Support Vector Machines (SVM)

1 Introduction

Support Vector Machines (SVM) are supervised learning models used for classification and regression analysis. Developed by Vladimir Vapnik and his colleagues, SVMs are based on the concept of decision planes that define decision boundaries.

2 Concept

The basic idea of SVM is to find a hyperplane that best divides a dataset into classes. The hyperplane is chosen to maximize the margin between the classes, which is the distance between the hyperplane and the nearest data point from either class. These nearest data points are called support vectors.

3 Mathematical Formulation

Given a training dataset with n samples (x_i, y_i) where $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$, the goal is to find a hyperplane defined by:

$$w \cdot x + b = 0$$

where w is the weight vector and b is the bias.

The optimization problem can be formulated as:

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

subject to:

$$y_i(w \cdot x_i + b) \geq 1 \quad \forall i$$

4 Kernel Trick

In cases where the data is not linearly separable, SVM can use a kernel function to map the data into a higher-dimensional space where it becomes linearly separable. Commonly used kernels include the polynomial kernel and the radial basis function (RBF) kernel.

5 Algorithm

The SVM algorithm can be summarized in the following steps:

1. **Input:** Training data (x_i, y_i), $i = 1, \ldots, n$, where $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$.

2. **Initialize:** Set up the optimization problem to find the hyperplane parameters w and b.

3. **Solve the optimization problem:** Use quadratic programming to solve:

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

   subject to:
$$y_i(w \cdot x_i + b) \geq 1 \quad \forall i$$

4. **Determine the support vectors:** Identify the data points that lie on the margin, i.e., the points for which $y_i(w \cdot x_i + b) = 1$.

5. **Construct the decision function:** The decision function for a new data point x is given by:
$$f(x) = \text{sign}(w \cdot x + b)$$

6 Conclusion

SVMs are powerful tools for classification and regression tasks, especially in high-dimensional spaces. They are effective in cases where the number of dimensions exceeds the number of samples and are versatile due to the use of kernel functions.