Introduction

An overview of unsupervised learning and clustering.

In the **Clustering** section, you will be using unsupervised learning methods, i.e. methods of extracting insights from unlabeled datasets. Specifically, you will learn about different *clustering* algorithms and how they are able to group together similar data observations.

A. Unsupervised learning

So far, we've only used supervised learning methods, since we've exclusively been dealing with labeled datasets. However, in the real world many datasets are completely unlabeled, since labeling datasets involves additional work and foresight. Rather than just ignoring all these unlabeled datasets, we can still extract meaningful insights using unsupervised learning.

Since we only have data observations to work with, and no labels, unsupervised learning methods are centered around finding similarities/differences between data observations and making inferences based on those findings. The most commonly used form of unsupervised learning is clustering (https://en.wikipedia.org/wiki/Cluster_analysis). As the name suggests, clustering algorithms will cluster the data into distinct groups (clusters), where each cluster consists of similar data observations.

Clustering is used in many different applications, from anomaly detection (i.e. detecting real vs. fraudulent data) to market research (e.g. grouping customers together based on their purchase history). In the upcoming chapters, you'll learn about a variety of commonly used clustering algorithms in data science, as well as other tools for finding similarities between data observations.

Mark as Completed

55% completed, meet the criteria and claim your course certificate!

Buy Certificate