CHAPTER 9

Python

Python is one of the preferred languages for data science in the industry primarily because of its simple syntax and the number of reusable machine learning/deep learning packages. These packages make it easy to develop data science products without getting bogged down with the internals of a particular algorithm or method. They have been written, debugged, and tested by the best experts in the field, as well as by a large supporting community of developers that contribute their time and expertise to maintain and improve them.

In this section, we will go through the foundations of programming with Python 3. This section forms a framework for working with higher-level packages such as NumPy, Pandas, Matplotlib, TensorFlow, and Keras. The programming paradigm we will cover in this chapter can be easily adapted or applied to similar languages, such as R, which is also commonly used in the data science industry.

The best way to work through this chapter and the successive chapters in this part is to work through the code by executing them on Google Colab or GCP Deep Learning VMs.

Data and Operations

Fundamentally, programming involves storing data and operating on that data to generate information. Techniques for efficient data storage are studied in the field called data structures, while the techniques for operating on data are studied as algorithms.

Data is stored in a memory block on the computer. Think of a memory block as a container holding data (Figure 9-1). When data is operated upon, the newly processed data is also stored in memory. Data is operated by using arithmetic and boolean expressions and functions.