18.1.3

Types of Unsupervised Learning: Transformations and Clustering

Unsupervised learning appears to be a great choice! You and Martha decided to follow your intuition and research which type of unsupervised learning is best: transformations or clustering.

There are two types of unsupervised learning: transformations and clustering algorithms.

Transformations

We use **transformations** when we need to take raw data and make it easier to understand. Transformations also can help prepare data so that it can be used for other machine learning algorithms.

For example, if you had cable TV viewer data, it might contain location, viewing duration, and viewer churn and retention during commercials and after programs ended. Transformations can reduce the dimensional representation, which simply means we'll be decreasing the number of features used for the model or analysis. After doing so, the data can either be processed for use in other algorithms or narrowed down so it can be viewed in 2D.

Clustering Algorithms

We use **clustering algorithms** to group similar objects into clusters. For example, if a cable service wants to group those with similar viewing habits, we would use a clustering algorithm.

Challenges of Unsupervised Learning

IMPORTANT

> Unsupervised learning isn't the solution for every data analytic challenge. Just because supervised learning might not work for one situation doesn't mean unsupervised learning will work instead. Understanding the data and what can be done with it is an important first step before choosing an algorithm.

Recall that unsupervised learning does not take in any pairing of input and outcomes from the data—it only looks at the data as a whole. This can cause some challenges when running the algorithm. Since we won't know the outcome it's predicting, we might not know that the result is correct.

This can lead to issues where we're trying to decide if the model has provided any helpful information that we can use to make decisions in the real world. For example, our store owner might run a model that ends up grouping the type of people by how much they're buying. This could be useful in some contexts—for example, knowing who the top

spenders are—but it might not help the store owner better organize the store for maximum purchases per person, or understand the differences in product preferences between top purchasers.

The only way to determine what an unsupervised algorithm did with the data is to go through it manually or create visualizations. Since there will be a manual aspect, unsupervised learning is great for when you want to explore the data. Sometimes you'll use the information provided to you by the unsupervised algorithm to transition to a more targeted, supervised model.

As with supervised learning, data should be preprocessed into a correct format with only numerical values, null value determination, and so forth. The only difference is unsupervised learning doesn't have a target variable—it only has input features that will be used to find patterns in the data. It's important to carefully select features that could help to find those patterns or create groups.

The next section will cover data preprocessing and data wrangling, and provide a refresher on Pandas and data cleaning. First, you'll need to install the necessary libraries for practice.