

If we used such a data set to train an adaptive model, we could run into severe difficulties due to the small proportion of those in the cancer class. For instance, a classifier that assigned every point to the normal class would achieve 99.9% accuracy, and it may be difficult to avoid this trivial solution. Also, even a large data set will contain very few examples of skin images corresponding to cancer, and so the learning algorithm will not be exposed to a broad range of examples of such images and hence is not likely to generalize well. A balanced data set with equal numbers of examples from each of the classes would allow us to find a more accurate model. However, we then have to compensate for the effects of our modifications to the training data. Suppose we have used such a modified data set and found models for the posterior probabilities. From Bayes' theorem (5.24), we see that the posterior probabilities are proportional to the prior probabilities, which we can interpret as the fractions of points in each class. We can therefore simply take the posterior probabilities obtained from our artificially balanced data set, divide by the class fractions in that data set, and then multiply by the class fractions in the population to which we wish to apply the model. Finally, we need to normalize to ensure that the new posterior probabilities sum to one. Note that this procedure cannot be applied if we have learned a discriminant function directly instead of determining posterior probabilities.

Combining models. For complex applications, we may wish to break the problem into a number of smaller sub-problems each of which can be tackled by a separate module. For example, in our hypothetical medical diagnosis problem, we may have information available from, say, blood tests as well as skin images. Rather than combine all of this heterogeneous information into one huge input space, it may be more effective to build one system to interpret the images and a different one to interpret the blood data. If each of the two models gives posterior probabilities for the classes, then we can combine the outputs systematically using the rules of probability. One simple way to do this is to assume that, for each class separately, the distributions of inputs for the images, denoted by \mathbf{x}_I , and the blood data, denoted by \mathbf{x}_B , are independent, so that

$$p(\mathbf{x}_I, \mathbf{x}_B | \mathcal{C}_k) = p(\mathbf{x}_I | \mathcal{C}_k)p(\mathbf{x}_B | \mathcal{C}_k). \quad (5.26)$$

Section 11.2

This is an example of a *conditional independence* property, because the independence holds when the distribution is conditioned on the class \mathcal{C}_k . The posterior probability, given both the image and blood data, is then given by

$$\begin{aligned} p(\mathcal{C}_k | \mathbf{x}_I, \mathbf{x}_B) &\propto p(\mathbf{x}_I, \mathbf{x}_B | \mathcal{C}_k)p(\mathcal{C}_k) \\ &\propto p(\mathbf{x}_I | \mathcal{C}_k)p(\mathbf{x}_B | \mathcal{C}_k)p(\mathcal{C}_k) \\ &\propto \frac{p(\mathcal{C}_k | \mathbf{x}_I)p(\mathcal{C}_k | \mathbf{x}_B)}{p(\mathcal{C}_k)}. \end{aligned} \quad (5.27)$$

Thus, we need the class prior probabilities $p(\mathcal{C}_k)$, which we can easily estimate from the fractions of data points in each class, and then we need to normalize