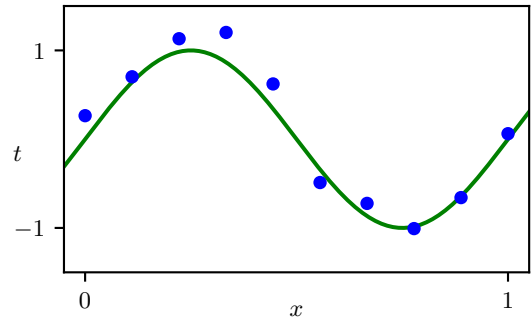


Figure 1.4 Plot of a training data set of $N = 10$ points, shown as blue circles, each comprising an observation of the input variable x along with the corresponding target variable t . The green curve shows the function $\sin(2\pi x)$ used to generate the data. Our goal is to predict the value of t for some new value of x , without knowledge of the green curve.



derived labels. Since large volumes of text are available from multiple sources, this approach allows for scaling to very large training sets and associated very large neural networks.

Large language models can exhibit extraordinary capabilities that have been described as the first indications of emerging artificial general intelligence (Bubeck *et al.*, 2023), and we discuss such models at length later in the book. On the next page, we give an illustration of language generation, based on a model called GPT-4 (OpenAI, 2023), in response to an input prompt ‘Write a proof of the fact that there are infinitely many primes; do it in the style of a Shakespeare play through a dialogue between two parties arguing over the proof.’.

Chapter 12

1.2. A Tutorial Example

For the newcomer to the field of machine learning, many of the basic concepts and much of the terminology can be introduced in the context of a simple example involving the fitting of a polynomial to a small synthetic data set (Bishop, 2006). This is a form of supervised learning problem in which we would like to make a prediction for a target variable, given the value of an input variable.

1.2.1 Synthetic data

We denote the input variable by x and the target variable by t , and we assume that both variables take continuous values on the real axis. Suppose that we are given a training set comprising N observations of x , written x_1, \dots, x_N , together with corresponding observations of the values of t , denoted t_1, \dots, t_N . Our goal is to predict the value of t for some new value of x . The ability to make accurate predictions on previously unseen inputs is a key goal in machine learning and is known as *generalization*.

We can illustrate this using a synthetic data set generated by sampling from a sinusoidal function. Figure 1.4 shows a plot of a training set comprising $N = 10$ data points in which the input values were generated by choosing values of x_n , for $n = 1, \dots, N$, spaced uniformly in the range $[0, 1]$. The associated target data values were obtained by first computing the values of the function $\sin(2\pi x)$ for each value of x