Unsupervised Learning: Clustering

ML Instruction Team, Fall 2022

CE Department
Sharif University of Technology

Clustering: An Overview

- Clustering algorithms can be classified into different categories, based on the following criteria:
  - Whether each point is assigned to exactly one cluster or several clusters with certain probabilites that add up to 1:
    - **Hard**
    - **Soft**
  - Whether all clusters are on the same level or several clusters are built in a hierarchical way:
    - **Partitional**
    - **Hierarchical**

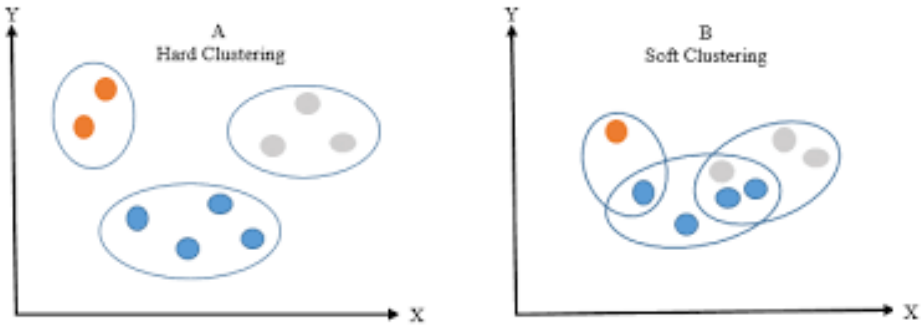

Figure: Hard vs Soft [1].

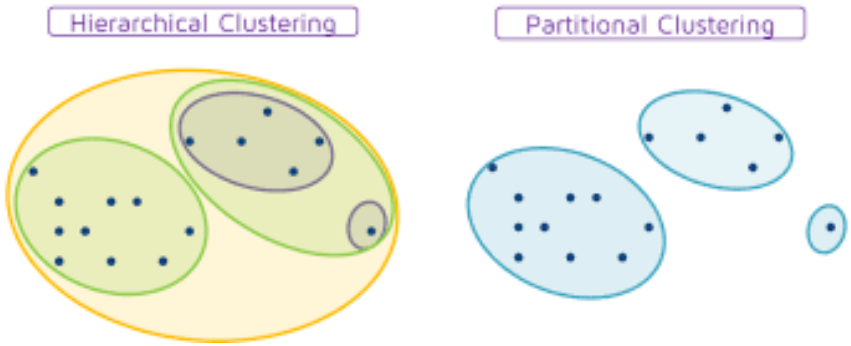

Figure: Partitional vs Heirarchical [2].

Clustering: An Overview

- Hierarchical clustering is usually done in two different ways:
  - **Agglomerative**: This is a "bottom-up" approach, Each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.
  - **Divisive**: This is a "top-down" approach, All observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

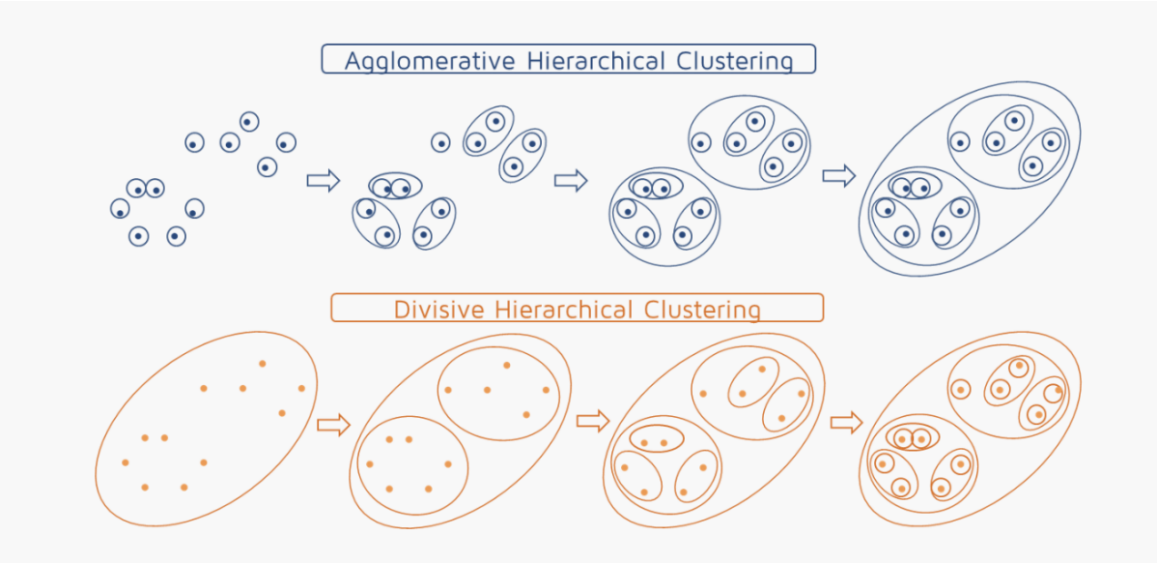

Figure: Agglomerative vs Divisive [3].

Hard Partitional Clustering: K-Means

- A particularly simple method for clustering is K-means, The idea is **to represent each cluster k by a center point $\mathbf{c}_k$ and assign each data point $\mathbf{x}_n$ to one of the clusters k** which can be written in terms of index sets $\mathcal{C}_k$
- The center points and the assignment are then chosen such that the mean squared distance between data points and center points **is minimized**:

$$J := \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|\mathbf{x}_n - \mathbf{c}_k\|^2$$

- Here we introduced a corresponding binary indicator variable $r_{nk} \in \{0, 1\}$ where $k = 1, 2, \ldots, K$ describing which of the K clusters the data point $\mathbf{x_n}$ is assigned to, so that if data point $\mathbf{x_n}$ is assigned to cluster k then $r_{nk} = 1$, and $r_{nj} = 0$ for $j \neq k$
- Now, Our goal is to find values for the $\{r_{nk}\}$ and the $\{\mathbf{c}_k\}$ so as to minimize J. we can do this through an **Iterative Procedure**

Hard Partitional Clustering: K-Means

- To minimize J through iterating, we have to do the following algorithm :
  1. **Initialize $\mathbf{c_k}$** with **Random Value** for all $k = 1, 2, \dots, K$, It could be chosen from data values either.
  2. Minimize J with respect to r_{nk}, keeping the c_k fixed. because J is a linear function of r_{nk} this optmization can be performed easily to give a closed form solution:

$$r_{nk} = \begin{cases} 1 & \text{if } k = \texttt{argmin}_j \|\mathbf{x}_n - \mathbf{c}_j\|^2 \\ 0 & \text{otherwise} \end{cases}$$

  3. Minimize J with respect to $\mathbf{c_k}$, keeping the r_{nk} fixed. if the assignment is fixed, it is easy to show that the optimal choice of the center positions is given by:

$$\mathbf{c_k} = \frac{\sum_n r_{nk}\mathbf{x}_n}{\sum_n r_{nk}}$$

  4. Check the convergence criteria, otherwise go to step 2.

Hard Partitional Clustering: K-Means

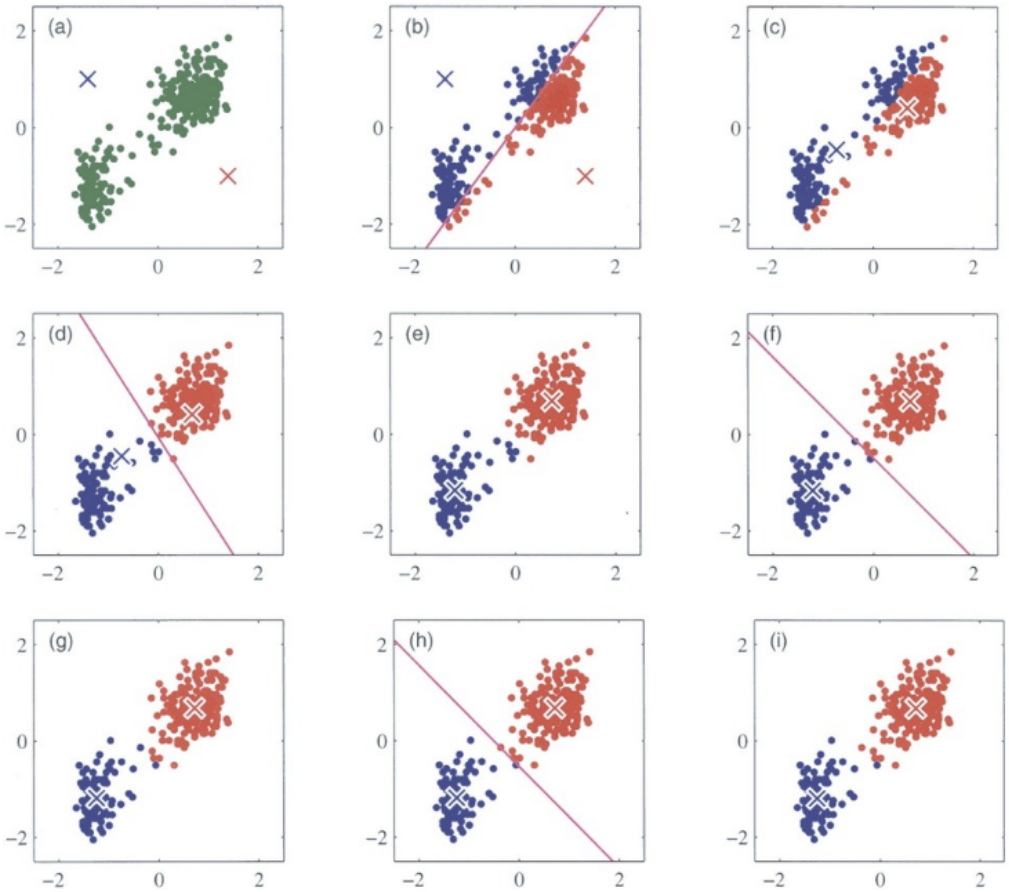

Figure: Illustration of K-Means with $K = 2$

Hard Partitional Clustering: K-Means

- Note that the result of the algorithm is **not necessarily a global optimum** of the objective function J
- It is therefore advisable to **run the algorithm several times** with different initial center locations and **pick the best result**.
- A drawback of this and many other clustering algorithms is that **the number of clusters is not determined**.
- One has to decide on a proper K in advance, or one simply runs the algorithm with several different K-values and picks the best according to some criterion.

References

- **[1]**. Raschka, Sebastian. "What Are Data Science and Machine Learning?" Dr. Sebastian Raschka, 3 Sept. 2022, sebastianraschka.com/faq/docs/datascience-ml.html.
- **[2]**. Raschka, Sebastian, and Vahid Mirjalili. Python Machine Learning: Machine Learning and Deep Learning With Python, Scikit-learn, and TensorFlow 2, 3rd Edition. 3rd ed., Packt Publishing, 2019.
- **[3]**. Peluffo, Diego. Dimensionality Reduction Effect Over an Artificial (3-dimensional) Spherical Shell Manifold. Resultant Embedded (2-dimensional) Data Is an Attempt to Unfolding the Original Data. Feb. 2017, www.researchgate.net/publication 313787026-Interactive-Data-Visualization-Using-Dimensionality-Reduction-and-Similarity-Based-Representations.
- **[4]**. Kumar, Ajitesh. "5 Common Ensemble Methods in Machine Learning." Data Analytics, 16 Aug. 2022, vitalflux.com/5-common-ensemble-methods-in-machine-learning.
- **[5]**. http://strijov.com/sources/demo-GLM.php
- **[6]**. www.researchgate.net/figure/Schematic-of-a-Decision-Tree-The-figure-shows-an-example-of-a-decision-tree-with-3-fig1-348456545. Accessed 8 Sept. 2022.

References

- **[7]**. Wikipedia contributors. “Support-vector Machine.” Wikipedia, 1 Sept. 2022, en.wikipedia.org/wiki/Support-vector-machine.
- **[8]**. www.researchgate.net/figure/Simple-directed-graphical-model-with-three-variables-To-illustrate-how-graphical-models-fig6-262407302. Accessed 8 Sept. 2022.
- **[9]**. Ashtari, Hossein. “What Is a Neural Network? Definition, Working, Types, and Applications in 2022.” Spiceworks, 3 Aug. 2022, www.spiceworks.com/tech/artificial-intelligence/articles/what-is-a-neural-network.
- **[10]**. Balaouras, Georgios. “Optimization Algorithms.” Georgios Balaouras, 21 Apr. 2022, mpalaourg.me/project/optimization-algorithms.
- **[11]**. GeeksforGeeks. “Introduction to Hill Climbing | Artificial Intelligence.” GeeksforGeeks, 23 Aug. 2022, www.geeksforgeeks.org/introduction-hill-climbing-artificial-intelligence.
- **[12]**. Agrawal, Sanidhya. “What Is Instance-Based Learning? - Sanidhya Agrawal.” Medium, 14 Dec. 2021, medium.com/@sanidhyaagrawal08/what-is-instance-based-learning-a9b06079e836.

Thank You!

Any Question?