

CHAPTER 15

Batch vs. Online Learning

Data is a vital component for building learning models. There are two design choices for how data is used in the modeling pipeline. The first is to build your learning model with data at rest (batch learning), and the other is when the data is flowing in streams into the learning algorithm (online learning). This flow can be as individual sample points in your dataset, or it can be in small batch sizes. Let's briefly discuss these concepts.

Batch Learning

In batch learning the machine learning model is trained using the entire dataset that is available at a certain point in time. Once we have a model that performs well on the test set, the model is shipped for production and thus learning ends. This process is also called *offline learning*. If in the process of time, new data becomes available, and there is need to update the model based on the new data, the model is trained from scratch all over again using both the previous data samples and the new data samples.

This pipeline is further illustrated in Figure 15-1.