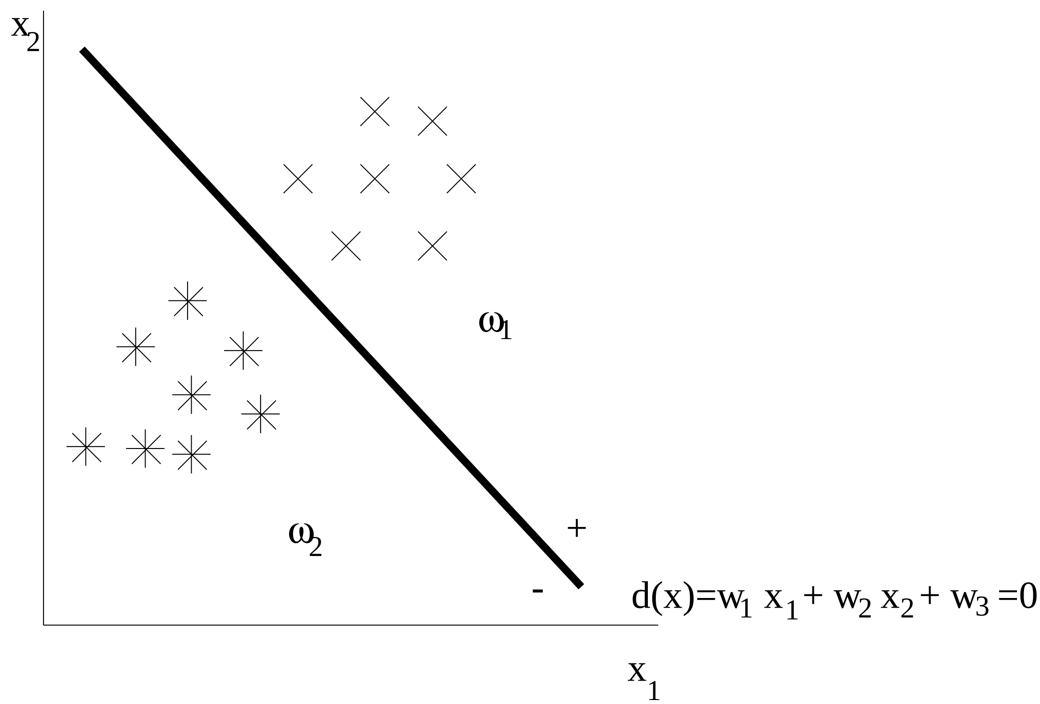
Decision Function

Decision Functions:

- Linear Decision Functions
- Pattern Space and Weight Space
- Geometrical Properties
- Implementation of Decision Functions

Decision Function:



Success:

(1) the form of $d(x)$.

(2) determination of coefficients

Linear Decision Functions

•

$$
\begin{aligned}
d(x) &= w_1 x_1 + \ldots + w_n x_n + w_{n+1} \\
&= w_0' x + w_{n+1}
\end{aligned}
$$

$w_0 = (w_1, \ldots, w_n)'$ – weight or parameter vector

- Introduce the augmented pattern and weight vectors:

$$
\begin{aligned}
x &= (x_1, \ldots, x_n, 1)' \\
w &= (w_1, \ldots, w_n, w_{n+1})'
\end{aligned}
$$

The decision function can be written as:

$$d(x) = w'x$$

For the two-class case, a decision function $d(x)$ suppose to have the property

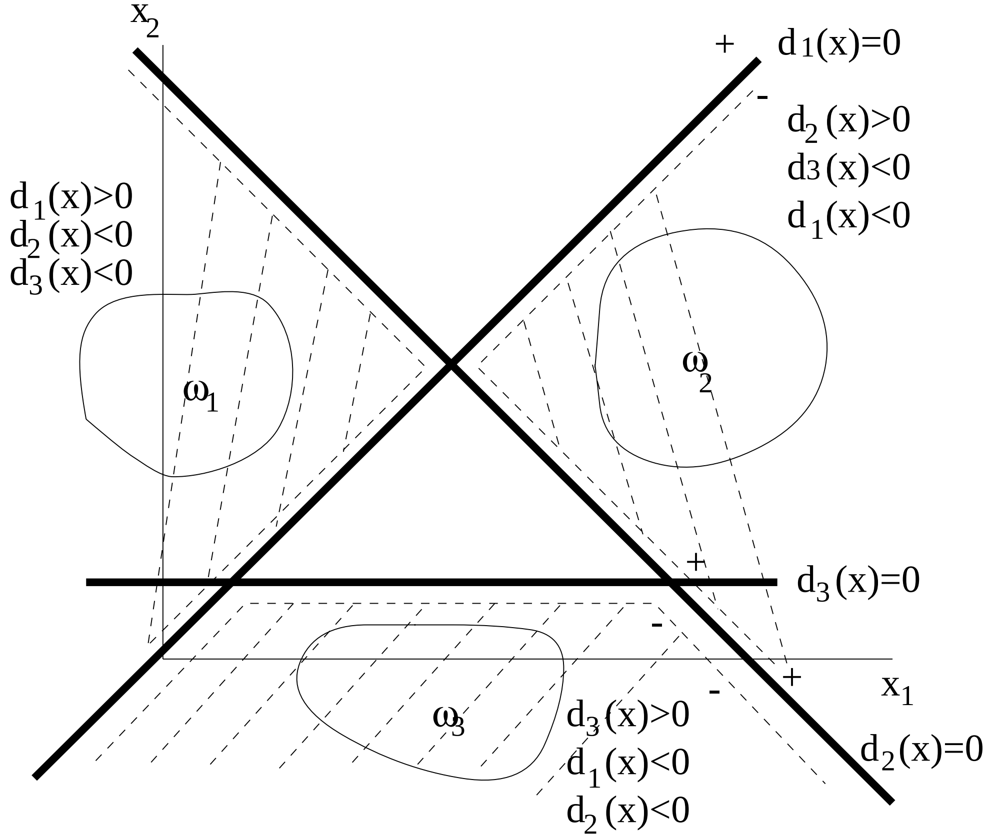
$$d(x) = w'x \begin{cases} > 0 & \text{if } x \in \omega_1 \\ < 0 & \text{if } x \in \omega_2 \end{cases}$$

For the multi-class case, denoted as $\omega_1, \ldots, \omega_M$, there are three cases to occur:

Case 1. Each pattern class separable from others by a single surface

$$d_i(x) = w_i' x \begin{cases} > 0 & \text{if } x \in \omega_i \\ < 0 & \text{otherwise} \end{cases}$$

$$i = 1, \ldots, M$$

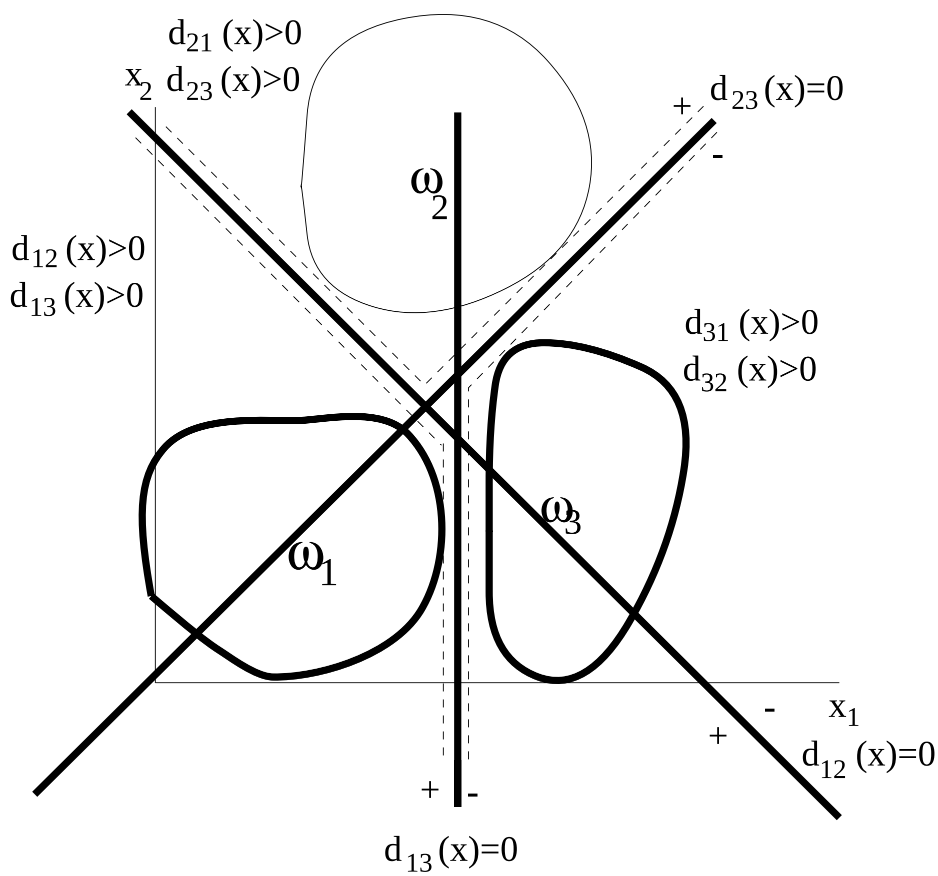


Case 2. Each pattern class separable from every other individual class by a distinct decision surface

$$d_{ij}(x) = w'_{ij}x > 0 \text{ if } x \in \omega_i$$
$$\text{for all } j \neq i$$

where $d_{ij}(x) = -d_{ji}(x)$ totally $\frac{M(M-1)}{2}$ decision surfaces

- No class separable from others by a single decision surface



Case 3. There exist M decision functions,

$$d_k(x) = w_k' x, \quad k = 1, \ldots, M$$

so that

$$\begin{aligned} &d_i(x) > d_j(x) \ \text{ if } x \in \omega_i \\ &\text{for all } j \neq i \end{aligned}$$

This is a special instance of Case 2 since we may define

$$\begin{aligned} d_{ij}(x) &= d_i(x) - d_j(x) \\ &= (w_i - w_j)' x \\ &= w_{ij}' x \end{aligned}$$

Case 2 is more general than Case 3!

General Decision Functions

- The pattern classes are linearly separable, if classifiable by any of linear decision function cases discussed above.

- One useful way to generalize the linear decision function concept is to consider the quadratic decision function

$$d(x) = x'Ax + x'b + c$$

Pattern Space and Weight Space

- Suppose we observe $k = k_1 + k_2$, pattern vectors from the pattern space E^n so that k_1 pattern vectors among them belong to ω_1 and k_2 pattern vectors among them belong to ω_2:

$$\left.\begin{array}{ll} x_1^{(1)}, \ldots, x_{k_1}^{(1)} : & \omega_1 \\ x_1^{(2)}, \ldots, x_{k_2}^{(2)} : & \omega_2 \end{array}\right\} \quad x_j^{(i)} \in E^n$$

- Find $w = (w_1, \ldots, w_{n+1})' \in E^{n+1}$:

$$\begin{cases} (x_j^{(1)'}, 1)w > 0, \\ j \leq k_1, \end{cases} \quad \begin{cases} (x_l^{(2)'}, 1)w < 0, \\ l \leq k_2, \end{cases}$$

or

$$\begin{cases} (x_j^{(1)'}, 1)w > 0, & j \leq k_1 \\ (-x_l^{(1)'}, -1)w > 0, & l \leq k_2 \end{cases}$$

- The desired weights w are the solution to k linear inequalities:

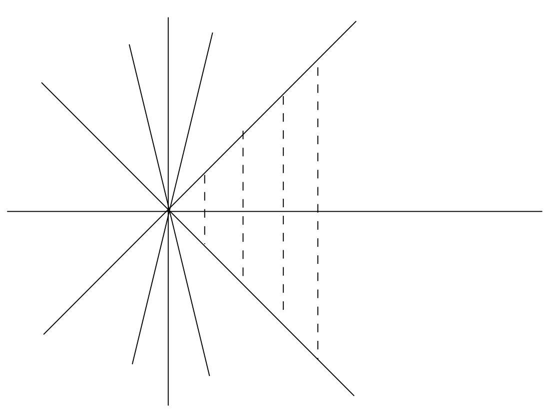
$$x_1' w > 0, \ldots, x_k' w > 0$$

Composes a convex polyhedral cone in E^{n+1}
where $x_1, \ldots, x_k, w \in E^{n+1}$.

- Convex: w_1, w_2 are solutions $\Rightarrow$
  $\alpha w_1 + (1-\alpha) w_2$ is a solution for $0 \leq \alpha \leq 1$.

- cone:
  w is a solution $\Rightarrow$
  λw is a solution for $\lambda > 0$.

- polyhedral:
  The solution to each individual linear inequality $x_j' w > 0$ composes a half space in E^{n+1}. The intersection of those k half spaces in E^{n+1} forms a polyhedral shape.



Geometric Properties

- For multiclass cases,

  the boundary between ω_i and all $\omega_{j \neq i}$ in case 1 is given by

$$d_i(x) = \sum_{j=1}^{n} wa_{ij} x_j + w_{i\ n+1} = 0.$$

The boundary between ω_i and ω_j in case 2 is given by

$$d_{ij}(x) = \sum_{k=1}^{n} w_{ijk} x_k + w_{ij\ n+1} = 0,$$

The boundary between ω_i and ω_j in case 3 is given by

$$\begin{aligned} d_{ij}(x) &= d_i(x) - d_j(x) \\ &= \Sigma_{k=1}^{n} (w_{ik} - w_{jk}) x_k + (w_{i\ n+1} - w_{j\ n+1}) \\ &= 0 \end{aligned}$$

- Geometric interpretation of

$$\begin{aligned} d(x) &= w_1 x_1 + \ldots + w_n x_n + w_{n+1} \\ &= w_0' x + w_{n+1} = 0 \end{aligned}$$

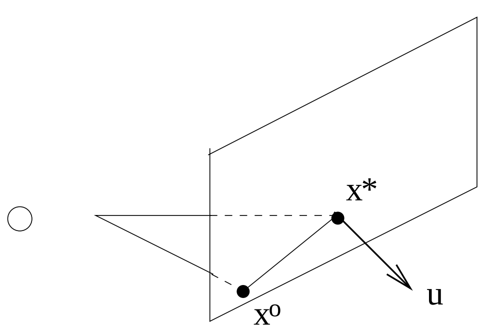
where

$$\begin{aligned} w_0 &= (w_1, \ldots, w_n)', \\ x &= (x_1, \ldots, x_n)'. \end{aligned}$$

Hyperplane: $w_o x + w_{n+1} = 0 \Rightarrow$

$$(\frac{w_0}{||w_0||})' x = -\frac{w_{n+1}}{||w_0||}$$

where $||w_0|| = \sqrt{w_1^2 + \ldots + w_n^2}$.
Let u denote the unit vector $\frac{w_0}{||w_0||}$. Then u is the normal to the hyperplane.
Proof: If x^*, x^o belong to the hyperplane, u is orthogonal to $x^* - x^o$.



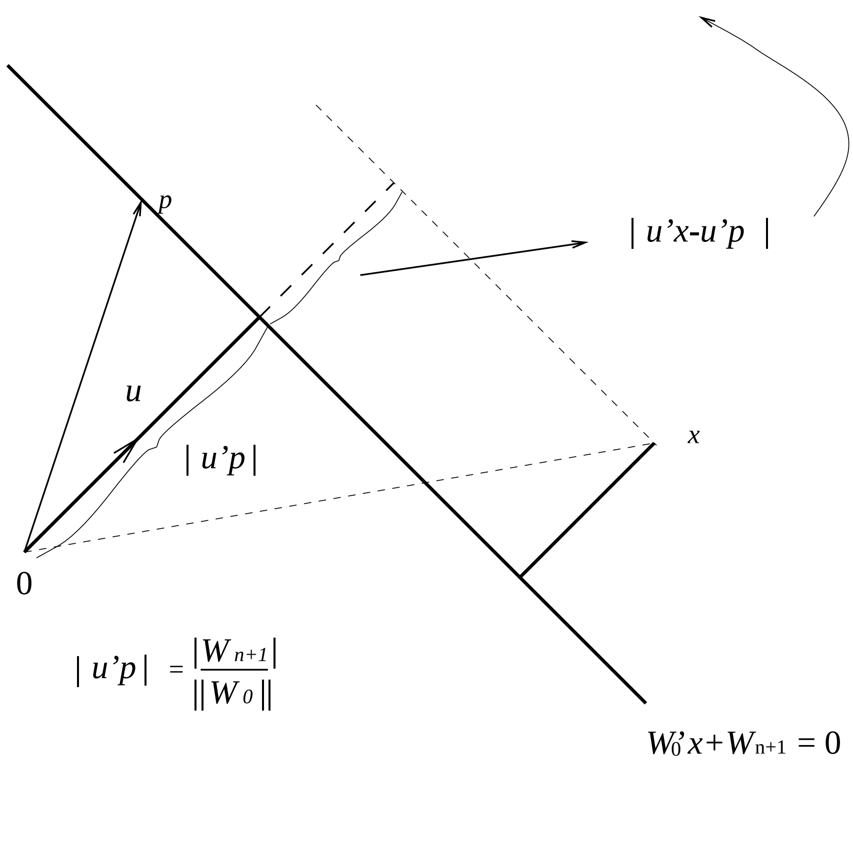
$$\left.\begin{aligned} u'x^* &= -\frac{w_{n+1}}{||w_0||} \\ u'x^o &= -\frac{w_{n+1}}{||w_0||} \end{aligned}\right\} u'(x^* - x^o) = 0$$

The distance from the origin to the hyperplane is given by

$$D_u = \frac{|w_{n+1}|}{||w_0||}.$$

The distance from an arbitrary point x to a hyperplane is given by

$$D_x = |\frac{w_0'x + w_{n+1}}{||w_0||}|$$



Implementation of Decision Functions

- In most of applications, the entire pattern recognition system is implemented in a computer.
- A multiclass linear pattern classifier (Case 3).