



Figure 2.1 An extension of the simple sine curve regression problem to two dimensions. (a) A plot of the function $y(x_1, x_2) = \sin(2\pi x_1)\sin(2\pi x_2)$. Data is generated by selecting values for x_1 and x_2 , computing the corresponding value of $y(x_1, x_2)$, and then adding Gaussian noise. (b) Plot of 100 data points in which x_2 is unobserved showing high levels of noise. (c) Plot of 100 data points in which x_2 is fixed to the value $x_2 = \frac{\pi}{2}$, simulating the effect of being able to measure x_2 as well as x_1 , showing much lower levels of noise.

Section 1.2

using an extension of the sine curve example to two dimensions in Figure 2.1.

As a practical example of this, a biopsy sample of the skin lesion is much more informative than the image alone and might greatly improve the accuracy with which we can determine if a new lesion is malignant. Given both the image and the biopsy data, the intrinsic uncertainty might be very small, and by collecting a large training data set, we may be able to reduce the systematic uncertainty to a low level and thereby make predictions of the class of the lesion with high accuracy.

Both kinds of uncertainty can be handled using the framework of *probability theory*, which provides a consistent paradigm for the quantification and manipulation of uncertainty and therefore forms one of the central foundations for machine learning. We will see that probabilities are governed by two simple formulae known as the *sum rule* and the *product rule*. When coupled with *decision theory*, these rules allow us, at least in principle, to make optimal predictions given all the information available to us, even though that information may be incomplete or ambiguous.

Section 2.1

Section 5.2

The concept of probability is often introduced in terms of frequencies of repeatable events. Consider, for example, the bent coin shown in Figure 2.2, and suppose that the shape of the coin is such that if it is flipped a large number of times, it lands concave side up 60% of the time, and therefore lands convex side up 40% of the time. We say that the *probability* of landing concave side up is 60% or 0.6. Strictly, the probability is defined in the limit of an infinite number of ‘trials’ or coin flips in this case. Because the coin must land either concave side up or convex side up, these probabilities add to 100% or 1.0. This definition of probability in terms of the frequency of repeatable events is the basis for the *frequentist* view of statistics.

Now suppose that, although we know that the probability that the coin will land concave side up is 0.6, we are not allowed to look at the coin itself and we do not