



Figure 5.13 The left-hand plot shows the class-conditional densities for two classes, denoted red and blue. On the right is the corresponding posterior probability $p(C_1|\mathbf{x})$, which is given by a logistic sigmoid of a linear function of \mathbf{x} . The surface in the right-hand plot is coloured using a proportion of red ink given by $p(C_1|\mathbf{x})$ and a proportion of blue ink given by $p(C_2|\mathbf{x}) = 1 - p(C_1|\mathbf{x})$.

are constant and so will be given by linear functions of \mathbf{x} , and therefore the decision boundaries are linear in input space. The prior probabilities $p(C_k)$ enter only through the bias parameter w_0 , so that changes in the priors have the effect of making parallel shifts of the decision boundary and more generally of the parallel contours of constant posterior probability.

For the general case of K classes, the posterior probabilities are given by (5.45) where, from (5.46) and (5.47), we have

$$a_k(\mathbf{x}) = \mathbf{w}_k^T \mathbf{x} + w_{k0} \quad (5.51)$$

in which we have defined

$$\mathbf{w}_k = \Sigma^{-1} \boldsymbol{\mu}_k \quad (5.52)$$

$$w_{k0} = -\frac{1}{2} \boldsymbol{\mu}_k^T \Sigma^{-1} \boldsymbol{\mu}_k + \ln p(C_k). \quad (5.53)$$

We see that the $a_k(\mathbf{x})$ are again linear functions of \mathbf{x} as a consequence of the cancellation of the quadratic terms due to the shared covariances. The resulting decision boundaries, corresponding to the minimum misclassification rate, will occur when two of the posterior probabilities (the two largest) are equal, and so will be defined by linear functions of \mathbf{x} . Thus, we again have a generalized linear model.

If we relax the assumption of a shared covariance matrix and allow each class-conditional density $p(\mathbf{x}|C_k)$ to have its own covariance matrix Σ_k , then the earlier cancellations will no longer occur, and we will obtain quadratic functions of \mathbf{x} , giving rise to a *quadratic discriminant*. The linear and quadratic decision boundaries are illustrated in Figure 5.14.

5.3.2 Maximum likelihood solution

Once we have specified a parametric functional form for the class-conditional densities $p(\mathbf{x}|C_k)$, we can then determine the values of the parameters, together with