Unsupervised Learning: Clustering

Abdelrahman Khaled

Machine Learning Research Cluster
German University in Cairo

January 23, 2019

Outline

Unsupervised Learning

In unsupervised learning our program attempts to learn from data that is unlabeled. meaning that our data X is only a set feature vectors with no label set Y.

We try to solve the problem considering that similar data points should be grouped together and dissimilar data points should be grouped in different groups. This process is called clustering where each cluster is a group of similar points. (in theory)

Unsupervised Learning

To make things simpler later on, let's introduce the standard notation.

Standard Notation

$X = \{x_1, x_2, \ldots, x_n\}$ is the set of all feature vectors.

Unlike in supervised learning, in unsupervised learning there is no agreed upon way to measure how well the algorithm is doing.

Clustering

Definition

"Grouping or segmenting a collection of objects into subsets or 'clusters,' such that those within within each cluster are more closely related to one another than objects assigned to different clusters."

But in order to cluster points we need to have a notion of similarity or dissimilarity.

Clustering

Definition

"Grouping or segmenting a collection of objects into subsets or 'clusters,' such that those within within each cluster are more closely related to one another than objects assigned to different clusters."

But in order to cluster points we need to have a notion of similarity or dissimilarity.
Let's define the dissimilarity D between two points to be the squared distance between those two points in space.

Clustering

Definition

"Grouping or segmenting a collection of objects into subsets or 'clusters,' such that those within within each cluster are more closely related to one another than objects assigned to different clusters."

But in order to cluster points we need to have a notion of similarity or dissimilarity.
Let's define the dissimilarity D between two points to be the squared distance between those two points in space.

$$d(x_{ij}, x_{i'j}) = (x_{ij} - x_{i'j})^2$$

$$D(x_i, x_{i'}) = \sum_j d(x_{ij}, x_{i'j})$$

Clustering

Now what?

Now we need to figure out a way to create clusters that contain the least dissimilar points ... but how do we do that?

Clustering

Now what?

Now we need to figure out a way to create clusters that contain the least dissimilar points ... but how do we do that?

The answer is: 'trial and error'
We try choosing many different numbers of clusters and see which one better segments the data. ¡3-¿

The first thing that comes to mind is brute force! For n datapoints try everything from 1 cluster to n clusters.

Clustering: The Problem

The problem with clustering and the brute force method is the enormous number of possibilities. It's almost impossible to go through all of them.
For K clusters and n data points, the algorithm will have this number of iterations

$$\frac{1}{k!}\sum_{k=1}^{K}(-1)^{K-k}\binom{K}{k}k^n$$

Which increases at an alarmingly fast rate the more data points we have.

K-Means Clustering

K-Means clustering is an algorithm that tries to only go through a small subset of all the possible clusters and chooses what it deems the best.

K-Means Algorithm

1. Choose a K
2. Randomly choose K points as initial "centroids" and assign each to a different cluster.
3. While the centroids change do
   1. Assign each point to the cluster that is assigned to the centroid closest to it.
   2. Average all the points in each cluster to get K new centroids.

Hierarchical Clustering: Agglomerative Clustering

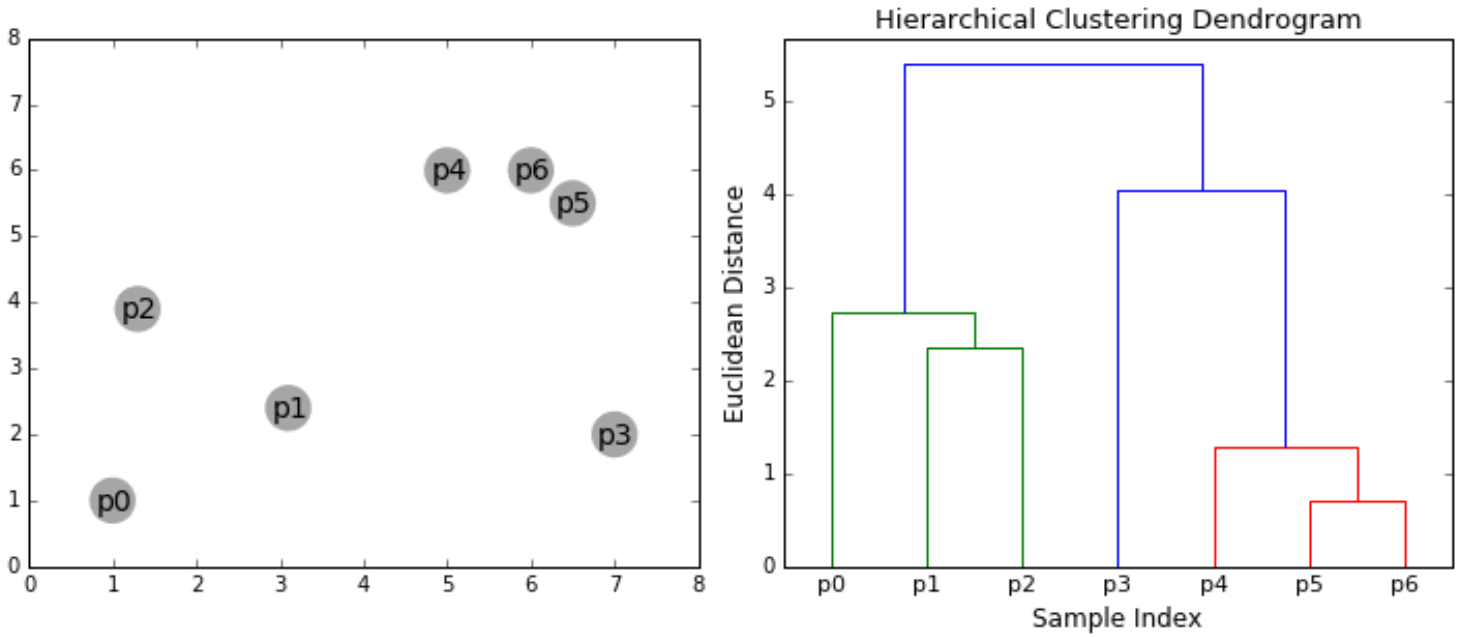

Figure: Agglomerative Clustering *Image Source*

Hierarchical Clustering: Agglomerative Clustering

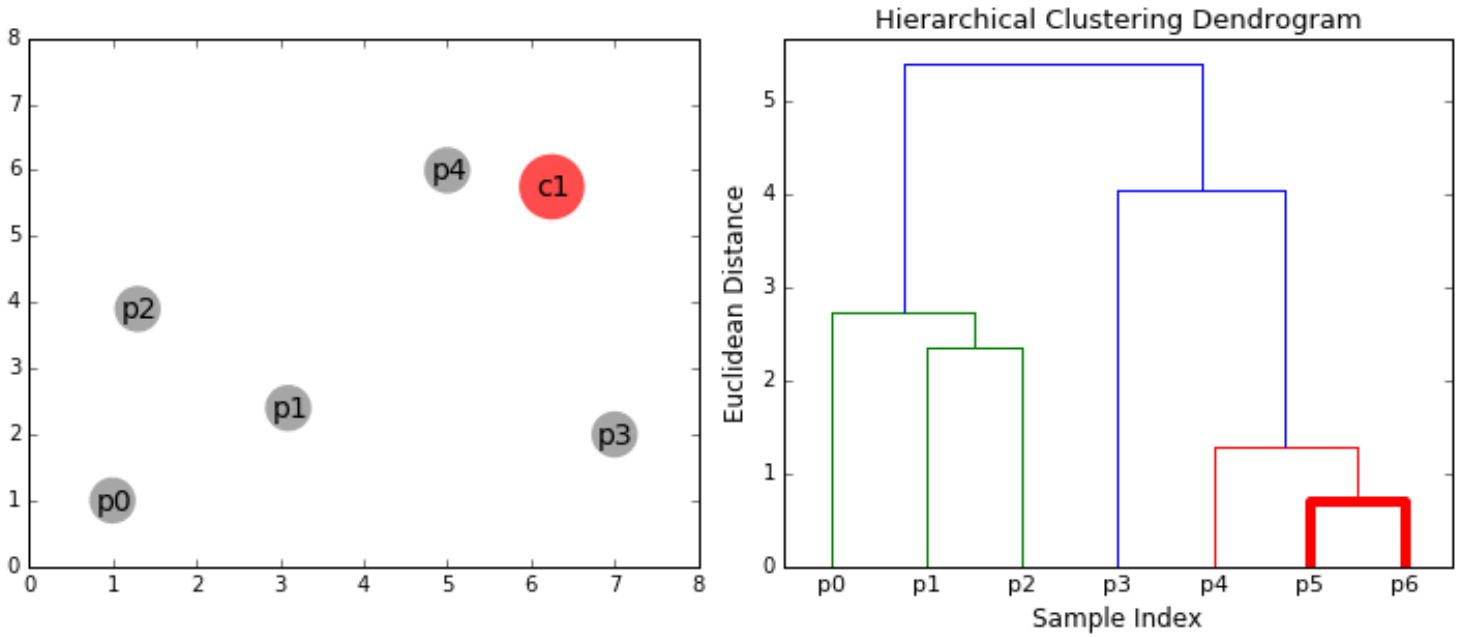

Figure: Agglomerative Clustering *Image Source*

Hierarchical Clustering: Agglomerative Clustering

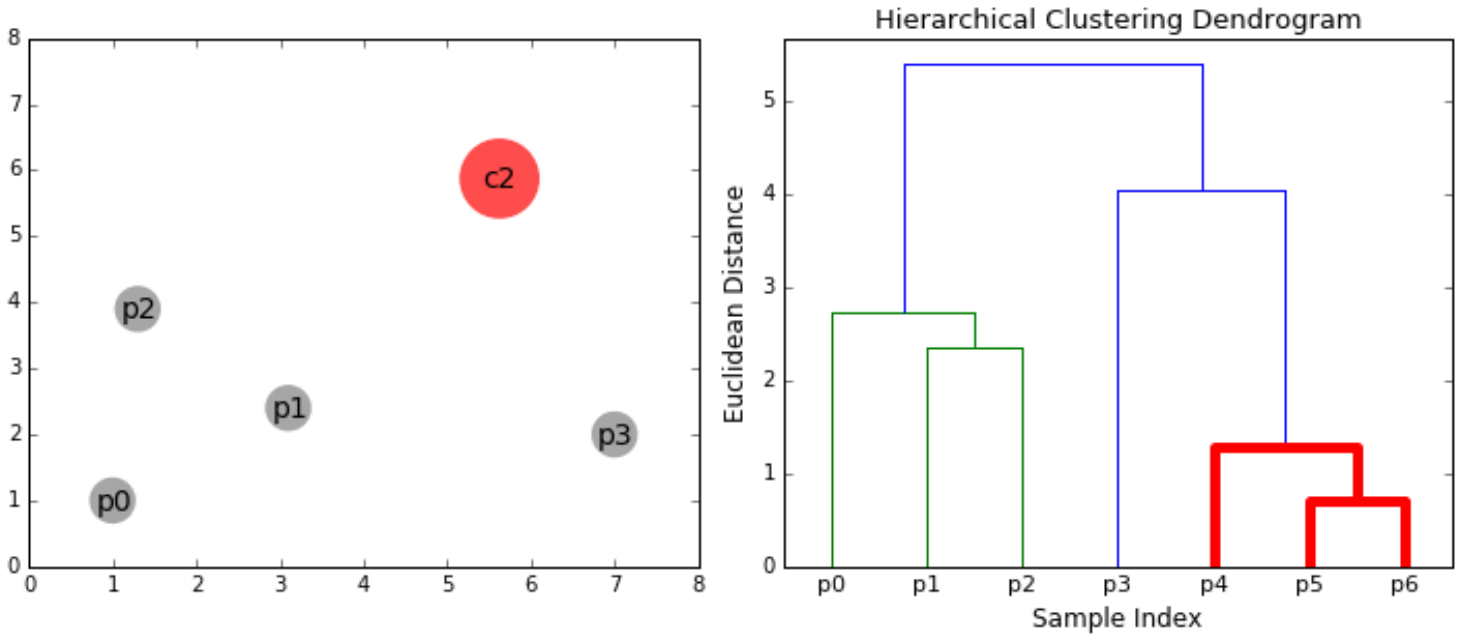

Figure: Agglomerative Clustering *Image Source*

Hierarchical Clustering: Agglomerative Clustering

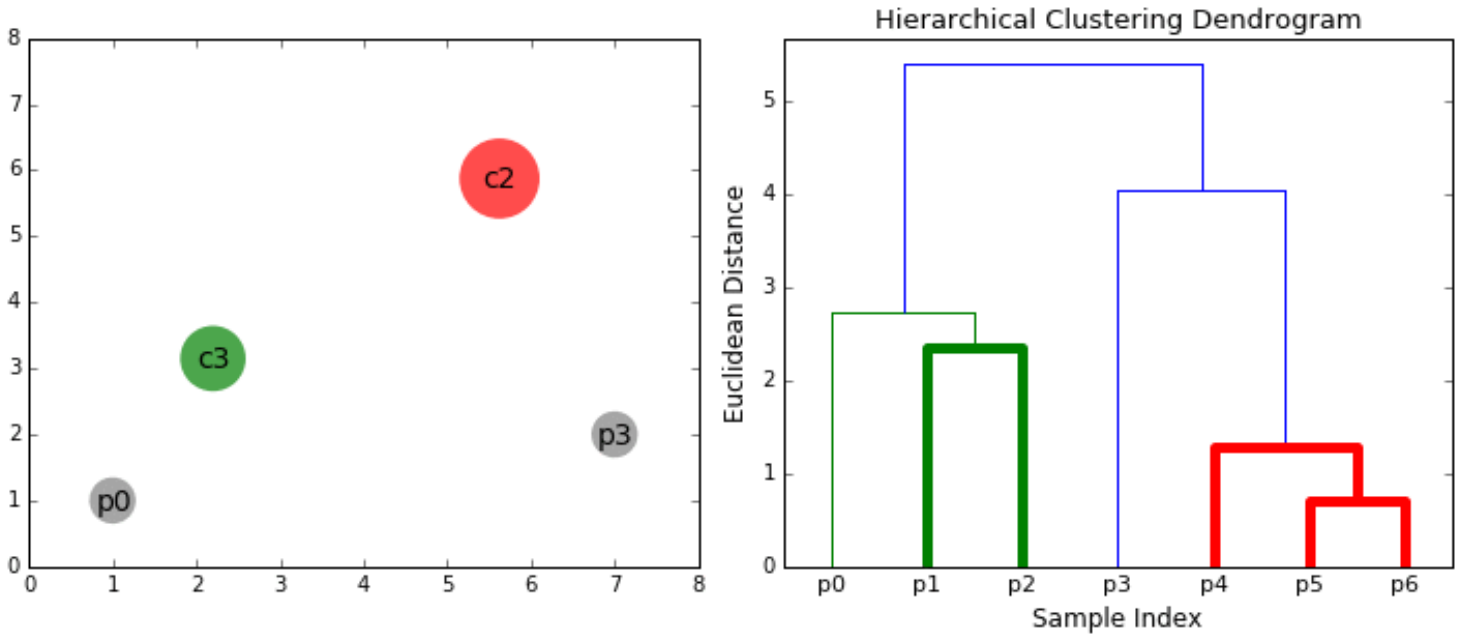

Figure: Agglomerative Clustering *Image Source*

Hierarchical Clustering: Agglomerative Clustering

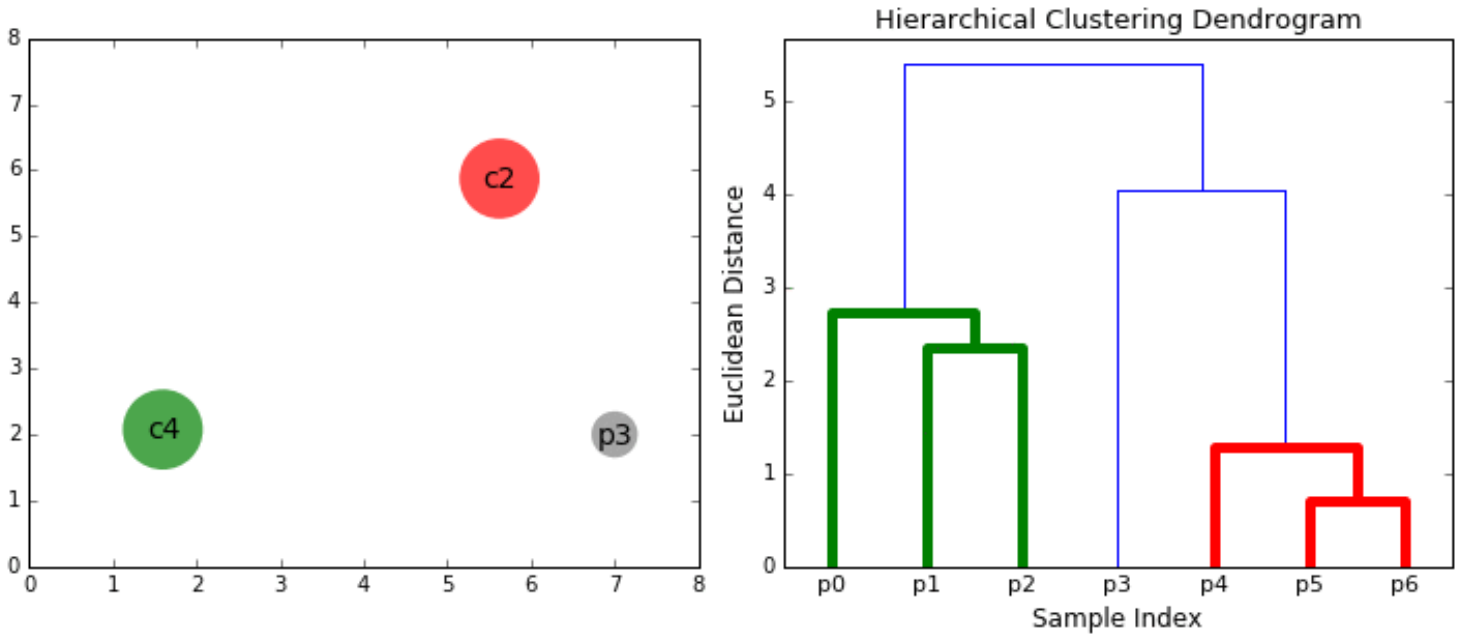

Figure: Agglomerative Clustering *Image Source*

Hierarchical Clustering: Agglomerative Clustering

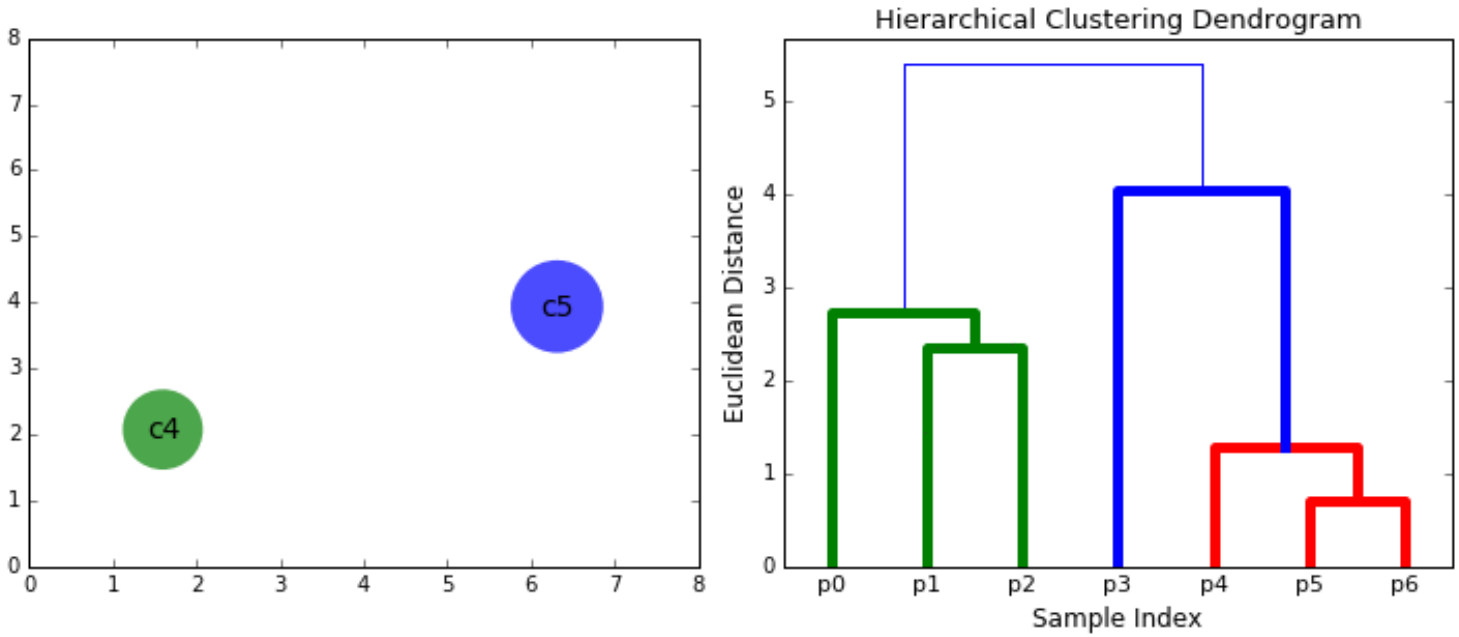

Figure: Agglomerative Clustering *Image Source*

Hierarchical Clustering: Agglomerative Clustering

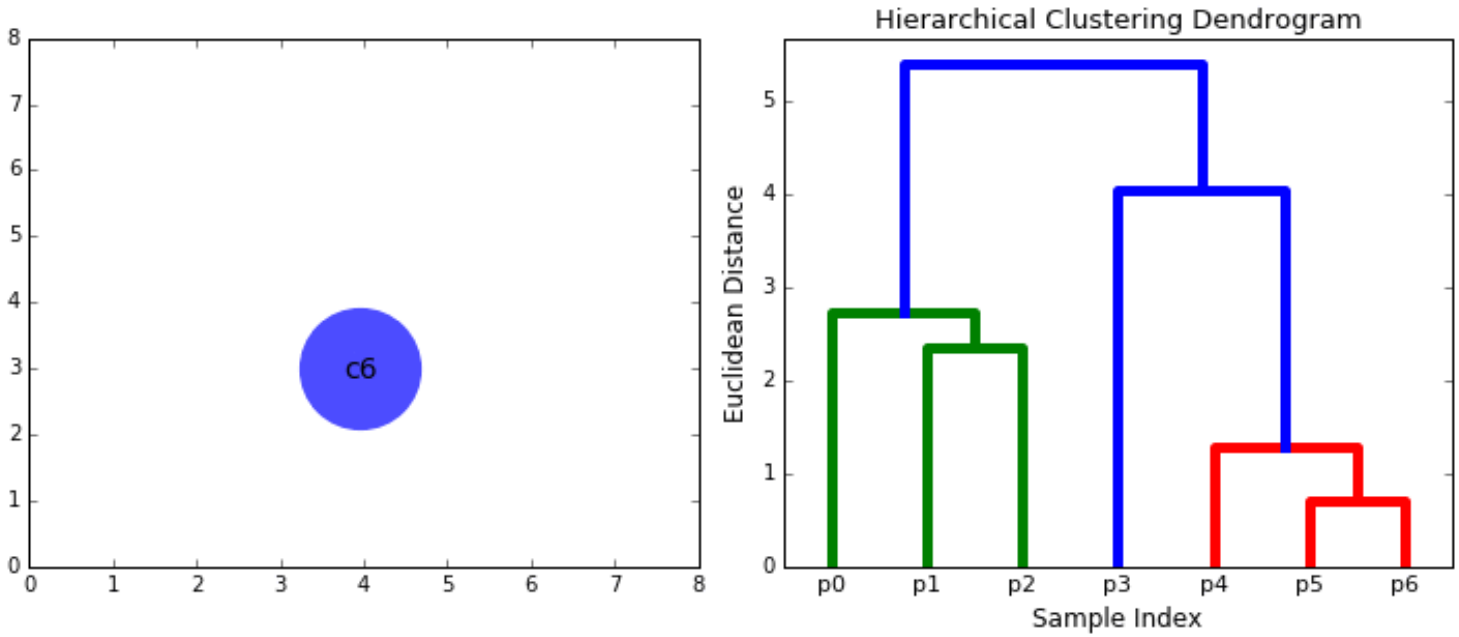

Figure: Agglomerative Clustering *Image Source*

Hierarchical Clustering: Agglomerative Clustering

There are three ways to decide how to cluster two clusters into one.

1. Single-Linkage Clustering: The minimum distance between all elements of both clusters. (The closest two elements)
2. Complete-Linkage Clustering: The maximum distance between all elements of both clusters. (The farthest two elements)
3. Average-Linkage Clustering: The average distance between all the elements of each cluster.

References

- MIT 6.0002 Introduction to Computational Thinking and Data Science, Fall 2016.
  Lecture 12: Clustering *link*
- *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition.* Trevor Hastie, Robert Tibshirani, Jerome Friedman. Springer, 2013.
  Section 14.3
- *Machine Learning An Algorithmic Perspective, Second Edition.* Stephen Marsland. Chapman and Hall, 2014.
  Section 14.1