Crafted by :

Rizki Fajar Nugroho

Machine Learning III

Unsupervised Learning Introduction
K-means clustering Algorithm
Hierarchical clustering Algorithm

Updated

Q2.2020

Hi, I'm Rizki Fajar Nugroho

Experiences:

- **Data Tech Specialist at IYKRA**
- **Machine Learning Engineer at Omdena, Singapore**
- **Graduated from Electrical and Electronic Engineering**

Objectives

- To understand the concept unsupervised learning
- To understand the concept unsupervised learning: K-Means
- To understand the concept unsupervised learning: Hierarchical Clustering

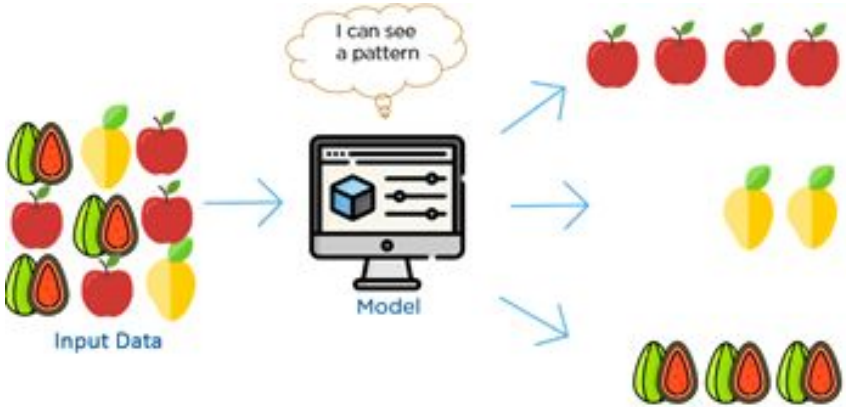
Outline

- Unsupervised Learning Introduction
- K-means Algorithm (Clustering)
- Hierarchical Clustering

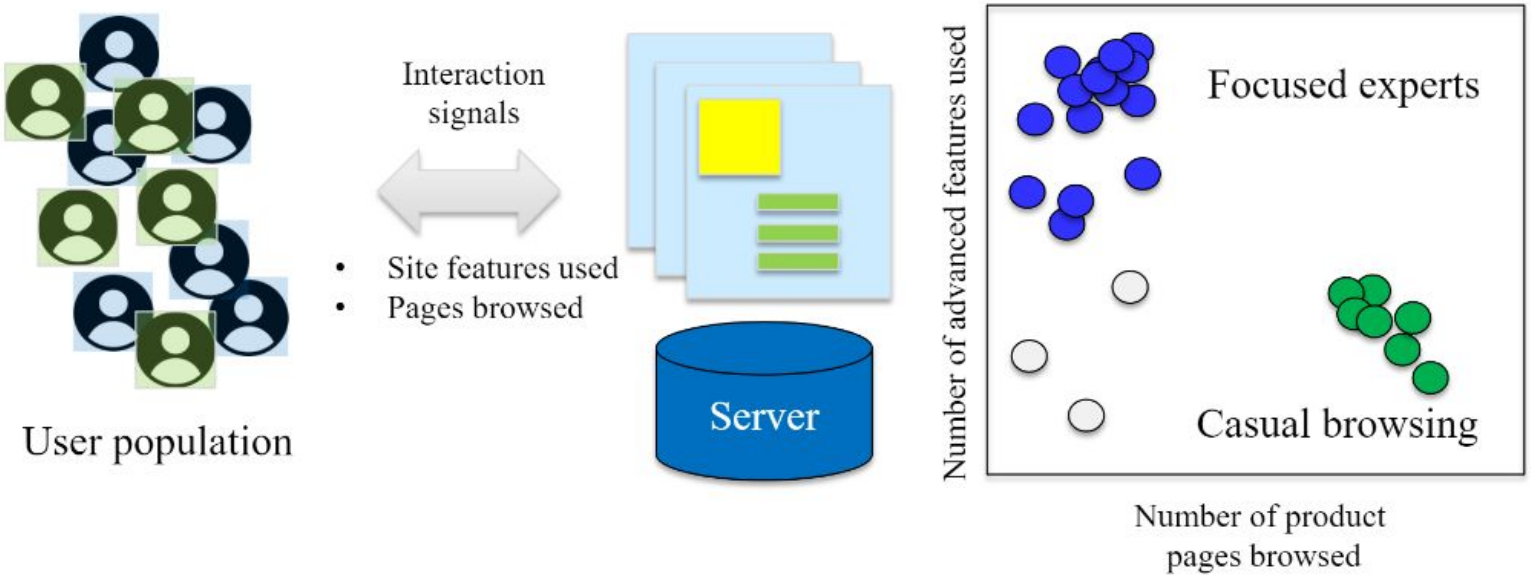
Unsupervised Learning Introduction

What is Unsupervised Learning?

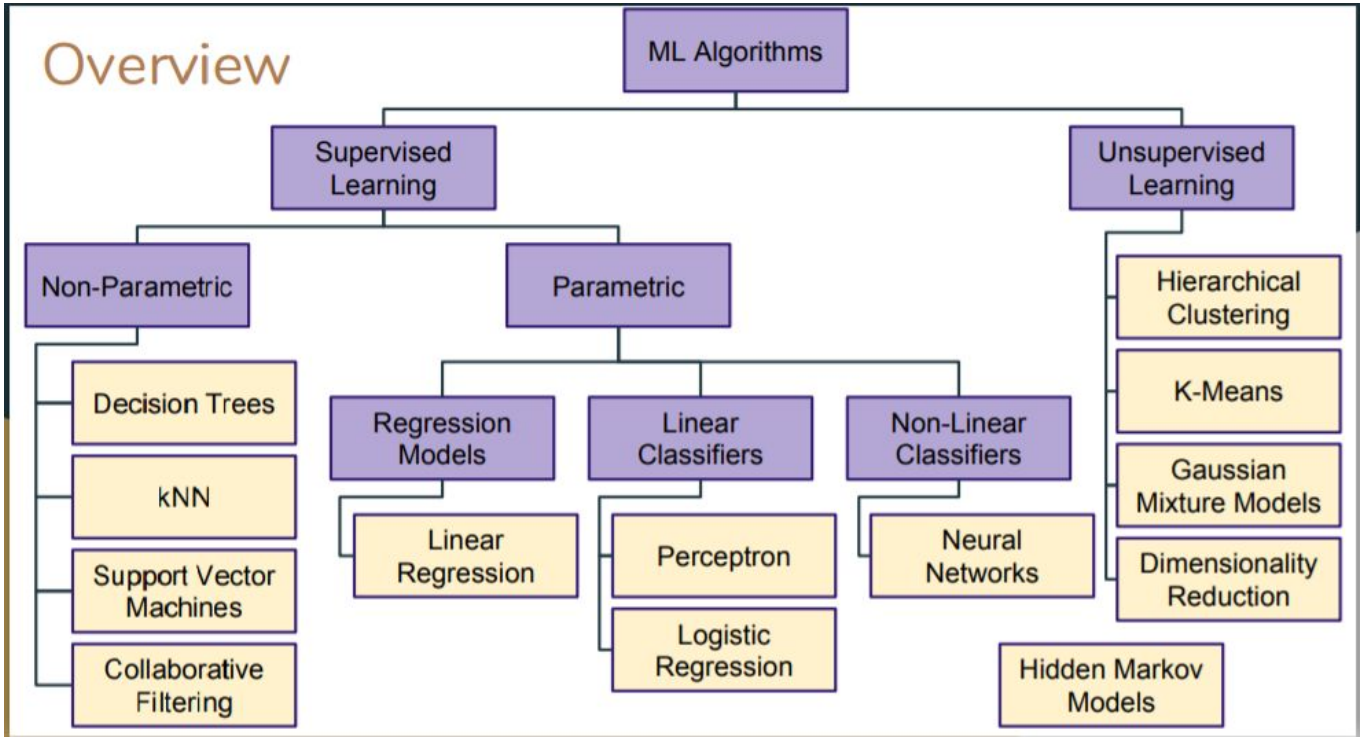
In unsupervised learning, **only input data** is provided in the dataset. There are **no labelled outputs** to aim for. But it may be surprising to know that it is still possible to find **many interesting and complex patterns hidden within data without any labels. The goal is to capture interesting structure / information**



What is Unsupervised Learning?

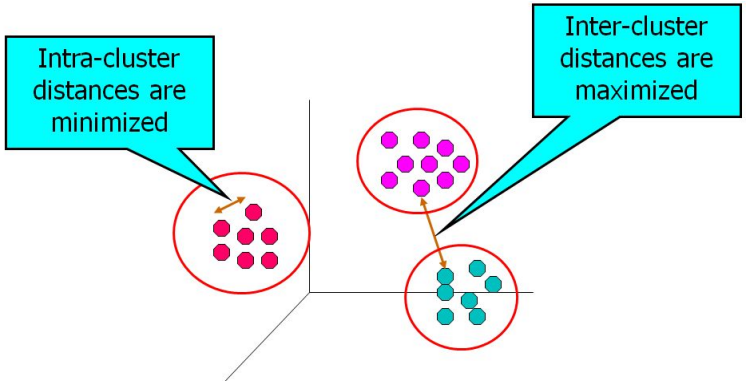


Unsupervised Learning Algorithms

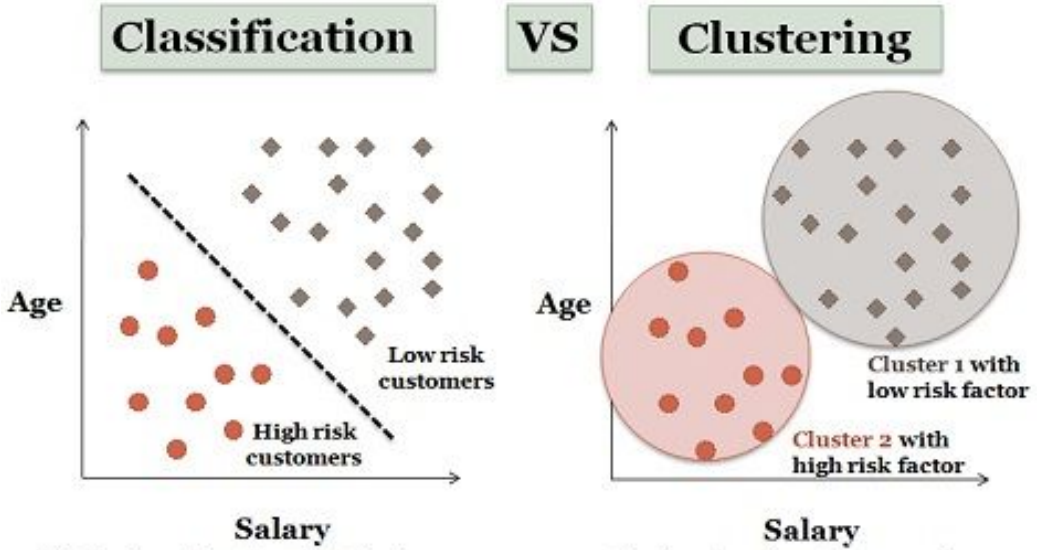


What is Clustering?

Clustering is the task of dividing the **data points** into a number of groups such that d**ata points in the same groups are more similar** to other data points in the same group **than those in other groups**. The aim is **to segregate groups with similar traits and assign them into clusters**



Difference between clustering with classification



Risk classification for the loan payees on the basis of customer salary

Difference between clustering with classification

The prior difference between classification and clustering is that classification is used in supervised learning technique where predefined labels are assigned to instances by properties, on the contrary, clustering is used in unsupervised learning **where similar instances are grouped, based on their features or properties**

Examples

- **Netflix**:

  A well-known application of clustering algorithms are Netflix recommendation systems. It is confirmed that there are about 2,000 clusters that have common audiovisual tastes. **Cluster 290** is the one that **includes people who like the series "Lost", "Black Mirror" and "Groundhog Day"**. Netflix uses these clusters **to refine its knowledge of the tastes of viewers and thus make better decisions in the creation of new original series.**

- **Banking Sector**: Classification is commonly used in the financial sector. In the era of online transactions where the use of cash has decreased markedly, it is necessary to determine whether movements made through cards are safe. **Entities can classify transactions as correct or fraudulent using historical data on customer behavior to detect fraud very accurately.**

The Application in Real-World Problems

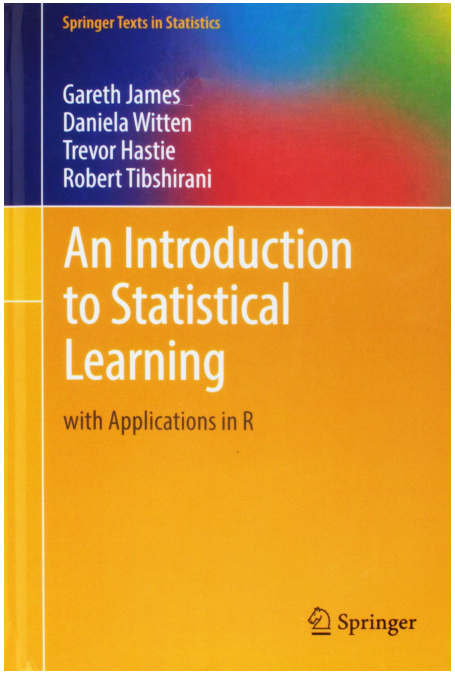
1. Customer Segmentation
2. Fraud / criminal activity Identification
3. Spam Email Identification

The Challenge of Unsupervised Learning

1. The **problem tends to be more subjective**, and **there is no simple goal for the analysis**
2. Unsupervised learning is **often performed as part of an exploratory data analysis**.
3. In unsupervised learning, **there is no way to check our result** because we don't know the true answer

K-Means (Clustering)

For the greater mathematical explanation



http://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf. Introduction to Statistical Learning. Chapter 10

K-means Clustering

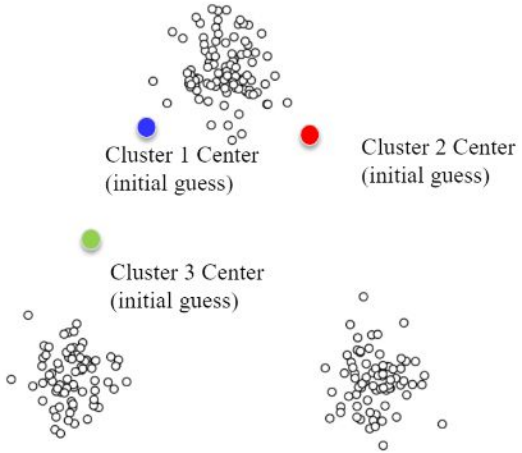
The k-means algorithm

Initialization Pick the number of clusters *k* you want to find. Then pick *k random* points to serve as an initial guess for the cluster centers.

Step A Assign each data point to the nearest cluster center.

Step B Update each cluster center by replacing it with the mean of all points assigned to that cluster (in step A).

Repeat steps A and B until the centers converge to a stable solution.



K-Means Clustering Intuition

Let’s watch this video together

How K-means clustering works?

- **Input:** N examples $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ ($\mathbf{x}_n \in \mathbb{R}^D$); the number of partitions K
- **Initialize:** K cluster centers $\mu_1, \ldots, \mu_K$. Several initialization options:
  - Randomly initialized anywhere in $\mathbb{R}^D$
  - Choose any K examples as the cluster centers
- **Iterate:**
  - Assign each of example $\mathbf{x}_n$ to its closest cluster center

$$\mathcal{C}_k = \{n : \quad k = \arg\min_k ||\mathbf{x}_n - \mu_k||^2\}$$

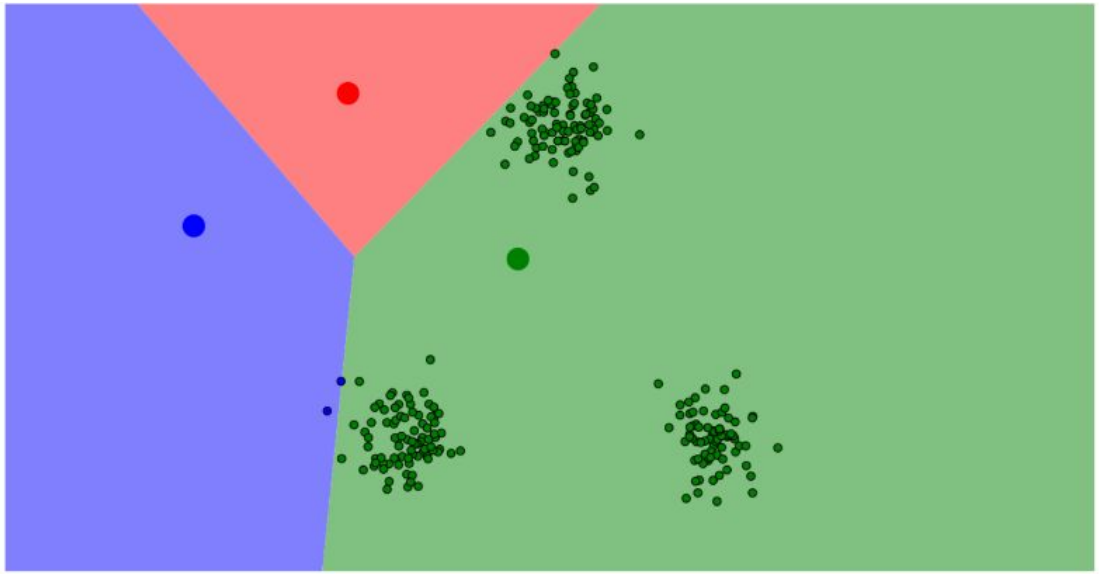
($\mathcal{C}_k$ is the set of examples closest to μ_k)

  - Recompute the new cluster centers μ_k (mean/centroid of the set $\mathcal{C}_k$)

$$\mu_k = \frac{1}{|\mathcal{C}_k|} \sum_{n \in \mathcal{C}_k} \mathbf{x}_n$$

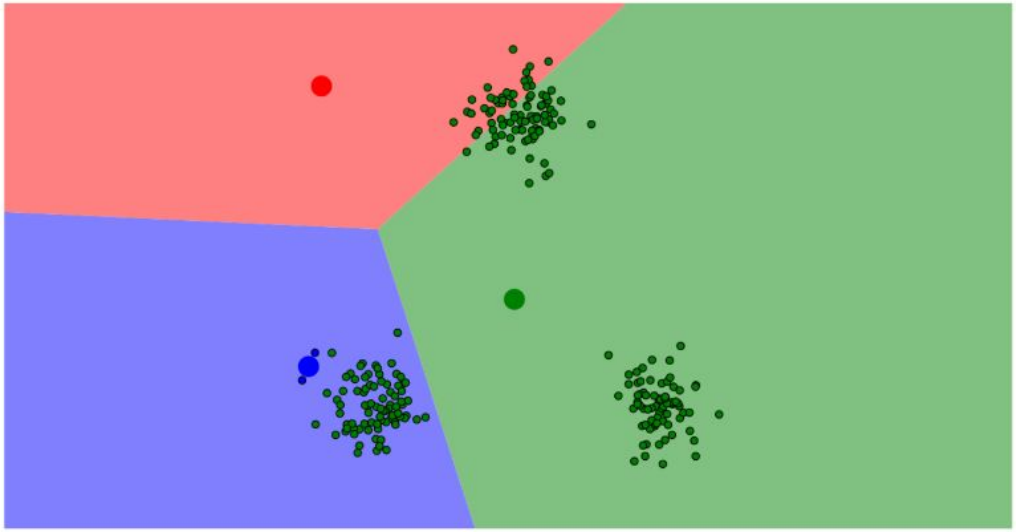
  - Repeat while not converged
- A possible convergence criteria: cluster centers do not change anymore

The Illustrations



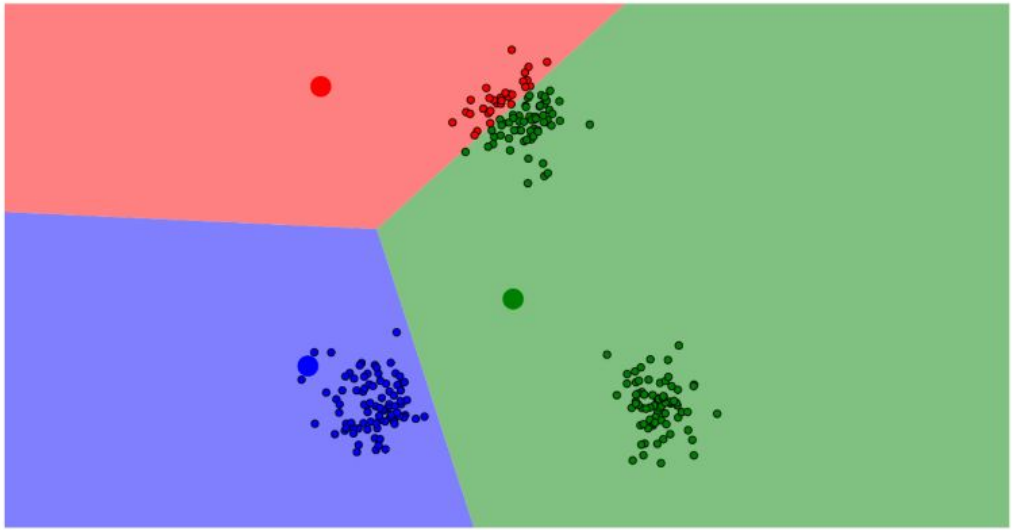
We want three clusters, so three centers are chosen randomly.

Data points are colored according to the closest center.

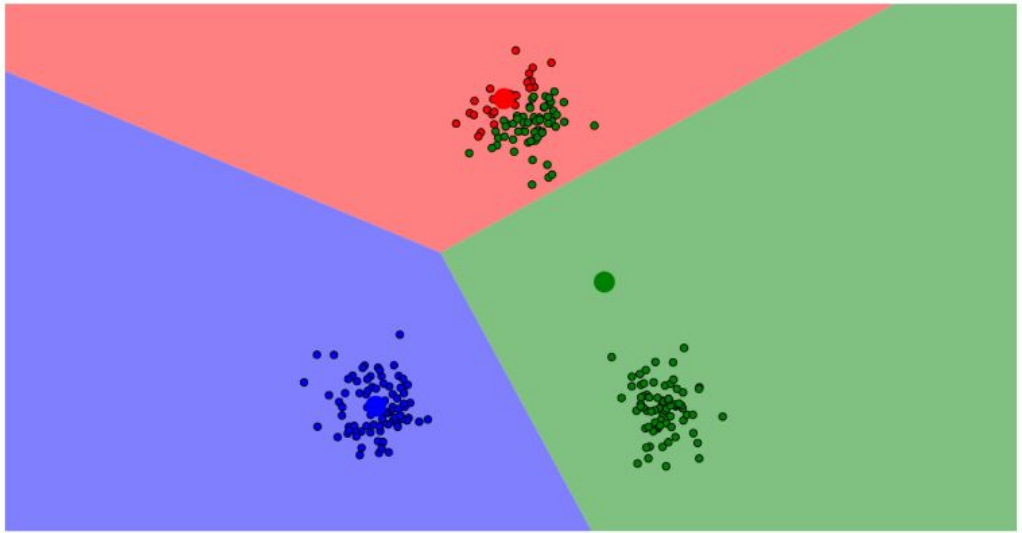


Each center is then updated…

… using the mean of all points assigned to that cluster.

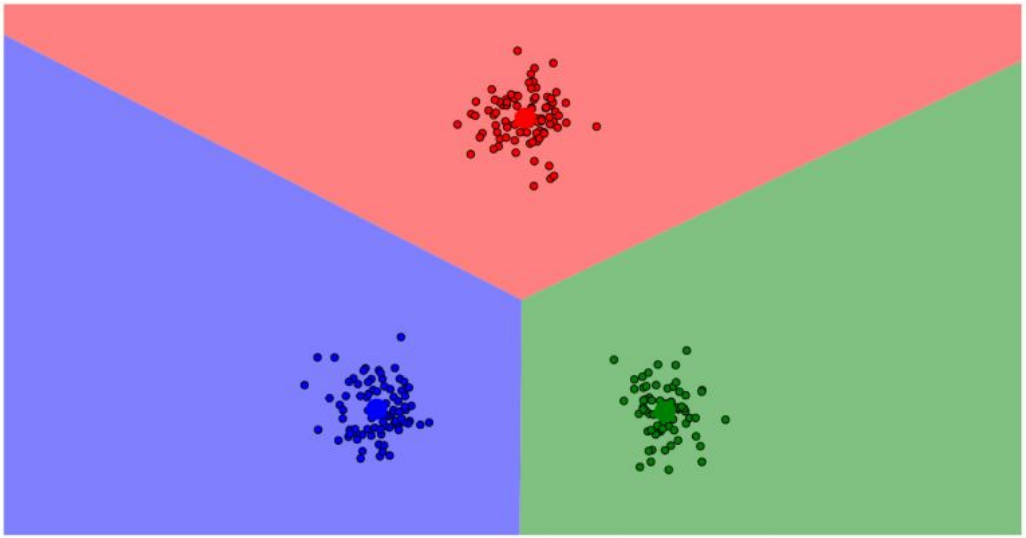


Data points are colored (again) according to the closest center.



Re-calculate all cluster centers.

The Illustrations



After repeating these steps for several more iterations…

The centers converge to a stable solution!

These centers define the final clusters.

How to choose the best K?

- There is no “easy” way for choosing the best ‘K’
- We can use the **elbow method** to calculate it

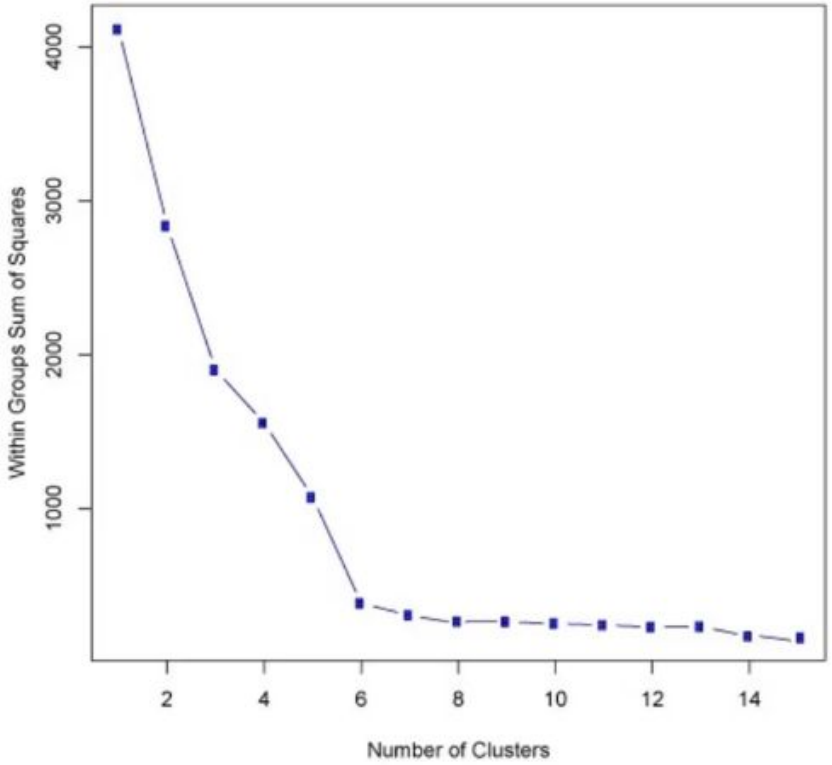
Elbow Method

- Compute the **SSE (sum squared error)** for **some values of K** (2, 4, 6, etc)
- The SSE is defined as **the sum of the squared distance between each member of the clusters and its centroids**

Elbow Method

- Once the SSE is being plotted against the K, **the error is decreases as the K gets larger**. This is because when the number of clusters increases, they should be smaller, hence the distortion is smaller
- The idea of elbow method is to **choose K at SSE decreases abruptly**

Elbow Method



The other elbow method metrics

1. silhouette
2. Calinski-harabasz

Check out the details explanation of these elbow method metrics at: Here and here

Advantages of K-means

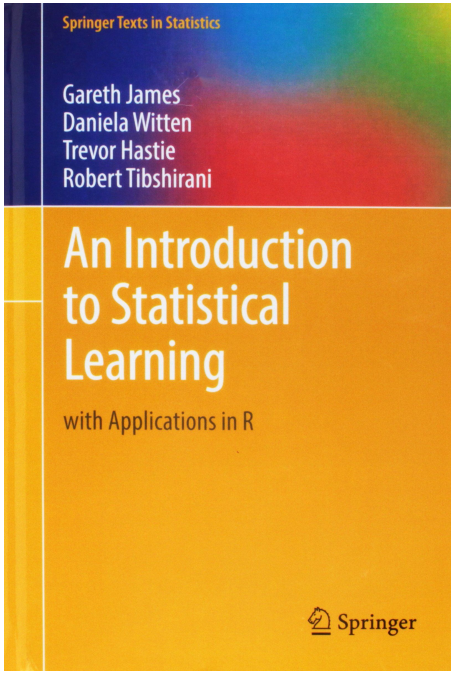
- **Relatively** simple to **implement**
- **Scales** to large data sets
- Guarantees **convergence**
- Can **warm-start** the **positions of centroids**
- Easily **adapts** to **new examples**

Disadvantages of K-means

- Choosing **k manually**
- Being **dependent on initial values**
- Clustering data **of varying sizes and density**
- Clustering **outliers**
- **Scaling with number of dimensions**

Hierarchical Clustering

For the greater mathematical explanation



http://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf. Introduction to Statistical Learning. Chapter 10

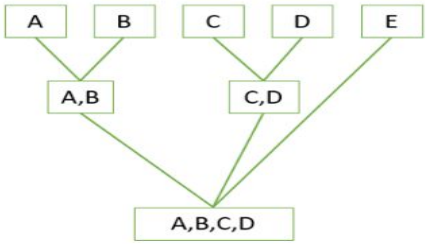
What is Hierarchical Clustering?

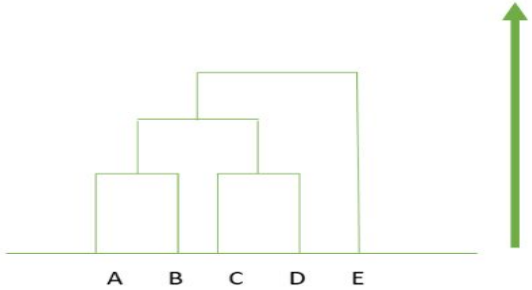
The number of clusters **is not predetermined**

There are two ways: **Buttom up,** or top down

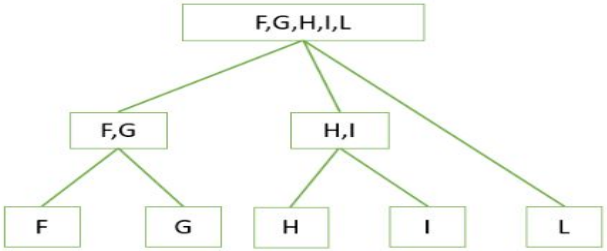
- Agglomerative - Bottom up approach. Start with many small clusters and merge them together to create bigger clusters.
- Divisive - Top down approach. Start with a single cluster rather than break it up into smaller clusters

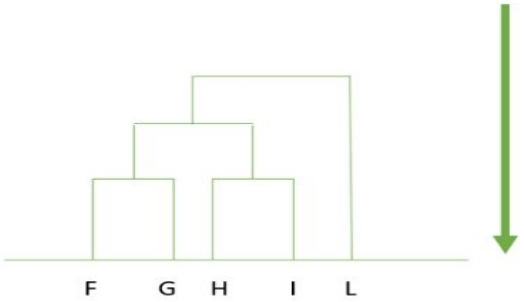
What is Hierarchical Clustering?



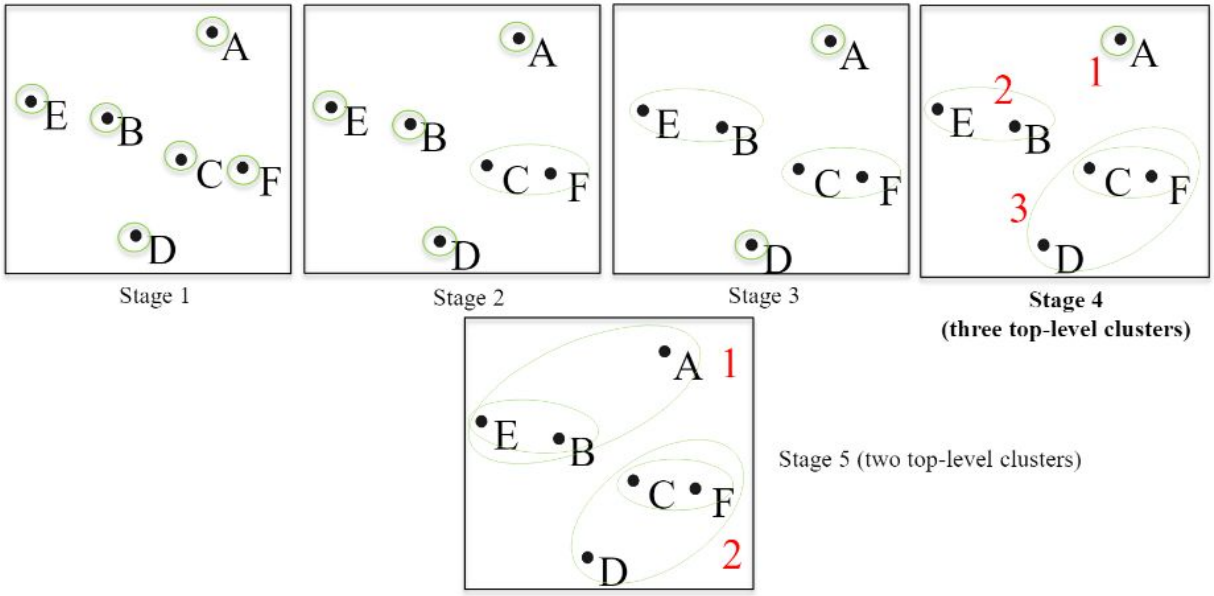


Agglomerative HC



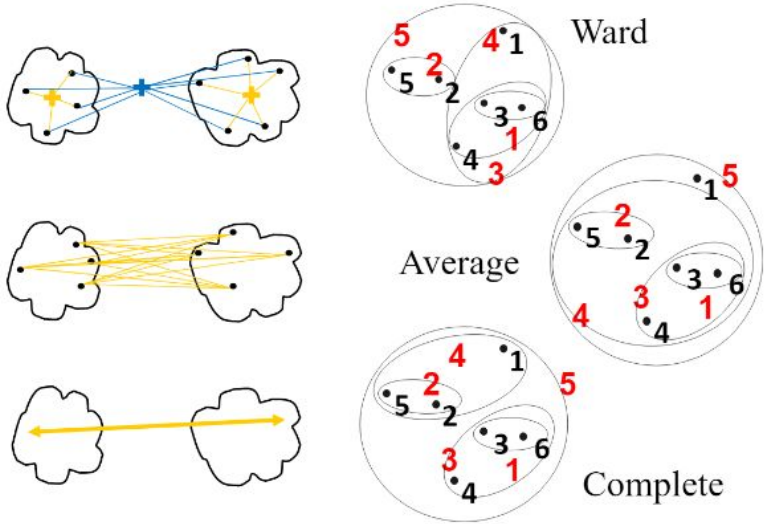


Agglomerative Clustering Example



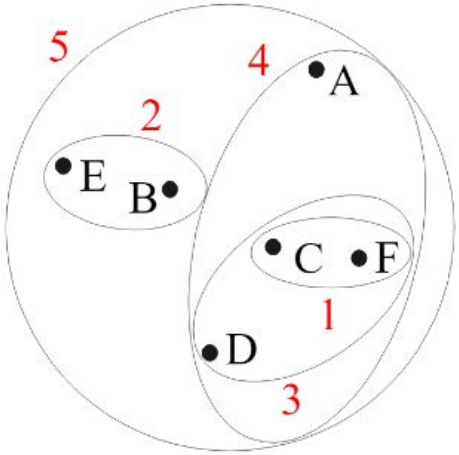
Linkage Criteria in Hierarchical Clustering

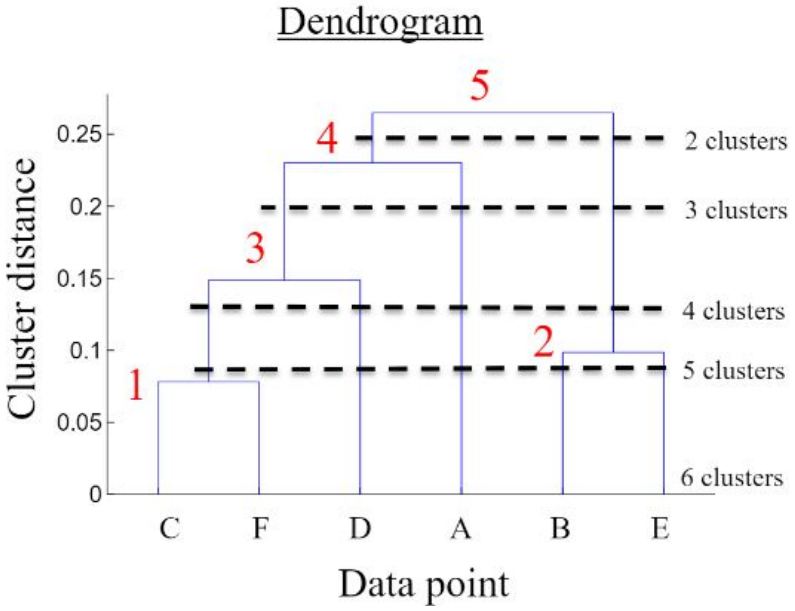
- **Ward's method**
  - *Least increase in total variance (around cluster centroids)*
- **Average linkage**
  - *Average distance between clusters*
- **Complete linkage**
  - *Max distance between clusters*



Check out the details explanation of these elbow method metrics at: **Introduction to Statistical Learning Page 395**

Dendrogram of Hierarchical Clustering





Advantages of Hierarchical Clustering

- **No assumption** of a particular number of clusters

  (i.e. k-means)
- May correspond to meaningful taxonomies

Disadvantages of Hierarchical Clustering

- Once a decision is made to combine two clusters, **it can't be undone**
- **Too slow** for large data sets

Tips and References

1. Introduction to Statistical Learning:
   Source
2. The Element of Statistical Learning:
   Source
3. Pattern Recognition and Machine Learning:
   Source
4. Interpretable ML Book:
   Source
5. Hands on Machine Learning with Scikit-Learn Book:
   Source

Thank you!