Optimization Problem

$$\begin{array}{ll} \text{Minimize} & f(\mathbf{w}, b) \equiv \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{subject to} & y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \quad i = 1, \ldots, m \end{array}$$

This is an optimization problem in $(n + 1)$ variables, with m linear inequality constraints.

Introducing Lagrange multipliers α_i, $i = 1, \ldots, m$ for the inequality constraints above gives the primal Lagrangian:

$$\begin{array}{ll} \text{Minimize} & L_P(\mathbf{w}, b, \boldsymbol{\alpha}) \equiv \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{m} \alpha_i [y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1] \\ \text{subject to} & \alpha_i \geq 0, \quad i = 1, \ldots, m \end{array}$$

Optimization Problem (continued)

Setting the gradients of L_P with respect to $\mathbf{w}, b$ equal to zero gives:

$$\frac{\partial L_P}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{m} \alpha_i y_i \mathbf{x}_i, \qquad \frac{\partial L_P}{\partial b} = 0 \Rightarrow \sum_{i=1}^{m} \alpha_i y_i = 0$$

Substituting the above in the primal gives the following dual problem:

$$\text{Maximize} \quad L_D(\boldsymbol{\alpha}) \equiv \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\text{subject to} \quad \sum_{i=1}^{m} \alpha_i y_i = 0; \qquad \alpha_i \geq 0, \;\; i = 1, \ldots, m$$

This is a convex quadratic programming problem in $\boldsymbol{\alpha}$.

Solution

The parameters $\mathbf{w}, b$ of the maximal margin classifier are determined by the solution $\boldsymbol{\alpha}$ to the dual problem:

$$\mathbf{w} = \sum_{i=1}^{m} \alpha_i y_i \mathbf{x}_i$$

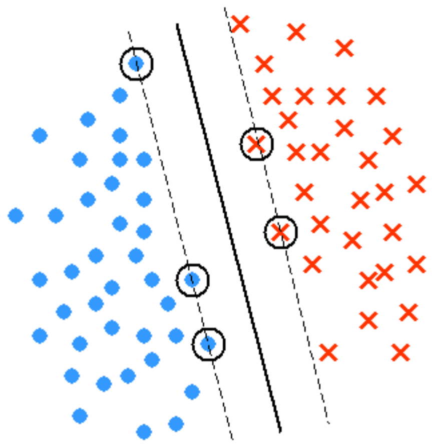
$$b = -\frac{1}{2}\left(\min_{y_i=+1} (\mathbf{w} \cdot \mathbf{x}_i) + \max_{y_i=-1} (\mathbf{w} \cdot \mathbf{x}_i)\right)$$

Support Vectors

Due to certain properties of the solution (known as the Karush-Kuhn-Tucker conditions), the solution $\boldsymbol{\alpha}$ must satisfy

$$\alpha_i[y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1] = 0, \quad i = 1, \ldots, m.$$

Thus, $\alpha_i > 0$ only for those points x_i that are closest to the classifying hyperplane. These points are called the **support vectors**.



Non-Separable Case

Want to relax the constraints

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1.$$

Can introduce slack variables ξ_i:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i,$$

where $\xi_i \geq 0\ \forall i$. An error occurs when $\xi_i > 1$.

Thus we can assign an extra cost for errors as follows:

Minimize	$f(\mathbf{w}, b, \boldsymbol{\xi}) \equiv \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{m} \xi_i$
subject to	$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i; \quad \xi_i \geq 0, \quad i = 1, \ldots, m$

Non-Separable Case (continued)

Dual problem:

Maximize $L_D(\boldsymbol{\alpha}) \equiv \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j)$

subject to $\sum_{i=1}^{m} \alpha_i y_i = 0; \quad 0 \leq \alpha_i \leq C, \;\; i = 1, \ldots, m$

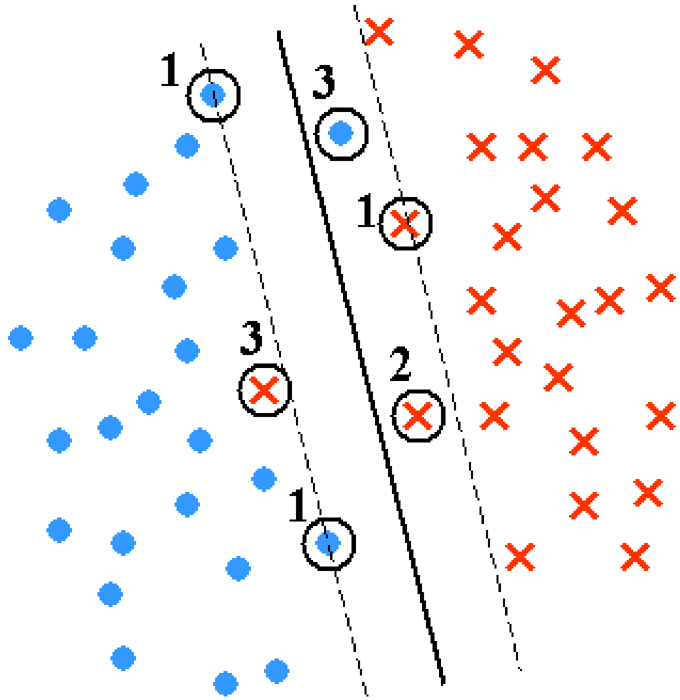
Solution:

The solution for $\mathbf{w}$ is again given by

$$\mathbf{w} = \sum_{i=1}^{m} \alpha_i y_i \mathbf{x}_i.$$

The solution for b is similar to that in the linear case.

Visualizing the Solution in the Non-Separable Case



1. Margin support vectors $\xi_i = 0$ Correct
2. Non-margin support vectors $\xi_i < 1$ Correct (in margin)
3. Non-margin support vectors $\xi_i > 1$ Error