Figure 1.2: Basic setup of the learning problem

chooses g that best matches f on the *training* examples of previous customers, with the hope that it will continue to match f on new customers. Whether or not this hope is justified remains to be seen. Figure 1.2 illustrates the components of the learning problem.

Exercise 1.1

Express each of the following tasks in the framework of learning from data by specifying the input space $\mathcal{X}$, output space $\mathcal{Y}$, target function $f\colon \mathcal{X} \to \mathcal{Y}$, and the specifics of the data set that we will learn from.

(a) Medical diagnosis: A patient walks in with a medical history and some symptoms, and you want to identify the problem.

(b) Handwritten digit recognition (for example postal zip code recognition for mail sorting).

(c) Determining if an email is spam or not.

(d) Predicting how an electric load varies with price, temperature, and day of the week.

(e) A problem of interest to you for which there is no analytic solution, but you have data from which to construct an empirical solution.