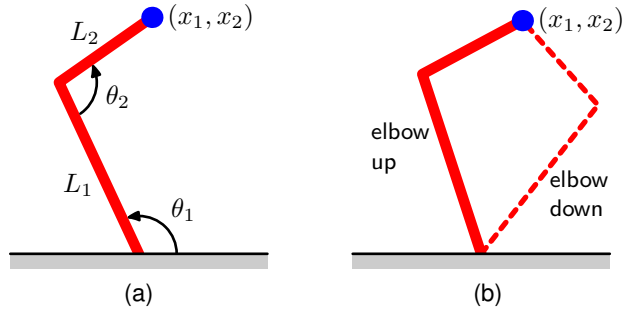


Figure 6.16 (a) A two-link robot arm, in which the Cartesian coordinates (x_1, x_2) of the end effector are determined uniquely by the two joint angles θ_1 and θ_2 and the (fixed) lengths L_1 and L_2 of the arms. This is known as the *forward kinematics* of the arm. (b) In practice, we have to find the joint angles that will give rise to a desired end effector position. This *inverse kinematics* has two solutions corresponding to ‘elbow up’ and ‘elbow down’.



In the robotics example, the kinematics is defined by geometrical equations, and the multimodality is readily apparent. However, in many machine learning problems the presence of multimodality, particularly in problems involving spaces of high dimensionality, can be less obvious. For tutorial purposes, however, we will consider a simple toy problem for which we can easily visualize the multimodality. The data for this problem is generated by sampling a variable x uniformly over the interval $(0, 1)$, to give a set of values $\{x_n\}$, and the corresponding target values t_n are obtained by computing the function $x_n + 0.3 \sin(2\pi x_n)$ and then adding uniform noise over the interval $(-0.1, 0.1)$. The inverse problem is then obtained by keeping the same data points but exchanging the roles of x and t . Figure 6.17 shows the data sets for the forward and inverse problems, along with the results of fitting two-layer neural networks having six hidden units and a single linear output unit by minimizing a sum-of-squares error function. Least squares corresponds to maximum likelihood under a Gaussian assumption. We see that this leads to a good model for the forward problem but a very poor model for the highly non-Gaussian inverse problem.

6.5.2 Conditional mixture distributions

We therefore seek a general framework for modelling conditional probability distributions. This can be achieved by using a mixture model for $p(t|x)$ in which

Figure 6.17 On the left is the data set for a simple forward problem in which the red curve shows the result of fitting a two-layer neural network by minimizing the sum-of-squares error function. The corresponding inverse problem, shown on the right, is obtained by exchanging the roles of x and t . Here the same network, again trained by minimizing the sum-of-squares error function, gives a poor fit to the data due to the multimodality of the data set.

