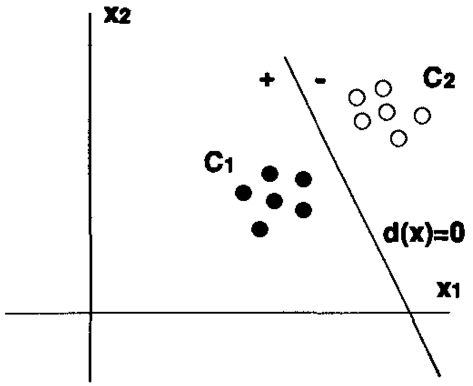
DECISION FUNCTIONS

BASIC CONCEPTS

We start with a simple example. Let C_1 and C_2 be two pattern classes, samples of which are shown in Fig. 2.1.1. Each sample pattern is a vector $\boldsymbol{x} = (x_1, x_2)^T$ in the $x_1 - x_2$ plane, denoted by either $\bullet(\boldsymbol{x} \in C_1)$ or $\circ(\boldsymbol{x} \in C_2)$.



The two populations can be clearly separated by a straight line. Let $d(\boldsymbol{x}) = 0$ be such a line. Then, its coefficients given by

$$d(\boldsymbol{x}) = w_1 x_1 + w_2 x_2 + w_3 = 0 \tag{2.1.1}$$

can be rearranged such that $d(x) > 0$ for all $\boldsymbol{x} \in C_1$ and $d(\boldsymbol{x}) < 0$ for all $\boldsymbol{x} \in C_2$. For any incoming $\boldsymbol{x}$ known *a priori* to belong to either C_1 or C_2, we can calculate $d(\boldsymbol{x})$ and *decide* that $\boldsymbol{x} \in C_1$ if $d(\boldsymbol{x}) > 0$ and $\boldsymbol{x} \in C_2$ if $d(\boldsymbol{x}) < 0$. Thus, $d(\boldsymbol{x})$ is a linear *decision function* of C_1.

This particular example can be easily extended to the case of two pattern classes C_1, C_2 in the n-dimensional Euclidean vector space R^n. Assume the classes to be geometrically separated by the hyperplane

$$d(\boldsymbol{x}) = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n + w_{n+1} = \boldsymbol{w}_0^T \boldsymbol{x} + w_{n+1} = 0 \tag{2.1.2}$$

where $\boldsymbol{w}_0 = (w_1, w_2, \ldots, w_n)^T$ is the *weight* vector, such that

$$d(\boldsymbol{x}) > 0, \quad \text{for} \quad \boldsymbol{x} \in C_1$$
$$d(\boldsymbol{x}) < 0, \quad \text{for} \quad \boldsymbol{x} \in C_2 \tag{2.1.3}$$

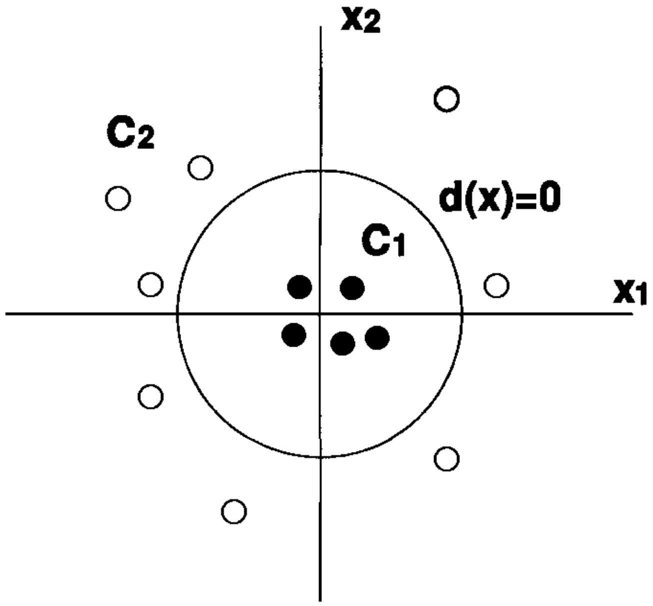
Then, for an arbitrary incoming $\boldsymbol{x}$ at $C_1 \cup C_2$, we can decide

$$d(\boldsymbol{x}) > 0, \Rightarrow \boldsymbol{x} \in C_1$$
$$d(\boldsymbol{x}) > 0, \Rightarrow \boldsymbol{x} \in C_2 \tag{2.1.4}$$

Usually, $\boldsymbol{x}$ and $\boldsymbol{w}_0$ of Eq. (2.1.2) are replaced by the *augmented* pattern and weight vectors $\boldsymbol{x} = (x_1, x_2, \ldots, x_n, 1)^T$ and $\boldsymbol{w} = (w_1, w_2, \ldots, w_{n+1})^T$ for which one gets

$$d(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} \tag{2.1.5}$$

A decision function may not be linear. In Fig. 2.1.2 the two pattern classes are separated by the circumference $d(\boldsymbol{x}) = 1 - x_1^2 - x_2^2 = 0$. Since $d(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in C_1$ and $d(\boldsymbol{x}) < 0$ for all $\boldsymbol{x} \in C_2$, $d(\boldsymbol{x})$ is a *nonlinear* decision function of C_1. The membership of an incoming $\boldsymbol{x}$ in either C_1 or C_2 will be decided by using Eq. (2.1.4).



In general there are m pattern classes $\{C_1, C_2, \ldots, C_m\}$ in R^n and a decision function is defined as follows.

■ **Definition 2.1.1** Let $C_1, C_2, \ldots, C_m$ be m pattern classes in R^n. If a surface $d(\boldsymbol{x}) = 0$, $\boldsymbol{x} \in R^n$ separates between some C_i and the remaining C_j, $j \neq i$, i.e.

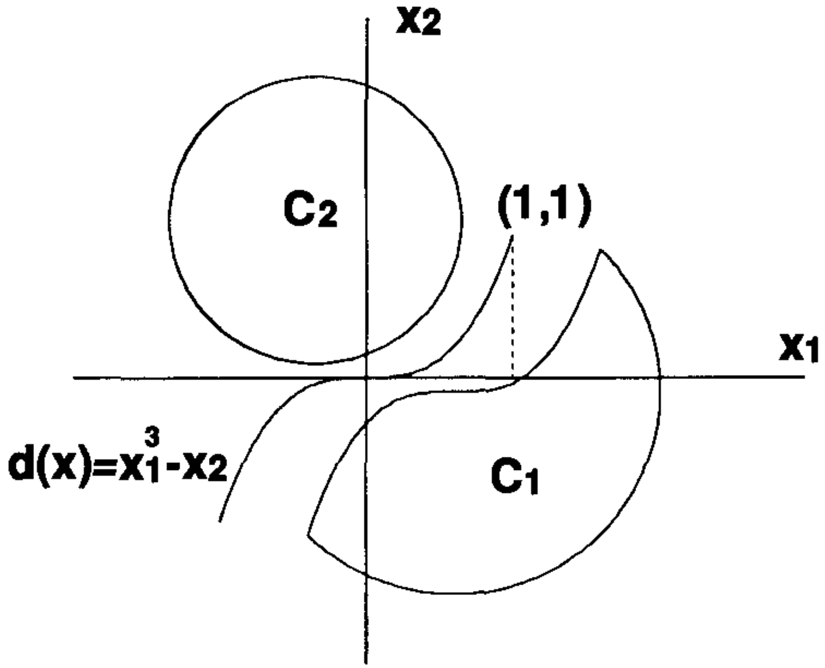
$$\begin{aligned} &d(\boldsymbol{x}) > 0, \quad \boldsymbol{x} \in C_i \\ &d(\boldsymbol{x}) < 0, \quad \boldsymbol{x} \in C_j, \quad j \neq i \end{aligned} \tag{2.1.6}$$

then $d(\boldsymbol{x})$ will be called a *decision function* of C_i.

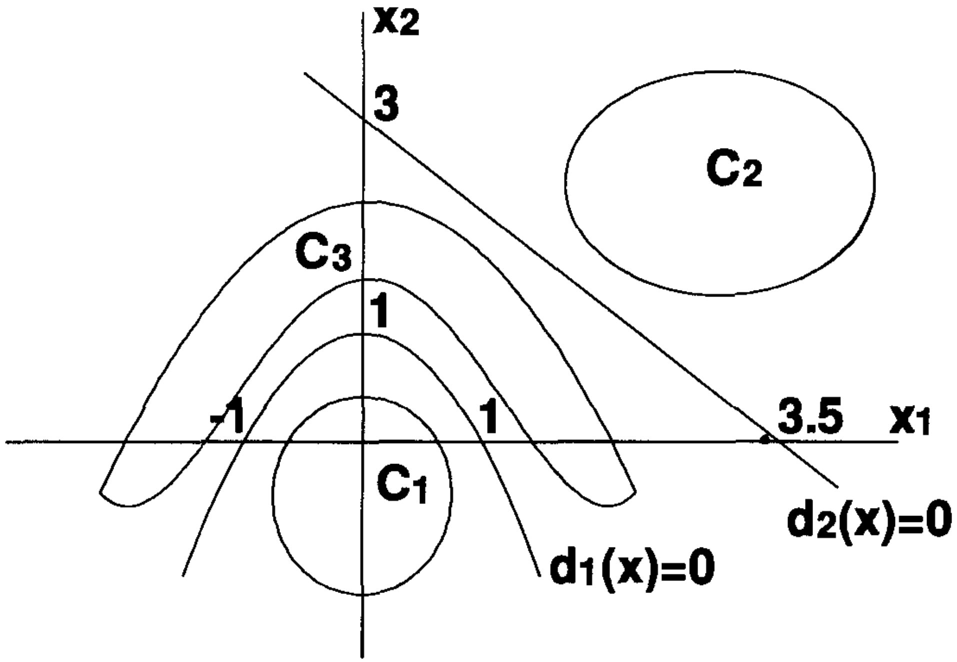
Naturally, the domain of definition for $d(\boldsymbol{x})$ must include the union of $C_1, C_2, \ldots, C_m$.

For the sake of simplicity, pattern classes would be often denoted in figures by the boundaries of the regions where the given sample patterns fall.

■ **Example 2.1.1** Let C_1 and C_2 be the pattern classes of Fig. 2.1.3. The parabola $x_1^3 - x_2 = 0$ is a decision function of C_1. Usually, the number of legitimate decision functions is infinite. In this particular case, $d^*(\boldsymbol{x}) = x_1 - x_2$ is also a possible decision function.



■ **Example 2.1.2** Let C_1, C_2, C_3 be the pattern classes of Fig. 2.1.4. The parabola $d_1(\boldsymbol{x}) = 1 - x_1^2 - x_2 = 0$ is a decision function for C_1, while $d_2(\boldsymbol{x}) = 6x_1 + 7x_2 - 21$ is a decision function of C_2. Unlike in the previous example, a linear decision function for C_1 does not exist.



2.2 LINEAR DECISION FUNCTIONS

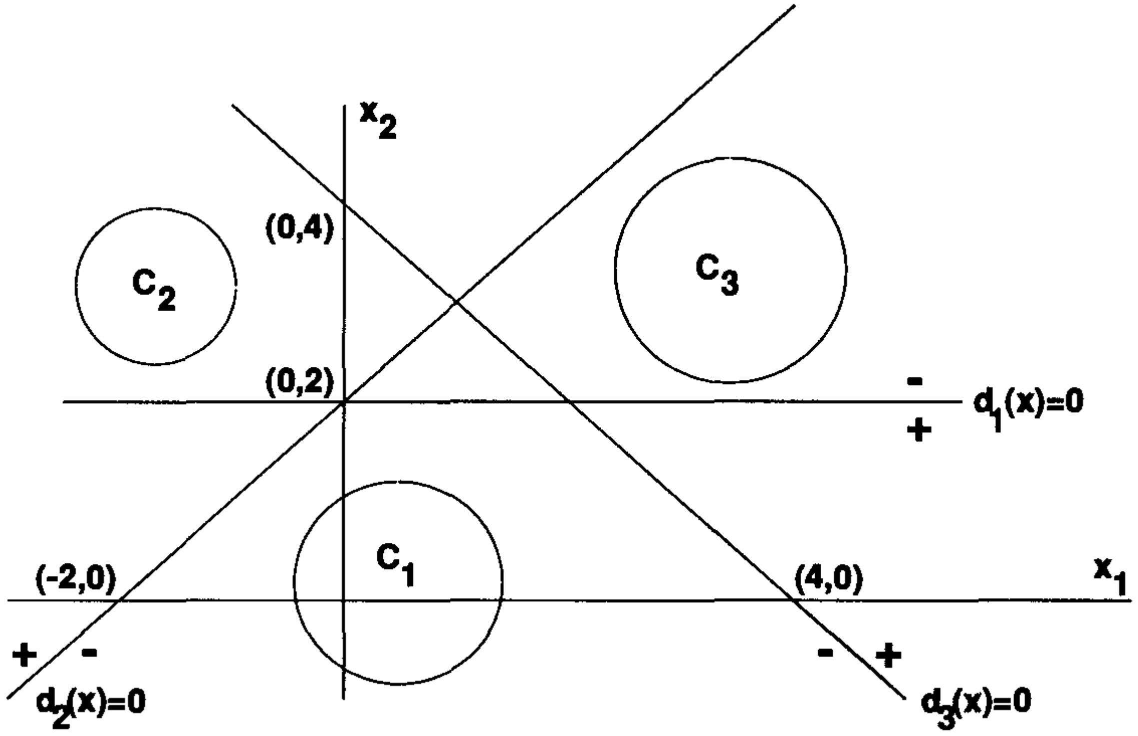
Given m pattern classes $C_1, C_2, \ldots, C_m$ in R^n we distinguish between two cases.

I. Absolute separation

If each pattern class C_i has a linear decision function $d_i(\boldsymbol{x})$, i.e.

$$d_i(\boldsymbol{x}) = \boldsymbol{w}_i^T \boldsymbol{x} = \begin{cases} >0, & \boldsymbol{x} \in C_i \\ <0, & otherwise \end{cases} \tag{2.2.1}$$

■ **Example 2.2.1** Consider the pattern classes C_1, C_2, C_3 in Fig. 2.2.1. The straight lines $d_1(\boldsymbol{x}) = 2 - x_2 = 0$, $d_2(\boldsymbol{x}) = -x_1 + x_2 - 2 = 0$ and $d_3(\boldsymbol{x}) = x_1 + x_2 - 4 = 0$ provide decision functions for C_1, C_2, C_3 respectively , i.e. C_1, C_2 and C_3 are absolutely separable.



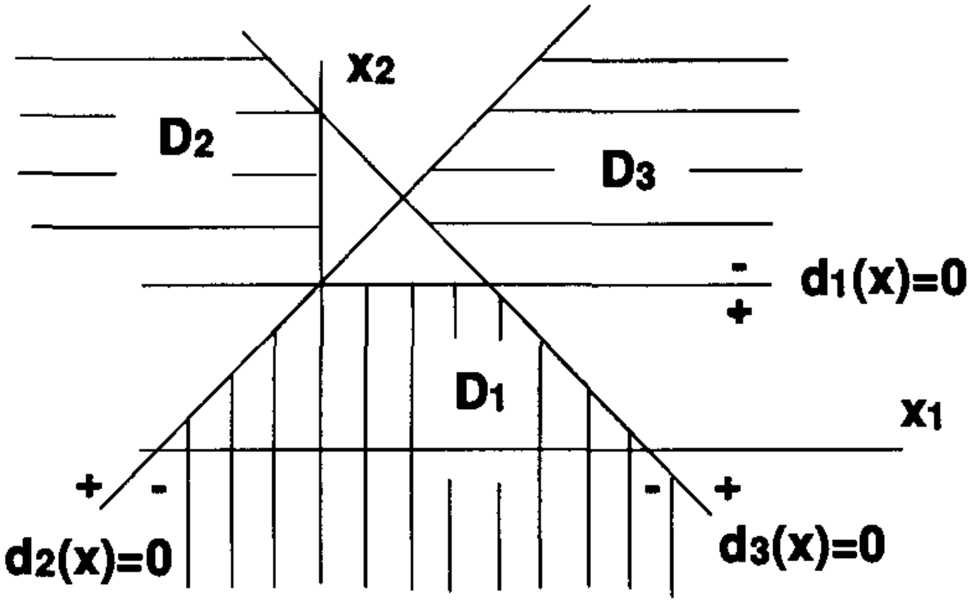
■ **Definition 2.2.1** (decision region): Let the pattern classes $\{C_i\}_{i=1}^{m}$ be absolutely separable by the linear decision functions $d_1(\boldsymbol{x}), d_2(\boldsymbol{x}), \ldots, d_m(\boldsymbol{x})$ respectively. Then the vector sets

$$D_i = \left\{ \boldsymbol{x} \mid d_i(\boldsymbol{x}) > 0;\ \ d_j(\boldsymbol{x}) < 0,\ \ j \neq i \right\},\quad 1 \leq i \leq m \tag{2.2.2}$$

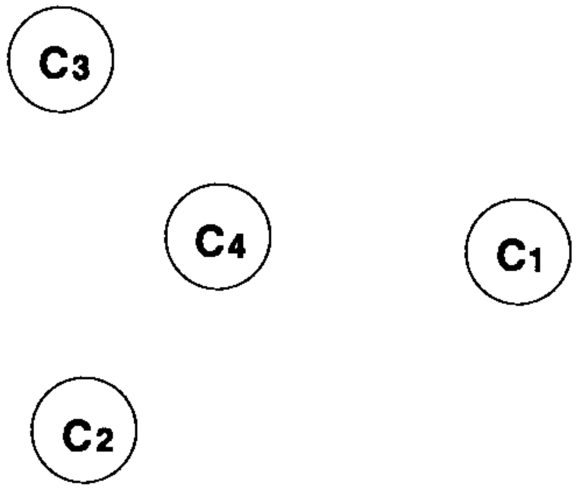
are called the *decision regions* of $C_1, C_2, \ldots, C_m$ respectively.

Note that each pattern class C_i, is a subset of its associated decision region D_i and that decision regions depend directly on the particular choice of decision functions.

■ **Example 2.2.2** The decision regions associated with the previous example are the shaded are as in Fig. 2.2.2.



■ **Example 2.2.3** The pattern classes C_1, C_2, C_3, C_4 in Fig. 2.2.3 are such that no linear decision function exists for C_4. However, any three of the four classes are absolutely separable.



■ **Figure 2.2.3** A case with no absolute separation.

II. Pairwise separation

In the absence of absolute separation, partial separation between pattern classes, can still occur if each pair of them can be separated by a linear decision function. In this case the pattern classes are said to be *pairwise separable*. Each pair of classes C_i and C_j are associated with a linear decision function d_{ij} such that

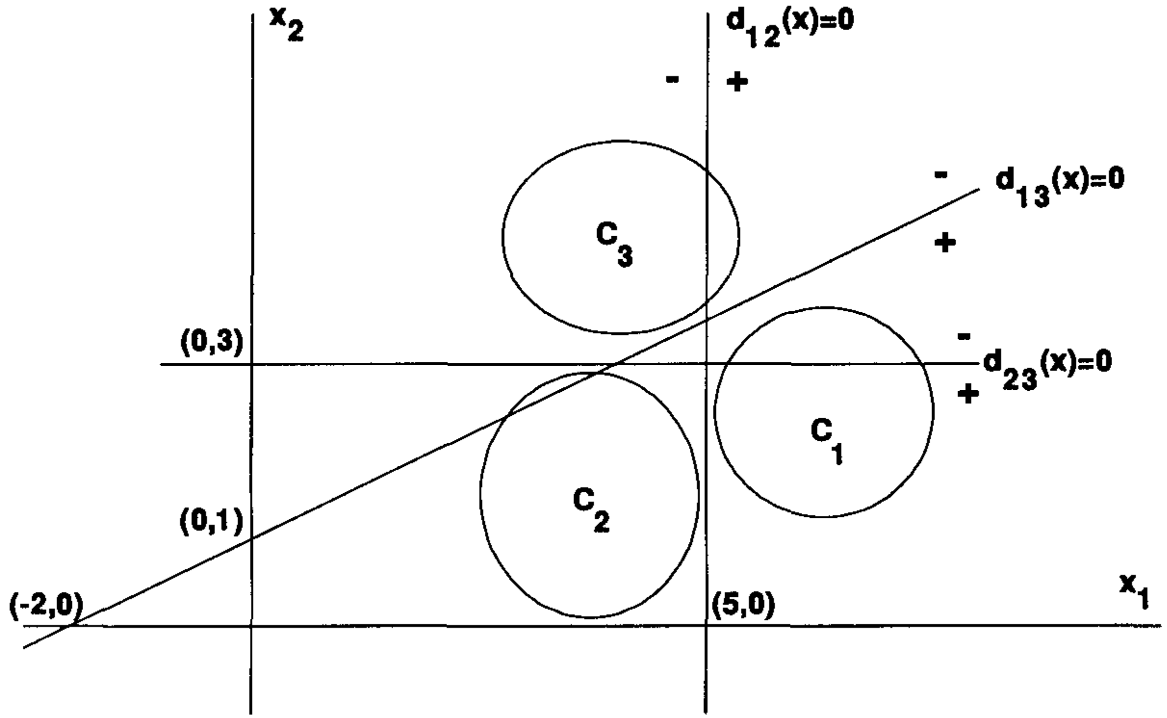
$$\begin{aligned} d_{ij}(\boldsymbol{x}) > 0 \ \ \textit{for all} \ \ \boldsymbol{x} \in C_i \\ d_{ij}(\boldsymbol{x}) < 0 \ \ \textit{for all} \ \ \boldsymbol{x} \in C_j \end{aligned} \qquad (2.2.3)$$

Consequently, for all $\boldsymbol{x} \in C_i$ we have

$$d_{ij}(\boldsymbol{x}) > 0 \ \ \textit{for all} \ \ j \neq i \qquad (2.2.4)$$

Also, for all i and j: $d_{ji}(\boldsymbol{x}) = -d_{ij}(\boldsymbol{x})$.

■ **Example 2.2.4** Let C_1, C_2 and C_3 be the pattern classes shown in Fig. 2.2.4. The linear decision functions $d_{12}(\boldsymbol{x}) = x_1 - 5$, $d_{23}(\boldsymbol{x}) = -x_2 + 3$ and $d_{13}(\boldsymbol{x}) = x_1 - 2x_2 + 2$ separate between the pairs (C_1, C_2), (C_2, C_3) and (C_1, C_3) respectively. Therefore C_1, C_2 and C_3 are pairwise separable.

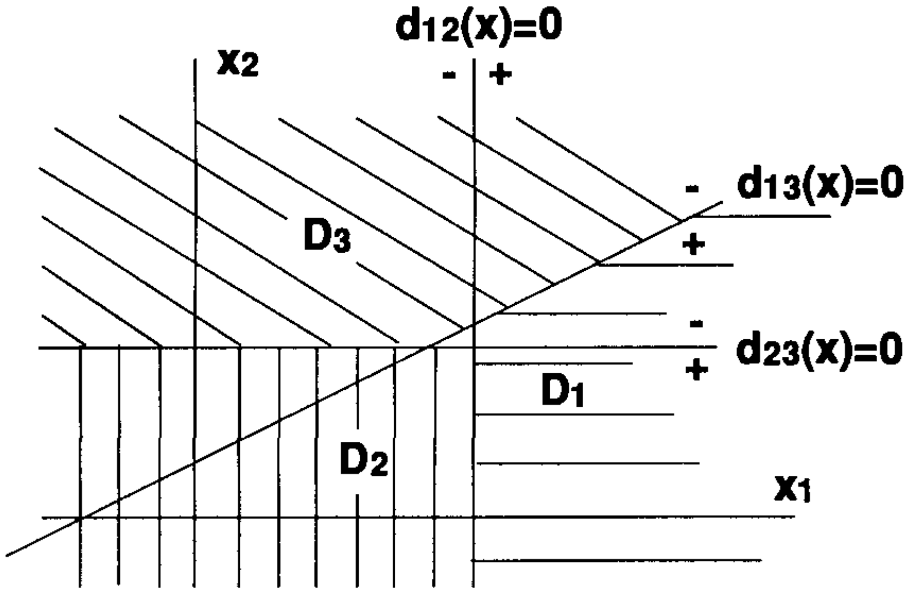


■ **Definition 2.2.2** (decision region). Let the pattern classes $C_1, C_2, \ldots, C_m$ be pairwise separable by the linear decision functions $\{d_{ij}(\boldsymbol{x})\}_{i,j=1}^{m}$. Then the vector sets

$$D_i = \{\boldsymbol{x} \mid d_{ij}(\boldsymbol{x}) > 0,\ j \neq i\},\ 1 \leq i \leq m \tag{2.2.5}$$

are called the decision regions of $C_1, C_2, \ldots, C_m$ respectively.

■ **Example 2.2.5** The decision regions of C_1, C_2, C_3 in the previous example are shown in Fig. 2.2.5. In order to get D_2 we take $d_{21}(\boldsymbol{x}) = -d_{12}(\boldsymbol{x})$, and to obtain D_3 we use $d_{31}(\boldsymbol{x}) = -d_{13}(\boldsymbol{x})$ and $d_{32}(\boldsymbol{x}) = -d_{23}(\boldsymbol{x})$.



■ **Figure 2.2.5** Decision regions for Example 2.2.5.

A common particular case of pairwise separation occurs when linear functions $d_1(\boldsymbol{x}),\ d_2(\boldsymbol{x}),\ldots,d_m(\boldsymbol{x})$ such that for all $\boldsymbol{x} \in C_i,\ \ 1 \le i \le m$

$$d_i(\boldsymbol{x}) > d_j(\boldsymbol{x}) \ \ \textit{for all} \ \ j \ne i \qquad (2.2.6)$$

exist. It is easily seen that by defining $d_{ij}(\boldsymbol{x}) = d_i(\boldsymbol{x}) - d_j(\boldsymbol{x})$ for $1 \le i,\ j \le m$, we obtain a case of pairwise separation. However, the union of the decision regions is now the whole space, i.e. no ambiguous region exists. Indeed, for any incoming pattern $\boldsymbol{x}$ we can find i for which

$$d_i(\boldsymbol{x}) = \max\,[d_j(\boldsymbol{x})],\ 1 \le j \le m \qquad (2.2.7)$$

and then classify $\boldsymbol{x}$ as a member of C_i. If the maximum is achieved for several i's we choose (for example) the smallest.

If Eq. (2.2.6) holds, there is a simple geometric interpretation to the empty ambiguous region: The straight lines $d_{12}(\boldsymbol{x})$, $d_{23}(\boldsymbol{x})$, $d_{13}(\boldsymbol{x})$, intersect at one point. Indeed, if $d_1(\boldsymbol{x}) - d_2(\boldsymbol{x}) = 0$ and $d_2(\boldsymbol{x}) - d_3(\boldsymbol{x}) = 0$ then clearly $d_1(\boldsymbol{x}) - d_3(\boldsymbol{x}) = 0$ as well.