

CHAPTER 14

Principles of Learning

Machine learning is, for the most part, sub-divided into three components based on the approach to the learning problem. The three predominant categories of learning are the supervised, unsupervised, and reinforcement learning schemes. In this chapter, we will go over supervised learning schemes in detail and also touch upon unsupervised and reinforcement learning schemes to a lesser extent.

The focus on supervised learning is for a variety of reasons. Firstly, they are the predominant techniques used for building machine learning products in industry; secondly, as you will soon learn, they are easy to ground truth and assess their performances before being deployed as part of a large-scale production pipeline. Let's examine each of the three schemes.

Supervised Learning

To easily understand the concept of supervised learning, let's revisit the problem of identifying spam emails from a set of emails. We will use this example to understand key concepts that are central to the definition and the framing of a supervised learning problem, and they are

- Features
- Samples
- Targets

For this contrived example, let's assume that we have a dictionary of the top 4 words in the set of emails and we record the frequency of occurrence for each email sample. This information is represented in a tabular format, where each feature is a column and the rows are email samples. This tabular representation is called a dataset. Figure 14-1 illustrates this depiction.