Review

In this chapter, we covered the basics of hyperparameter optimization, the process of selecting values for model parameters that can't be learned automatically on the training data. In particular, we introduced random and grid hyperparameter search and demonstrated the use of such code for optimizing models on the Tox21 dataset introduced in the last chapter.

In Chapter 6, we will return to our survey of deep architectures and introduce you to convolutional neural networks, one of the fundamental building blocks of modern deep architectures.

CHAPTER 6

Convolutional Neural Networks

Convolutional neural networks allow deep networks to learn functions on structured spatial data such as images, video, and text. Mathematically, convolutional networks provide tools for exploiting the local structure of data effectively. Images satisfy certain natural statistical properties. Let's assume we represent an image as a two-dimensional grid of pixels. Parts of an image that are close to one other in the pixel grid are likely to vary together (for example, all pixels corresponding to a table in the image are probably brown). Convolutional networks learn to exploit this natural covariance structure in order to learn effectively.

Convolutional networks are a relatively old invention. Versions of convolutional networks have been proposed in the literature dating back to the 1980s. While the designs of these older convolutional networks were often quite sound, they required resources that exceeded hardware available at the time. As a result, convolutional networks languished in relative obscurity in the research literature.

This trend reversed dramatically following the 2012 ILSVRC challenge for object detection in images, where the convolutional AlexNet achieved error rates half that of its nearest competitors. AlexNet was able to use GPUs to train old convolutional architectures on dramatically larger datasets. This combination of old architectures with new hardware allowed AlexNet to dramatically outperform the state of the art in image object detection. This trend has only continued, with convolutional neural networks achieving tremendous boosts over other technologies for processing images. It isn't an exaggeration to say that nearly all modern image processing pipelines are now powered by convolutional neural networks.

There has also been a renaissance in convolutional network design that has moved convolutional networks well past the basic models from the 1980s. For one, networks have been getting much deeper with powerful state-of-the-art networks reaching hundreds of layers deep. Another broad trend has been toward generalizing convolu-