

## Chapter 6

# Deep Feedforward Networks

**Deep feedforward networks**, also often called **feedforward neural networks**, or **multilayer perceptrons** (MLPs), are the quintessential deep learning models. The goal of a feedforward network is to approximate some function  $f^*$ . For example, for a classifier,  $y = f^*(\mathbf{x})$  maps an input  $\mathbf{x}$  to a category  $y$ . A feedforward network defines a mapping  $\mathbf{y} = f(\mathbf{x}; \boldsymbol{\theta})$  and learns the value of the parameters  $\boldsymbol{\theta}$  that result in the best function approximation.

These models are called **feedforward** because information flows through the function being evaluated from  $\mathbf{x}$ , through the intermediate computations used to define  $f$ , and finally to the output  $\mathbf{y}$ . There are no **feedback** connections in which outputs of the model are fed back into itself. When feedforward neural networks are extended to include feedback connections, they are called **recurrent neural networks**, presented in chapter 10.

Feedforward networks are of extreme importance to machine learning practitioners. They form the basis of many important commercial applications. For example, the convolutional networks used for object recognition from photos are a specialized kind of feedforward network. Feedforward networks are a conceptual stepping stone on the path to recurrent networks, which power many natural language applications.

Feedforward neural networks are called **networks** because they are typically represented by composing together many different functions. The model is associated with a directed acyclic graph describing how the functions are composed together. For example, we might have three functions  $f^{(1)}$ ,  $f^{(2)}$ , and  $f^{(3)}$  connected in a chain, to form  $f(\mathbf{x}) = f^{(3)}(f^{(2)}(f^{(1)}(\mathbf{x})))$ . These chain structures are the most commonly used structures of neural networks. In this case,  $f^{(1)}$  is called the **first layer** of the network,  $f^{(2)}$  is called the **second layer**, and so on. The overall