Clustering - Introduction

(https://github.com/learn-co-curriculum/dsc-clustering-intro) (https://github.com/learn-co-curriculum/dsc-clustering-intro/issues/new)

Introduction

In this section, you'll learn about a useful unsupervised learning technique: clustering. This lesson summarizes the topics you'll be covering in this section.

Clustering

Clustering techniques are very powerful when you want to group data with similar characteristics together, but have no pre-specified labels. The main goal of clustering is to create clusters that have a high similarity between the data belonging to one cluster while aiming for minimal similarity between clusters.

K-Means Clustering

We start by providing a basic intuition of the K-means clustering algorithm. When using the K-means clustering algorithm, the number of clusters that you want to obtain is specified upfront and the algorithm aims at the most "optimal" cluster centers, given that there are *K* clusters.

Hierarchical Agglomerative Clustering

A second branch of clustering algorithms is hierarchical agglomerative clustering. Using hierarchical clustering, unlike K-means clustering, you don't decide on the number of clusters beforehand. Instead, you start with *n* clusters, where *n* is the number of data points, and at each step you join two clusters. You stop joining clusters when a certain criterion is reached.

Semi-Supervised Learning

Semi-supervised learning techniques, which are increasingly popular in machine learning, combine both concepts of supervised and unsupervised learning.

Market Segmentation with Clustering

A very common and useful application of clustering is market segmentation. You'll practice your clustering skills on a market segmentation dataset!

Summary

In this section, you'll learn how to use clustering techniques which are very useful for finding patterns and grouping unlabeled data together.