Support Vector Machine

Jerry Peng

August 23, 2023

1 Support Vector Machine

SVM (Support Vector Machine) is a supervised learning algorithm primarily used for classification and regression tasks. Its fundamental idea is to find a hyperplane (in high-dimensional space) that maximizes the margin between two classes. Figure 1 is a simple SVM

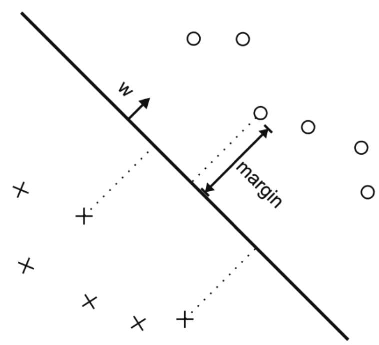

Figure 1: Simple SVM

Goal In a dataset with N dimensions (where N represents the number of features), the goal is to find the optimal decision boundary, also called a hyperplane, that accurately separates data points of different classes.

Support Vectors Support vectors are the core concept in the SVM algorithm. They are those data points in the dataset that lie close to the decision boundary defined by the SVM. Specifically, support vectors are the data points that determine the position of the maximum margin hyperplane. Only support vectors contribute to the decision function of the SVM model. This means that if you remove other points in the dataset but keep the support vectors, the SVM decision boundary will not change. However, if you remove or alter a support vector, the decision boundary might be affected.

Margin The minimum distance between the support vectors and the hyperplane is referred to as the margin.

1.1 Calculate the Hyperplane

Suppose we have a binary classification task. We have a dataset containing n labeled data points, each denoted as (x_i, y_i). Each $x_i \in \mathbb{R}^d$ is a feature vector, and each $y_i \in -1, 1$ is the corresponding label. Therefore, we can compute the hyperplane using the following formula:

$$f(x) = \mathbf{w}x + b = 0 \tag{1}$$

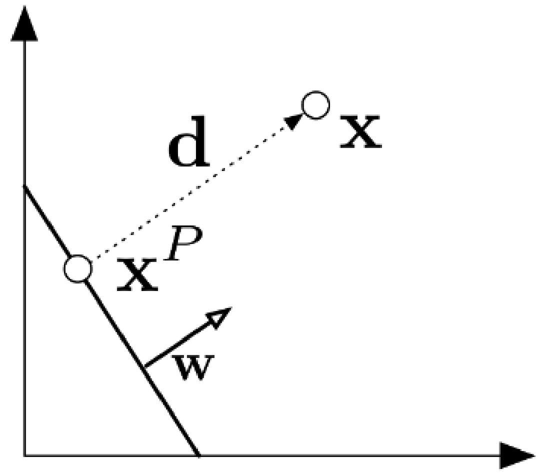

Figure 2: calculate margin

1.1.1 Margin

Before computing the specific Hyperplane, we need to calculate the Margin. Taking Figure 2 as an example. Suppose d is the vector of minimum length between hyperplane H and x, and x^P is a projection of x on H. Then, we can derive the formula: $x^P = x - d$. Since d is parallel to $\mathbf{w}$, we have $d = \alpha\mathbf{w}$ for some $\alpha \in \mathbb{R}$. Based on equation 1, we can derive $\mathbf{w}^T x + b = \mathbf{w}^T(x - d) + b = \mathbf{w}^T(x - \alpha\mathbf{w}) + b = 0$. Rearranging this expression, we can obtain:

$$\alpha = \frac{\mathbf{w}^T x + b}{\mathbf{w}^T\mathbf{w}} \tag{2}$$

Thus, we can determine the length of d, which is:

$$\|d\|_2 = \sqrt{d^T d} = \sqrt{\alpha^2 d^T d} = \frac{|\mathbf{w}^T x + b|}{\|\mathbf{w}\|_2} \tag{3}$$

Hence, the margin of H respect to D is

$$\gamma(\mathbf{w}, b) = \min_{x \in D} \frac{|\mathbf{w}^T x + b|}{\|\mathbf{w}\|_2} \tag{4}$$

1.1.2 Max Margin Classifier

We can formulate the search for the maximum margin separating hyperplane as a constrained optimization problem. The objective is to maximize the margin under the constraint that all data points must lie on the correct side of the hyperplane:

$$\max_{\mathbf{w},b} \gamma(\mathbf{w}, b) \text{ such that } \forall i \ y_i(\mathbf{w}^T x_i + b) \geq 0$$

Substituting $\gamma(\mathbf{w}, b)$ from Expression 4, we obtain:

$$\max_{\mathbf{w},b} \frac{1}{\|\mathbf{w}\|_2} \min_{x \in D} |\mathbf{w}^T x + b| \ \textbf{s.t.} \ \forall i \ y_i(\mathbf{w}^T x_i + b) \geq 0$$

Since the scale of the hyperplane is always invariant, we can arbitrarily fix the scale of $\mathbf{w}$ and b. Therefore, we set the scale such that the following expression holds:

$$\min_{x \in D} |\mathbf{w}^T x + b| = 1$$

Substituting this expression into the previously defined formula, we obtain:

$$\max_{\mathbf{w},b} \frac{1}{\|\mathbf{w}\|_2} \times 1 = \min_{\mathbf{w},b} \|\mathbf{w}\|_2 = \min_{\mathbf{w},b} \mathbf{w}^T \mathbf{w}$$

Thus, the updated optimization problem can be formulated as:

$$\min_{\mathbf{w},b} \mathbf{w}^T \mathbf{w} \textbf{ s.t. } \forall i \; y_i(\mathbf{w}^T x_i + b) \geq 0 \text{ and} \min_{x \in D} |\mathbf{w}^T x + b| = 1$$

The aforementioned optimization problem can be succinctly written as:

$$\min_{\mathbf{w},b} \mathbf{w}^T \mathbf{w} \textbf{ s.t. } \forall i \; y_i(\mathbf{w}^T x_i + b) \geq 1$$

References