

Separating Hyperplane Problem

Statement

Given a set of data points labeled as positive or negative, find a separating hyperplane that maximizes the margin between the hyperplane and the closest data points. Formulate this problem as a convex optimization problem and explain how it relates to Support Vector Machines (SVMs) for classification.

Mathematical Formulation

Assume $\mathbf{x}_i = (x_1, x_2, \dots, x_n)$ is the given data vector and y is either $+1$ or -1 . Let the separating hyperplane be

$$w^T \mathbf{x} + b = 0 \quad (1)$$

where w is a normal vector to the hyperplane and b is the bias term. The margin of point \mathbf{x}_i from the hyperplane is given by

$$\frac{|w^T \mathbf{x}_i + b|}{\|w\|} \quad (2)$$

The margin between the hyperplane and the closest data point is then the minimum of this quantity

$$\min_i \frac{|w^T \mathbf{x}_i + b|}{\|w\|} \quad (3)$$

Thus, we need to maximize

$$\min_i \left(\frac{|w^T \mathbf{x}_i + b|}{\|w\|} \right) \quad (4)$$

We can choose w and b such that

$$\min_i |w^T \mathbf{x}_i + b| = 1 \quad (5)$$

Since $w^T \mathbf{x} + b = 0$ is the separating plane of the data points based on the labels which have values only $+1$ and -1 , thus

$$\text{sign}(w^T \mathbf{x}_i + b) = y_i \quad (6)$$

Hence

$$\min_i (y_i (w^T \mathbf{x}_i + b)) = 1 \quad (7)$$

Now

$$\min_i \left(\frac{|w^T \mathbf{x}_i + b|}{\|w\|} \right) = \frac{1}{\|w\|} \quad (8)$$

Hence our problem becomes:

$$\text{maximize } \frac{1}{\|w\|} \text{ subject to } \min_i (y_i (w^T \mathbf{x}_i + b)) = 1.$$

This can be written another way:

$$\text{minimize } \frac{\|w\|^2}{2} \text{ subject to } y_i (w^T \mathbf{x}_i + b) \geq 1.$$

This optimization problem is convex.

Relation with SVM

The problem statement given, to find a separating hyperplane that maximizes the margin between the hyperplane and the closest data points, is the fundamental objective of Support Vector Machines (SVMs) for classification. SVMs aim to find the optimal hyperplane that separates data points of different classes while maximizing the margin between the hyperplane and the nearest data points from each class. This is achieved by formulating the problem as a convex optimization problem, exactly as described in the mathematical formulation section above.

Therefore, the problem statement described is directly related to the core objective of SVMs, making it a fundamental concept in SVM classification.