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.

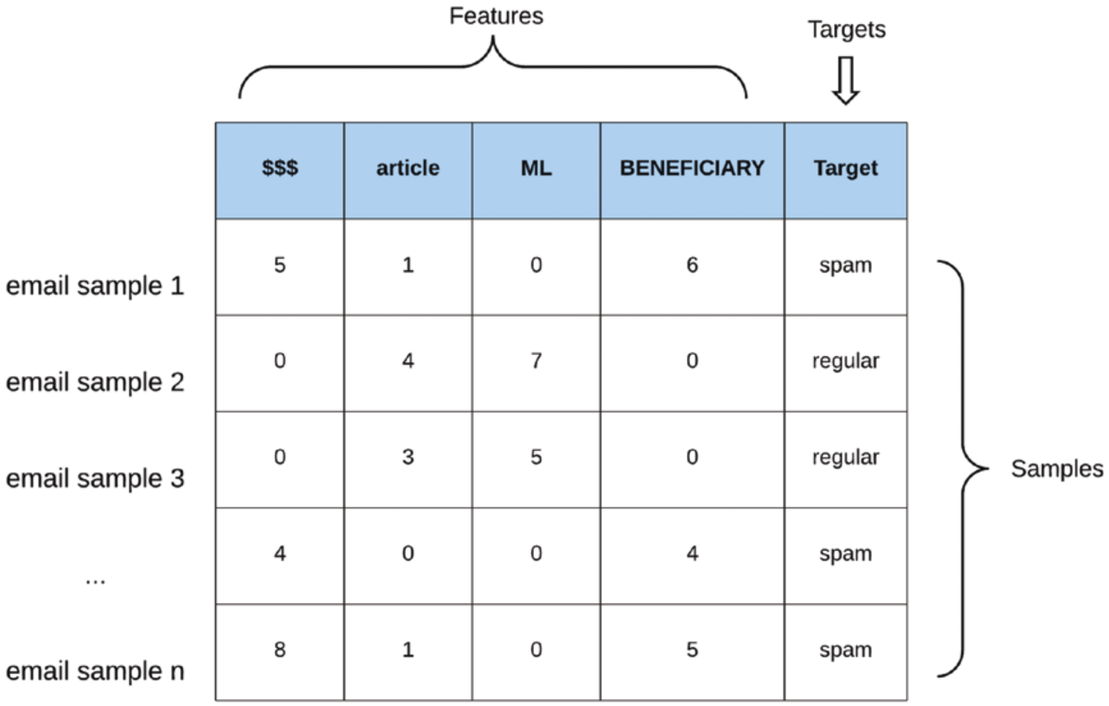
Supervised Learning

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.

E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform*,
https://doi.org/10.1007/978-1-4842-4470-8_14


Features
Targets
Samples

	$$$	article	ML	BENEFICIARY	Target
email sample 1	5	1	0	6	spam
email sample 2	0	4	7	0	regular
email sample 3	0	3	5	0	regular
...	4	0	0	4	spam
email sample n	8	1	0	5	spam

Figure 14-1. *Dataset representation*

The fundamental concept behind supervised machine learning is that each sample is associated with a target variable, and the goal is to teach the computer to learn the patterns from the dataset features that results in a target as a prediction outcome. The columns of a dataset in machine learning are referred to as features; other names you may find commonly used are **variables** or **attributes** of the dataset, but in this book, we will use the term features to describe the measurement units of a data sample. Moreover, the samples of a dataset are also referred to as rows, data points, or observations, but we will use the term samples throughout this book.

Hence, in supervised learning, a set of features are used to build a learning model that will predict the outcome of a target variable as shown in Figure 14-1.

Next, we will cover important modeling considerations for building supervised learning models.

Regression vs. Classification

In supervised learning, we typically have two types of modeling task, and they are **regression** and **classification**.