18.1.2

Unsupervised Learning

You and Martha are convinced: Unsupervised learning is the best way forward for your dataset because you are looking for any groupings, trends, or other information that could help you pitch cryptocurrencies to her firm.

Now the fun part of the project can start: learning how to use unsupervised algorithms. You know it might take a few tries to get it right, but if you can convince Accountability Accounting to invest in this cutting-edge financial system, you will truly be making a difference.

In supervised learning, the input data already has a paired outcome, which is plugged in to train the model to predict outcomes in new datasets. For example, we want to build a model that, when given unfamiliar data, can accurately predict the outcomes.

In **unsupervised learning**, there are two key differences from the above approach:

- There are no paired inputs and outcomes.
- The model uses a whole dataset as input.

Unsupervised learning is used in one of the following two ways:

- Transform the data to create an intuitive representation for analysis or to use in another machine learning model; or
- Cluster or determine patterns in a grouping of data, rather than to predict a classification.

For example, imagine a store owner is planning to stock up on back-to-school supplies. The owner's challenges are selecting from a wide range of school supplies, devoting considerable space to the inventory, and maintaining a fast inventory turnover rate, from stocking to selling.

The owner wants to predict the best time to start selling school supplies so staff can make room for the next big items to sell. The owner decides to focus on the previous year's data, which includes the number of items sold, when they were sold, and the area schools' start dates for 10, 20, and 30 days prior to the items' sell dates.

Since the owner knows what they are looking for—the best time to start selling school supplies—what would be the best approach?

○ Logistic regression

○ Classification

○ Unsupervised learning

○ Linear regression

Check Answer

Finish ▶

Now, imagine the owner notices that when customers come in to buy school supplies, they also purchase non-school items, giving other categories a bump in sales. There are a wide variety of products to offer different consumers, so the owner wants to know which consumers might buy more products and what items they might want to buy. We don't really know *how* to group the data, but we know that grouping the data will be instructive. You'll need to factor in many data points, such as age, race, gender, location, and much more.

Since the owner doesn't know exactly how many groups or what kinds of customers those groups contain, what would be the best method for finding out?

○ Logistic regression

○ Classification

○ Unsupervised learning

○ Linear regression

Check Answer

Finish ▶