

the image, and the output variable  $t$  will represent the absence of cancer, which we denote by the class  $C_1$ , or the presence of cancer, which we denote by the class  $C_2$ . We might, for instance, choose  $t$  to be a binary variable such that  $t = 0$  corresponds to class  $C_1$  and  $t = 1$  corresponds to class  $C_2$ . We will see later that this choice of label values is particularly convenient when working with probabilities. The general inference problem then involves determining the joint distribution  $p(\mathbf{x}, C_k)$ , or equivalently  $p(\mathbf{x}, t)$ , which gives us the most complete probabilistic description of the variables. Although this can be a very useful and informative quantity, ultimately, we must decide either to give treatment to the patient or not, and we would like this choice to be optimal according to some appropriate criterion (Duda and Hart, 1973). This is the *decision* step, and the aim of decision theory is that it should tell us how to make optimal decisions given the appropriate probabilities. We will see that the decision stage is generally very simple, even trivial, once we have solved the inference problem. Here we give an introduction to the key ideas of decision theory as required for the rest of the book. Further background, as well as more detailed accounts, can be found in Berger (1985) and Bather (2000).

Before giving a more detailed analysis, let us first consider informally how we might expect probabilities to play a role in making decisions. When we obtain the skin image  $\mathbf{x}$  for a new patient, our goal is to decide which of the two classes to assign the image to. We are therefore interested in the probabilities of the two classes, given the image, which are given by  $p(C_k|\mathbf{x})$ . Using Bayes' theorem, these probabilities can be expressed in the form

$$p(C_k|\mathbf{x}) = \frac{p(\mathbf{x}|C_k)p(C_k)}{p(\mathbf{x})}. \quad (5.19)$$

Note that any of the quantities appearing in Bayes' theorem can be obtained from the joint distribution  $p(\mathbf{x}, C_k)$  by either marginalizing or conditioning with respect to the appropriate variables. We can now interpret  $p(C_k)$  as the prior probability for the class  $C_k$  and  $p(C_k|\mathbf{x})$  as the corresponding posterior probability. Thus,  $p(C_1)$  represents the probability that a person has cancer, before the image is taken. Similarly,  $p(C_1|\mathbf{x})$  is the posterior probability, revised using Bayes' theorem in light of the information contained in the image. If our aim is to minimize the chance of assigning  $\mathbf{x}$  to the wrong class, then intuitively we would choose the class having the higher posterior probability. We now show that this intuition is correct, and we also discuss more general criteria for making decisions.

### 5.2.1 Misclassification rate

Suppose that our goal is simply to make as few misclassifications as possible. We need a rule that assigns each value of  $\mathbf{x}$  to one of the available classes. Such a rule will divide the input space into regions  $\mathcal{R}_k$  called *decision regions*, one for each class, such that all points in  $\mathcal{R}_k$  are assigned to class  $C_k$ . The boundaries between decision regions are called *decision boundaries* or *decision surfaces*. Note that each decision region need not be contiguous but could comprise some number of disjoint regions. To find the optimal decision rule, consider first the case of two classes, as in the cancer problem, for instance. A mistake occurs when an input vector belonging