classifiers.

1. Hyperplane in SVM

A hyperplane is a decision boundary that separates different classes in the feature space. For an SVM classifier in an n-dimensional space, a hyperplane is an $(n-1)$-dimensional subspace.

Linear Hyperplane

For a linearly separable dataset, the hyperplane can be expressed as:
$\mathbf{w} \cdot \mathbf{x} + b = 0$

where:

- $\mathbf{w}$ is the weight vector perpendicular to the hyperplane.
- b is the bias term.
- $\mathbf{x}$ is a point in the feature space.

Decision Rule

The decision rule for classifying a new data point $\mathbf{x}$ is:
$f(\mathbf{x}) = \text{sign}(\mathbf{w} \cdot \mathbf{x} + b)$

- If $f(\mathbf{x}) = 1$, the point is classified as one class.
- If $f(\mathbf{x}) = -1$, the point is classified as the other class.

2. Margin in SVM

The margin is the distance between the hyperplane and the nearest data points from each class. These nearest points are called support vectors. SVM aims to find the hyperplane that maximizes this margin.

Margin Definition

The margin M is given by:
$M = \frac{2}{\|\mathbf{w}\|}$

The goal is to maximize M, which translates to minimizing $\|\mathbf{w}\|$.

3. Hard Margin Classifier

A hard margin classifier is used when the data is linearly separable. It finds the hyperplane that perfectly separates the two classes with the maximum margin.

Formulation

The optimization problem for a hard margin SVM is:
$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2$

subject to:
$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \quad \forall i$

where y_i is the class label of the i-th sample, and $\mathbf{x}_i$ is the i-th sample.

Key Points

- Suitable for linearly separable data.
- No tolerance for misclassification or outliers.
- Guarantees the maximum margin hyperplane.

4. Soft Margin Classifier

A soft margin classifier is used when the data is not linearly separable. It introduces slack variables to allow some misclassification and handle outliers.

Slack Variables

Slack variables $\xi_i \geq 0$ measure the degree of misclassification of the i-th sample. The soft margin classifier balances maximizing the margin and minimizing the classification error.

Formulation

The optimization problem for a soft margin SVM is:
$\min_{\mathbf{w},b,\xi} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n} \xi_i$

subject to:
$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \forall i$
$\xi_i \geq 0, \quad \forall i$

where C is a regularization parameter controlling the trade-off between margin maximization and classification error.

Key Points

- Suitable for non-linearly separable data.
- Allows some misclassification to achieve a better generalization.
- The parameter C controls the penalty for misclassification: a high C means a high penalty (less tolerance for errors), while a low C means a low penalty (more tolerance for errors).