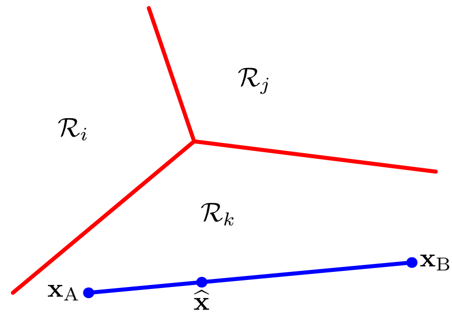

Figure 5.3 Illustration of the decision regions for a multi-class linear discriminant, with the decision boundaries shown in red. If two points $\mathbf{x}_\mathrm{A}$ and $\mathbf{x}_\mathrm{B}$ both lie inside the same decision region $\mathcal{R}_k$, then any point $\widehat{\mathbf{x}}$ that lies on the line connecting these two points must also lie in $\mathcal{R}_k$, and hence, the decision region must be singly connected and convex.

hence corresponds to a $(D-1)$-dimensional hyperplane defined by

$$(\mathbf{w}_k - \mathbf{w}_j)^\mathrm{T}\mathbf{x} + (w_{k0} - w_{j0}) = 0. \tag{5.8}$$

This has the same form as the decision boundary for the two-class case discussed in Section 5.1.1, and so analogous geometrical properties apply.

The decision regions of such a discriminant are always singly connected and convex. To see this, consider two points $\mathbf{x}_\mathrm{A}$ and $\mathbf{x}_\mathrm{B}$ both of which lie inside decision region $\mathcal{R}_k$, as illustrated in Figure 5.3. Any point $\widehat{\mathbf{x}}$ that lies on the line connecting $\mathbf{x}_\mathrm{A}$ and $\mathbf{x}_\mathrm{B}$ can be expressed in the form

$$\widehat{\mathbf{x}} = \lambda\mathbf{x}_\mathrm{A} + (1-\lambda)\mathbf{x}_\mathrm{B} \tag{5.9}$$

where $0 \leqslant \lambda \leqslant 1$. From the linearity of the discriminant functions, it follows that

$$y_k(\widehat{\mathbf{x}}) = \lambda y_k(\mathbf{x}_\mathrm{A}) + (1-\lambda)y_k(\mathbf{x}_\mathrm{B}). \tag{5.10}$$

Because both $\mathbf{x}_\mathrm{A}$ and $\mathbf{x}_\mathrm{B}$ lie inside $\mathcal{R}_k$, it follows that $y_k(\mathbf{x}_\mathrm{A}) > y_j(\mathbf{x}_\mathrm{A})$ and that $y_k(\mathbf{x}_\mathrm{B}) > y_j(\mathbf{x}_\mathrm{B})$, for all $j \neq k$, and hence $y_k(\widehat{\mathbf{x}}) > y_j(\widehat{\mathbf{x}})$, and so $\widehat{\mathbf{x}}$ also lies inside $\mathcal{R}_k$. Thus, $\mathcal{R}_k$ is singly connected and convex.

Note that for two classes, we can either employ the formalism discussed here, based on two discriminant functions $y_1(\mathbf{x})$ and $y_2(\mathbf{x})$, or else use the simpler but essentially equivalent formulation based on a single discriminant function $y(\mathbf{x})$.

Section 5.1.1

5.1.3 1-of-K coding

For regression problems, the target variable $\mathbf{t}$ was simply the vector of real numbers whose values we wish to predict. In classification, there are various ways of using target values to represent class labels. For two-class problems, the most convenient is the binary representation in which there is a single target variable $t \in \{0, 1\}$ such that $t = 1$ represents class $\mathcal{C}_1$ and $t = 0$ represents class $\mathcal{C}_2$. We can interpret the value of t as the probability that the class is $\mathcal{C}_1$, with the probability values taking only the extreme values of 0 and 1. For $K > 2$ classes, it is convenient to use a 1-of-K coding scheme, also known as the one-hot encoding scheme, in which $\mathbf{t}$ is a vector of length K such that if the class is $\mathcal{C}_j$, then all elements t_k of $\mathbf{t}$ are zero