Hyperplane Classification

Consider the class of hyperplanes, i.e., dot product plus a bias:

$$(\mathbf{w} \cdot \mathbf{x}) + b = 0$$

and corresponding activation function:

$$\text{sign}((\mathbf{w} \cdot \mathbf{x}) + b)$$

For linearly separable examples, there a unique optimal hyperplane defined by maximizing the margin, which is the minimum distance of an example from the decision boundary.

This can be expressed as:

$$\max_{\mathbf{w},b} \min\{\|\mathbf{x} - \mathbf{x}_i\| \mid (\mathbf{w} \cdot \mathbf{x}) + b = 0\}$$

The optimal hyperplane can be found by solving the problem:

$$\text{minimize } \|\mathbf{w}\|^2/2$$
$$\text{s.t. } y_i((\mathbf{w} \cdot \mathbf{x}_i) + b) \geq 1, i \in \{1, \ldots, m\}$$

That is, we require the smallest weights such that positive examples ≥ 1 and negative example ≤ -1.

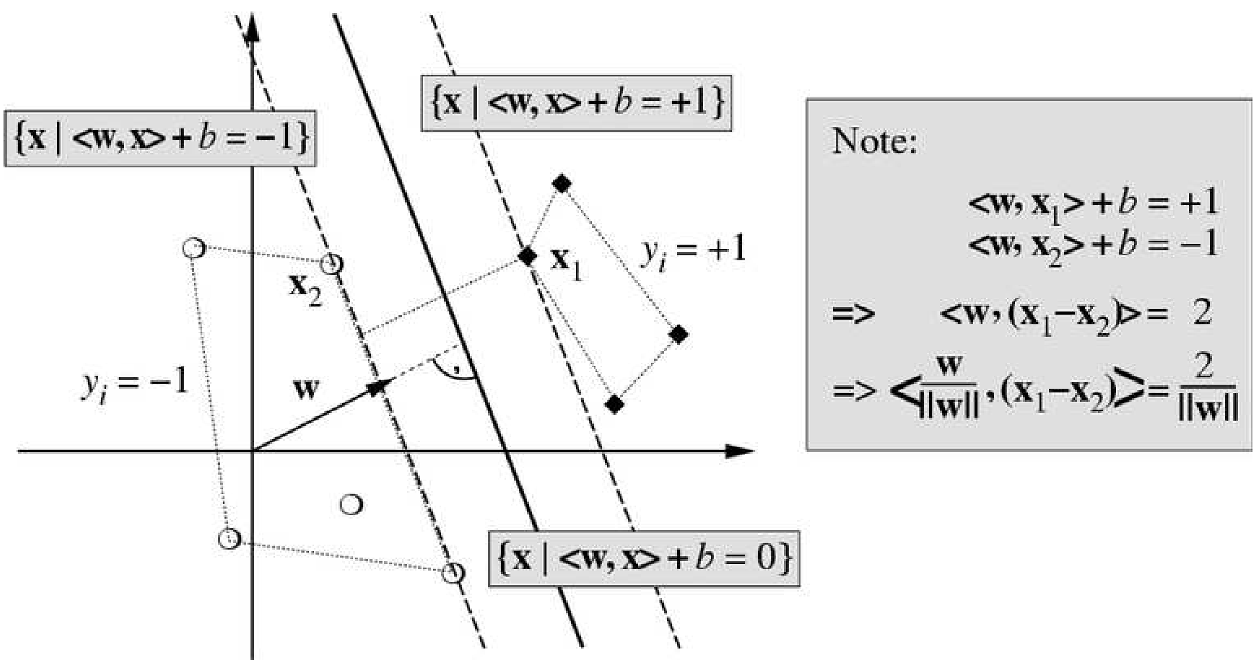
The smallest weights correspond to the maximum margin. Note that:

$$1 = \mathbf{w} \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|^2}$$

so the length of the margin is equal to:

$$\frac{\|\mathbf{w}\|}{\|\mathbf{w}\|^2} = \frac{1}{\|\mathbf{w}\|}$$

Hyperplane Example



Also see examples in book (p. 149 and p. 154).

Math Tricks

Through a series of math tricks, the problem can be transformed to the decision function:

$$f_a(\mathbf{x}) = \text{sign}\left(b + \sum_{i=1}^{m} y_i \alpha_i (\mathbf{x} \cdot \mathbf{x}_i)\right)$$

The α_i weights are found by solving:

$$\text{maximize} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\text{subject to } \alpha_i \geq 0 \text{ and } \sum_{i=1}^{m} \alpha_i y_i = 0$$