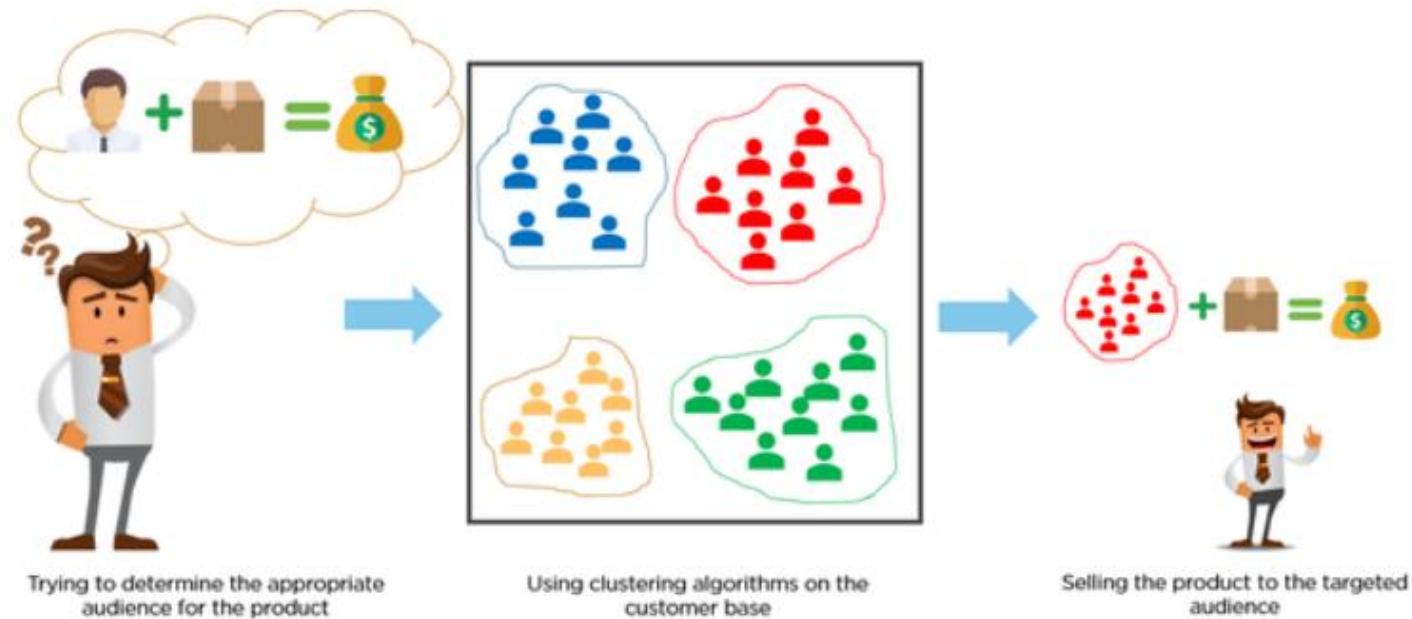


Unsupervised and Unstructured Machine Learning

BA820 – Mohannad Elhamod

Intro to Clustering

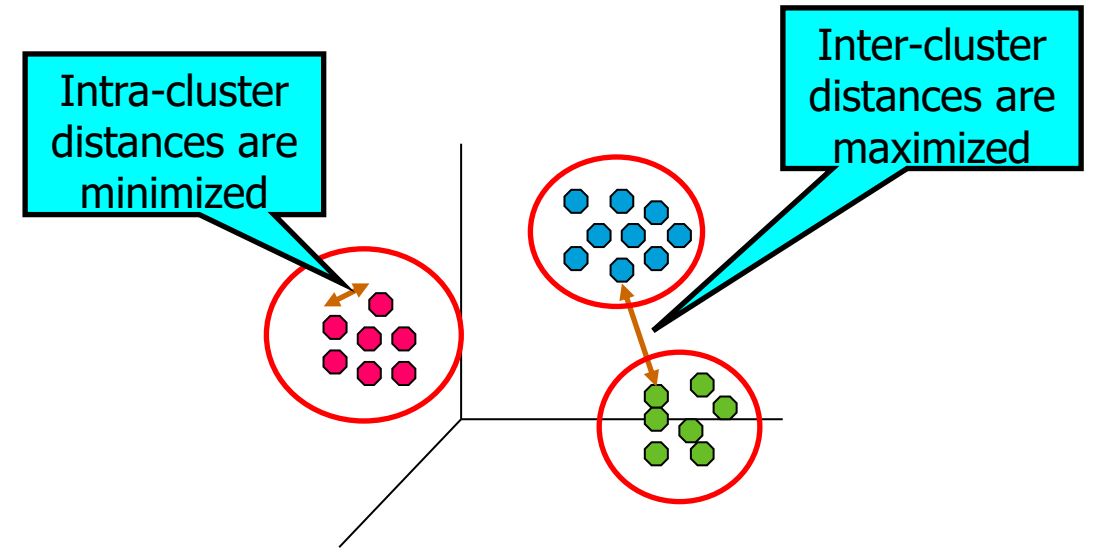
Cluster Analysis



medium.com

What is Cluster Analysis?

- Placing objects in groups such that:
 - the objects in a group are similar (or related) to one another.
 - They are different from (or unrelated to) the objects in other groups.
- We need a (metric/measure/objective function) to measure the (distance/similarity) of the (objects/clusters).



Clusters are in the eye of the beholder



How many
clusters?



Six Clusters

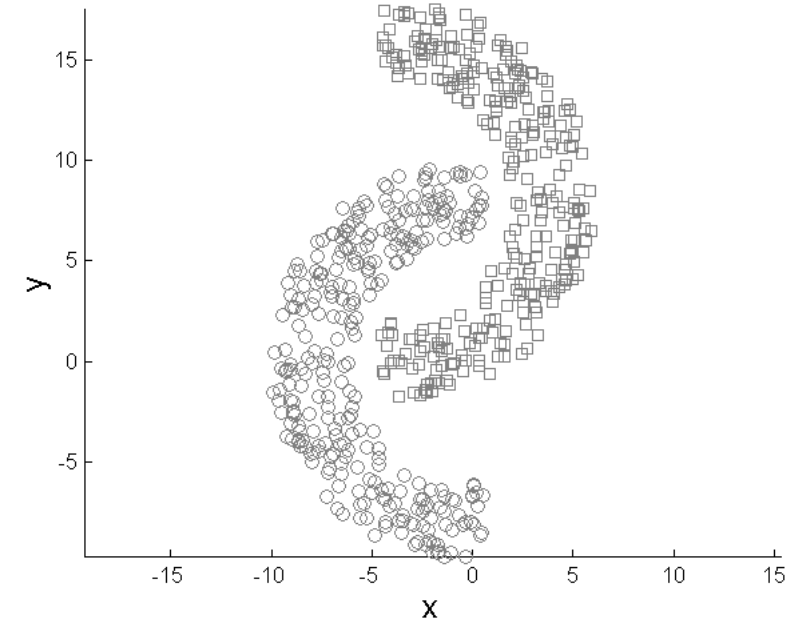
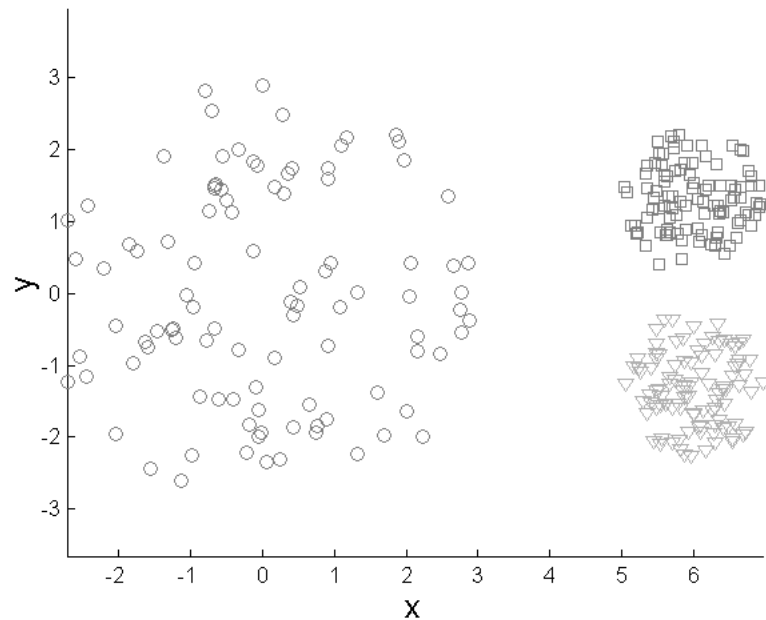


Two Clusters



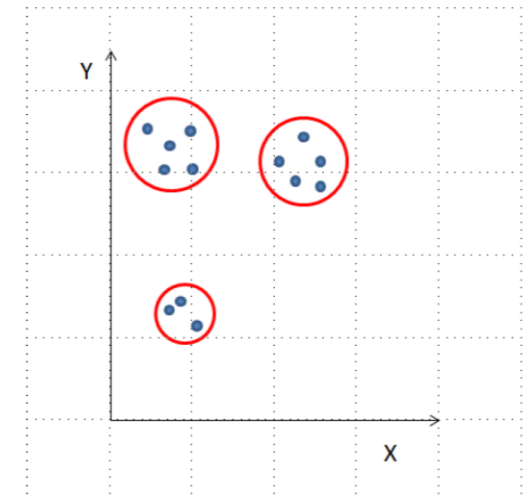
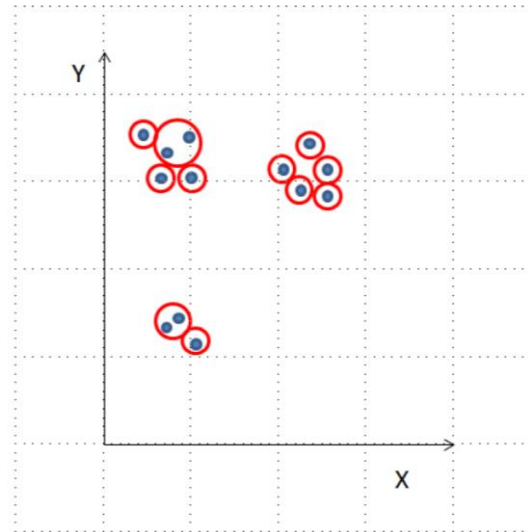
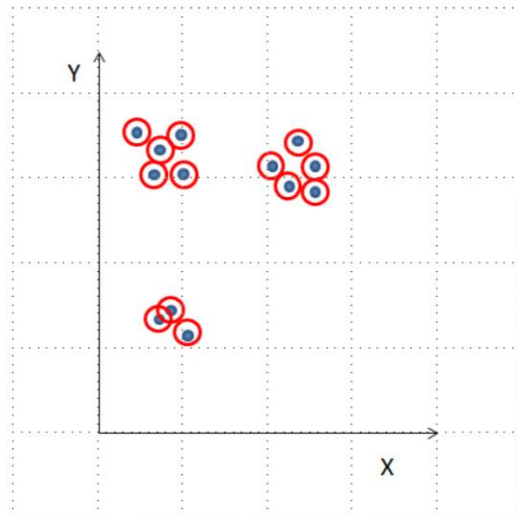
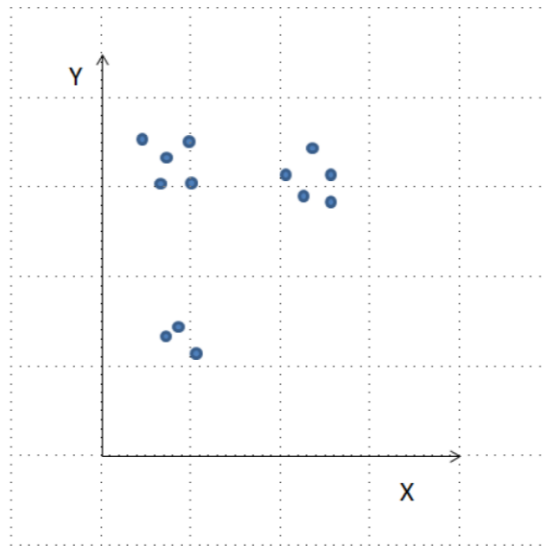
Four Clusters

Clusters come in all shapes and sizes



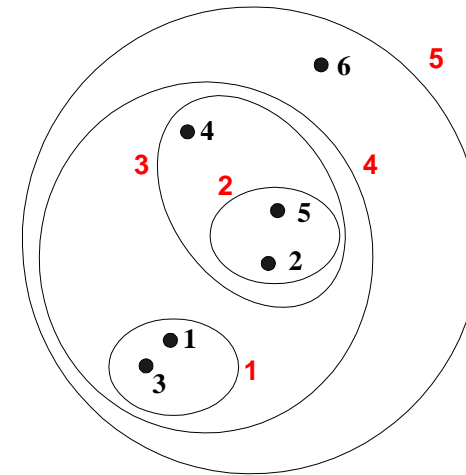
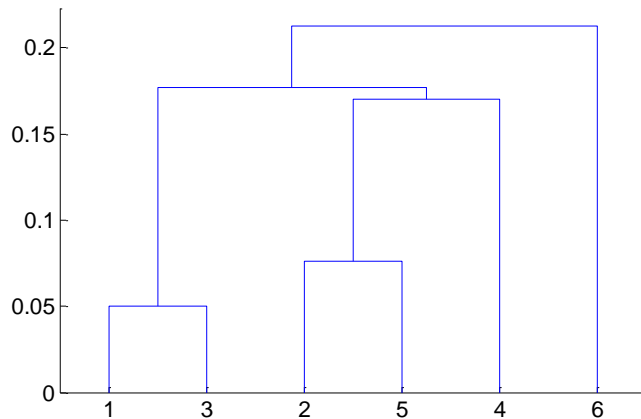
Hierarchical Clustering

Hierarchical Clustering

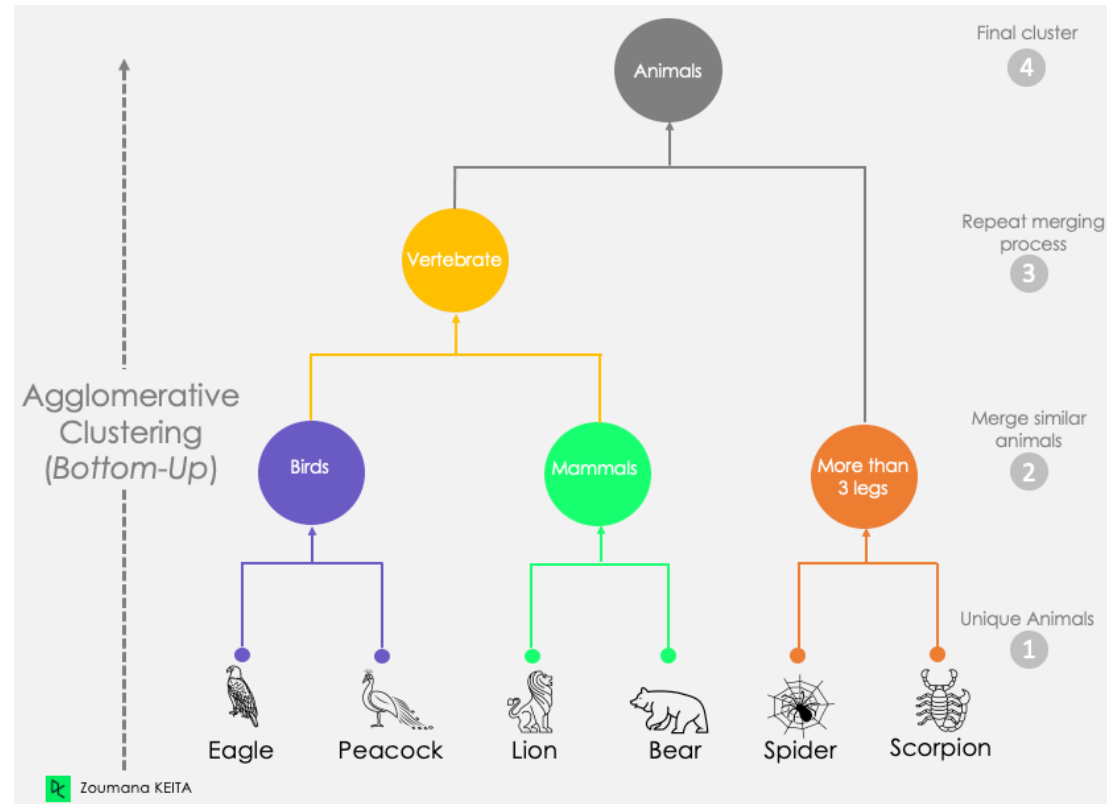


Hierarchical Clustering

- Produces a set of nested clusters organized as a hierarchical tree
- Can be visualized as a dendrogram
 - A tree like diagram that records the sequences of merges or splits.



Hierarchical Clustering

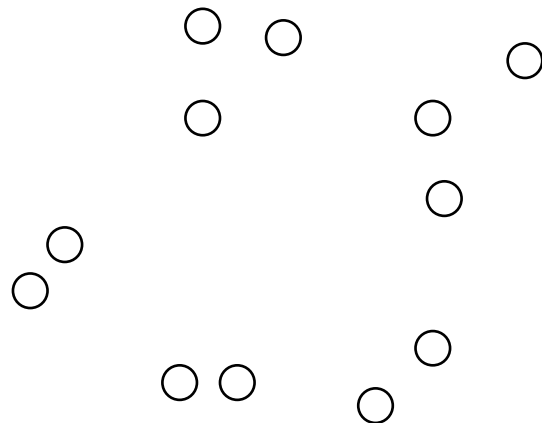


Hierarchical Clustering

- **Key Idea: Successively merge closest clusters**
- Basic algorithm
 1. Compute the proximity matrix
 2. Let each data point be a cluster
 3. **Repeat**
 4. Merge the two closest clusters
 5. Update the proximity matrix
 6. **Until** only a single cluster remains
- Key operation is the computation of the proximity of two clusters
 - Different approaches to defining the distance between clusters distinguish the different algorithms

Steps 1 and 2

- Start with clusters of individual points and a proximity matrix



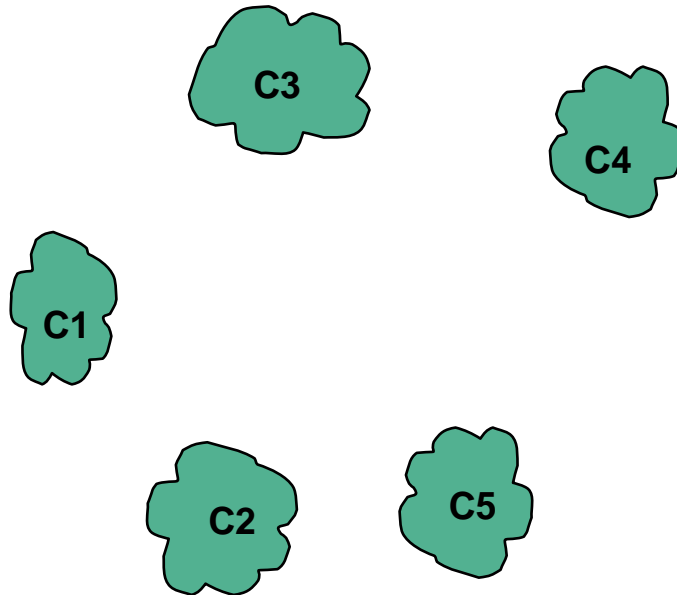
	p1	p2	p3	p4	p5	. . .
p1						
p2						
p3						
p4						
p5						
.						
.						
.						

Proximity Matrix

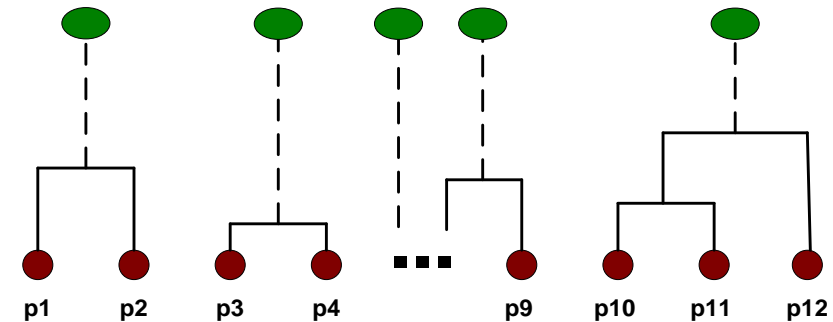


Intermediate Situation

After some merging steps, we have some clusters

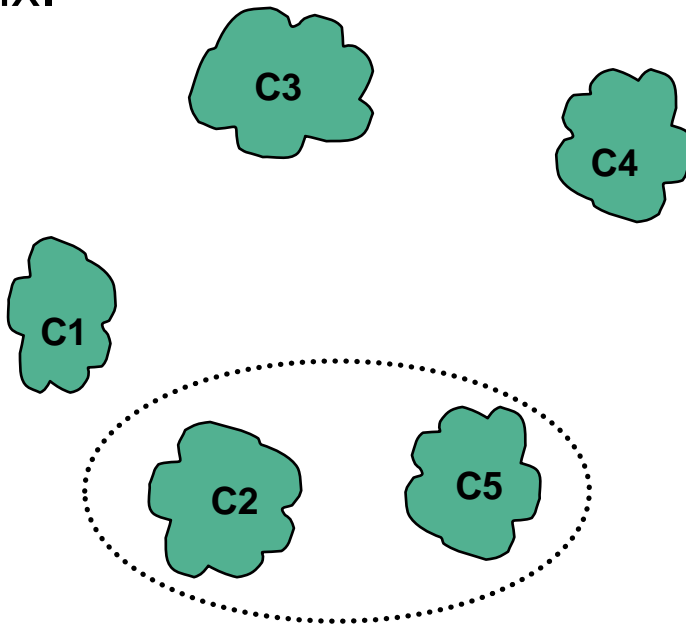


	C1	C2	C3	C4	C5
C1					
C2					
C3					
C4					
C5					



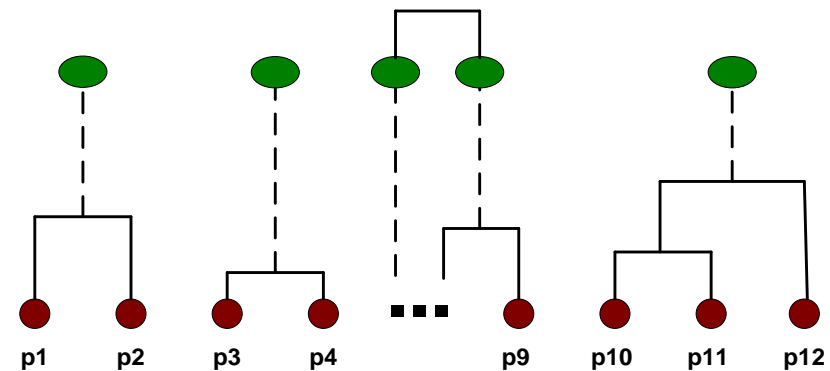
Step 4

We want to merge the two closest clusters (C2 and C5) and update the proximity matrix.



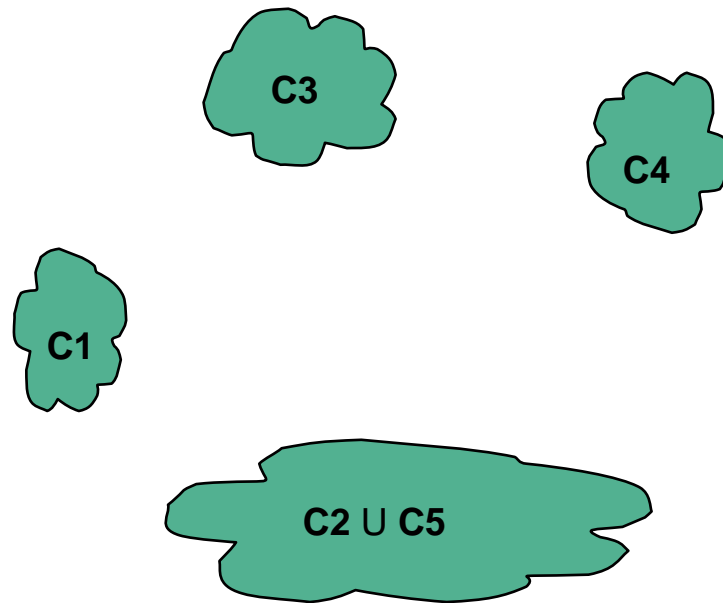
	C1	C2	C3	C4	C5
C1					
C2					
C3					
C4					
C5					

Proximity Matrix



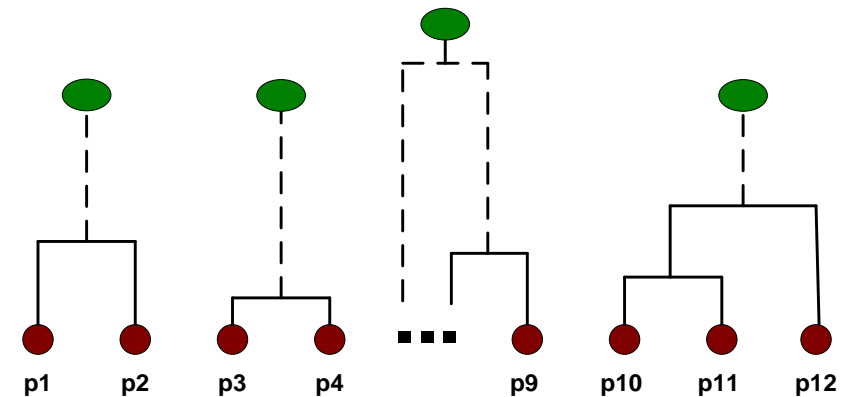
Step 5

The question is “How do we update the proximity matrix?”



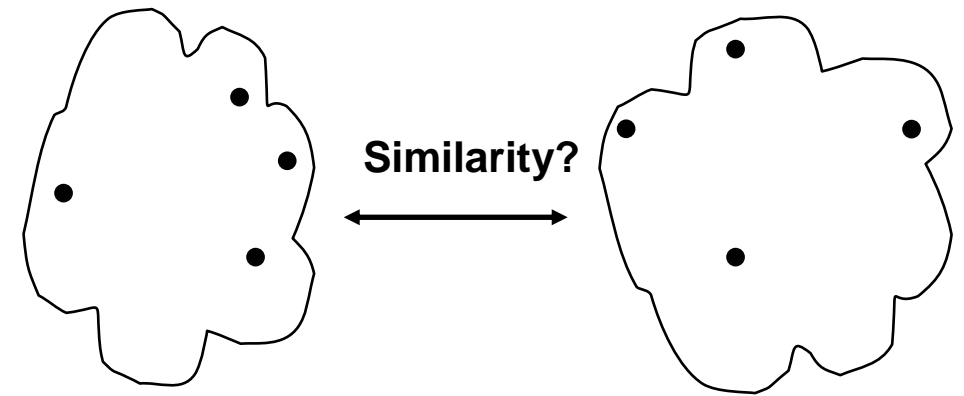
		$C2 \cup C5$	C3	C4
C1		?		
$C2 \cup C5$?	?	?	?
C3		?		
C4		?		

Proximity Matrix



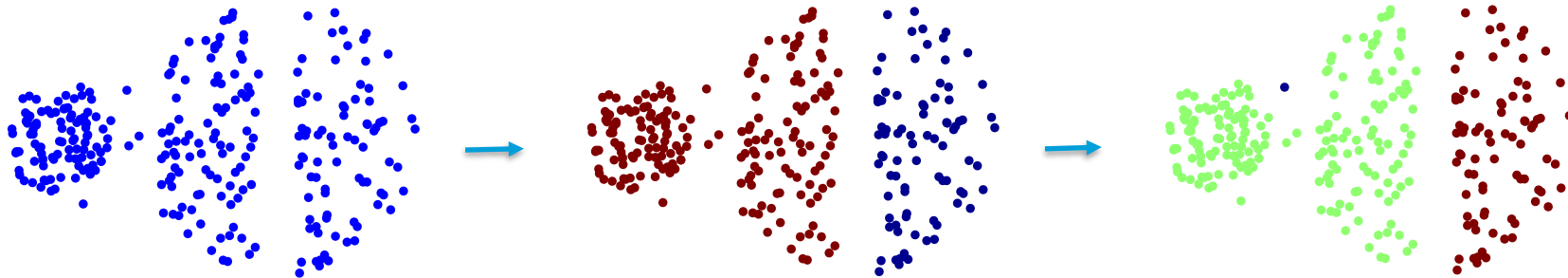
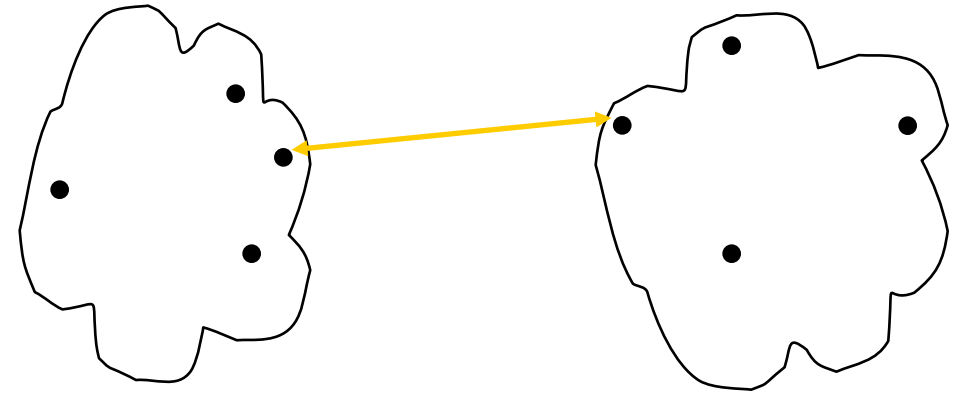
How to Define Inter-Cluster Distance

- MIN (Single Link)
- MAX (Complete Linkage)
- Group Average
 - Ward's Method uses squared error
- Distance Between Centroids



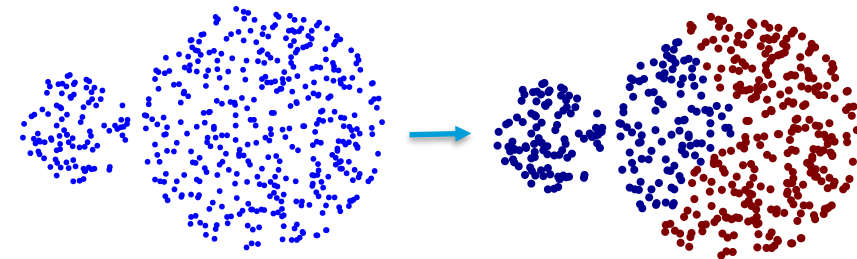
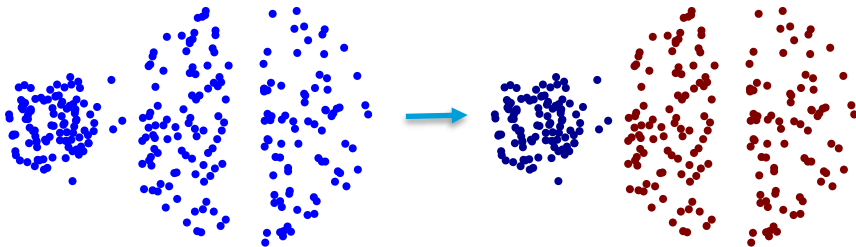
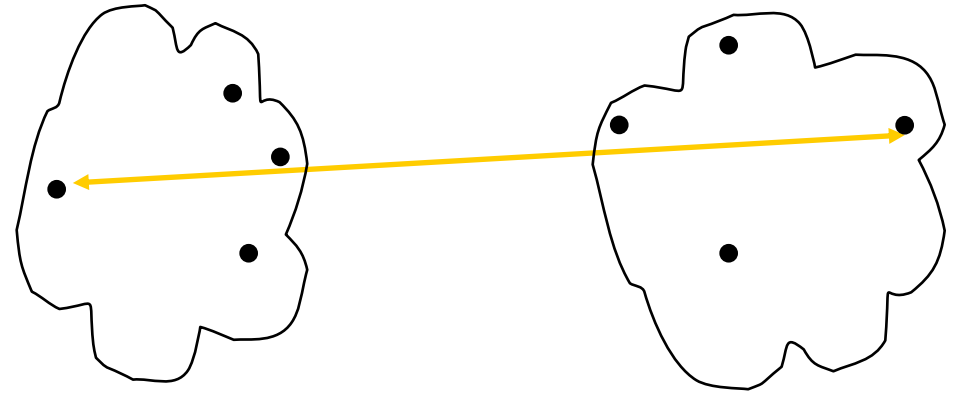
How to Define Inter-Cluster Similarity

- MIN (Single Link)
 - Sensitive to noise.



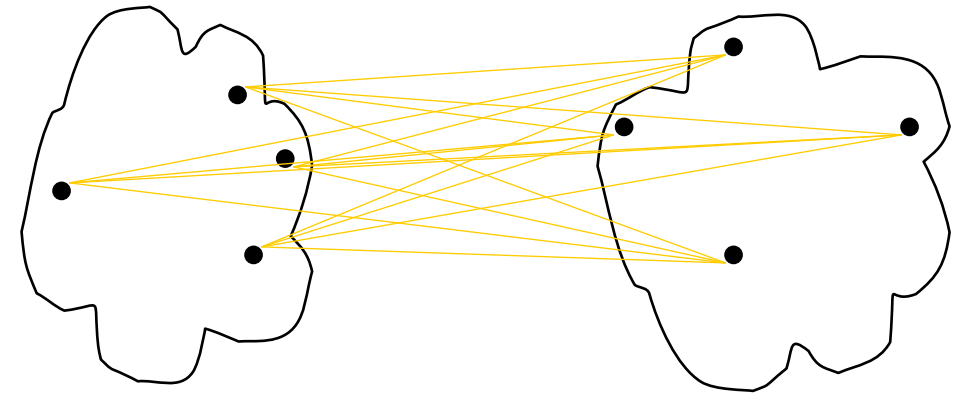
How to Define Inter-Cluster Similarity

- MAX (Complete Linkage)
 - Less susceptible to noise.
 - Breaks larger clusters.



How to Define Inter-Cluster Similarity

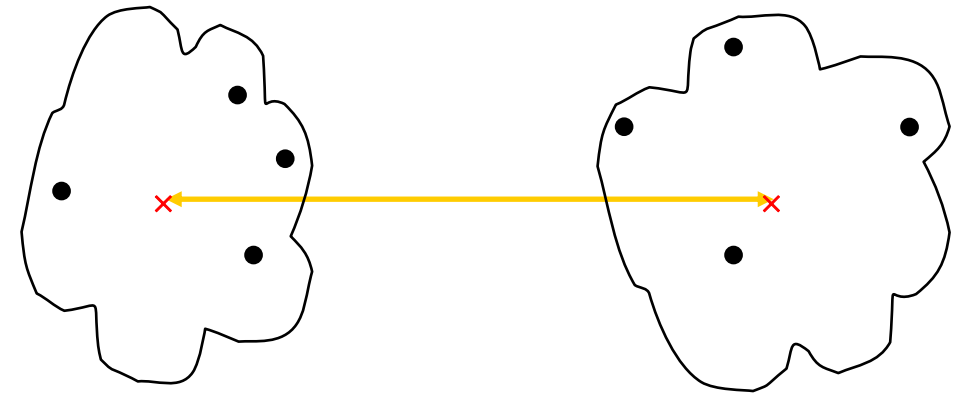
- Group Average
 - Middle ground between MIN and MAX.
 - If square distance is used, it is called Ward method.



$$\text{proximity}(\text{Cluster}_i, \text{Cluster}_j) = \frac{\sum_{\substack{p_i \in \text{Cluster}_i \\ p_j \in \text{Cluster}_j}} \text{proximity}(p_i, p_j)}{|\text{Cluster}_i| \times |\text{Cluster}_j|}$$

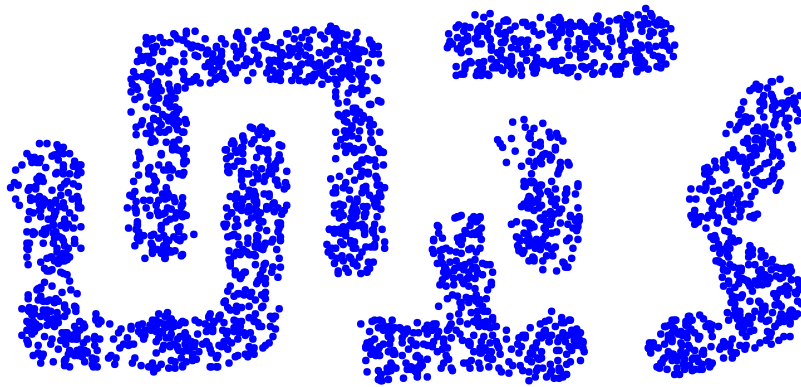
How to Define Inter-Cluster Similarity

- Distance Between Centroids

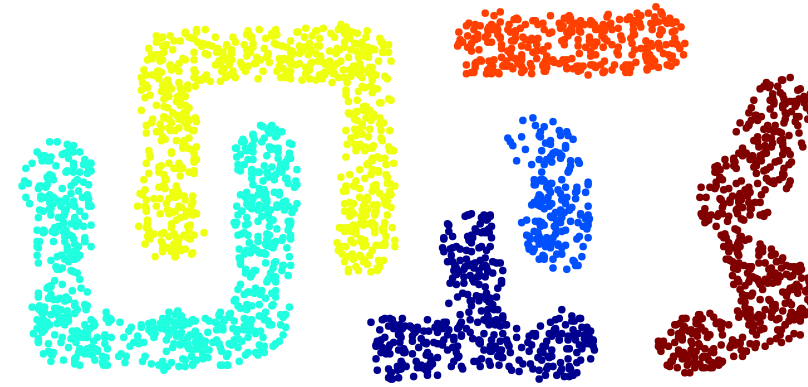


Hierarchical Clustering

Can handle non-elliptical shapes



Original Points



Six Clusters

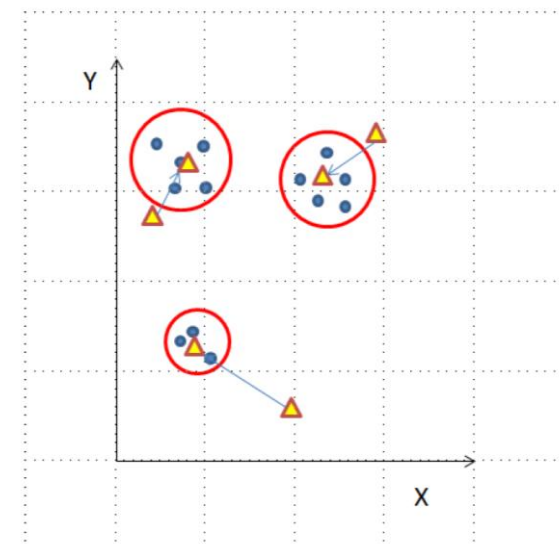
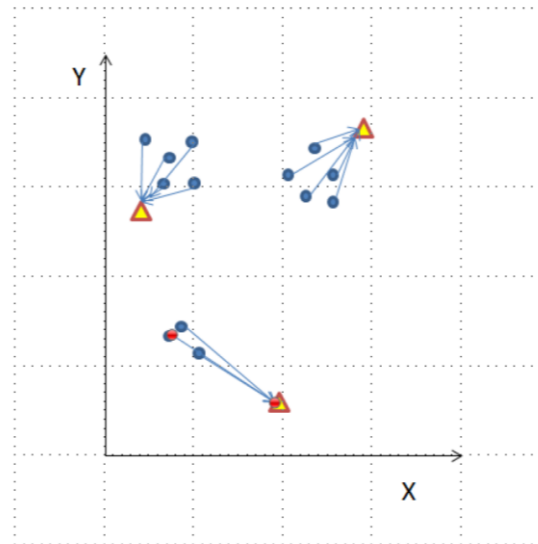
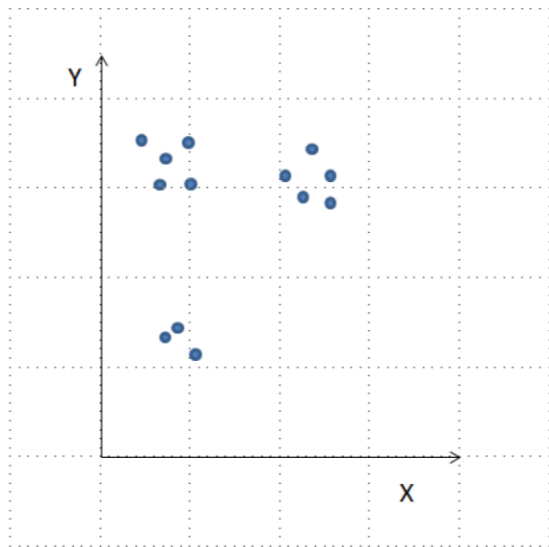
Hierarchical Clustering

- Visualization + dendograms (easy to find answer when changing number of clusters)
- Hierarchical has a high time complexity (polynomial $O(n^3)$).

Demo time!

Partitional Clustering: K-means

K-means Clustering

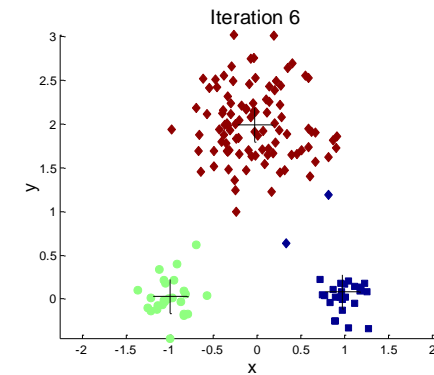
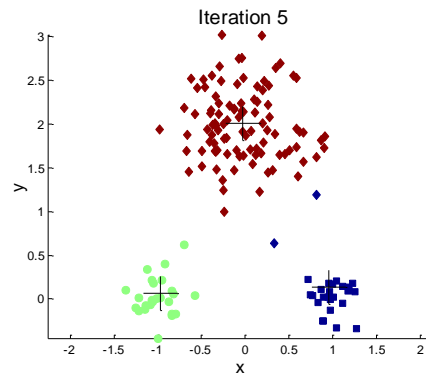
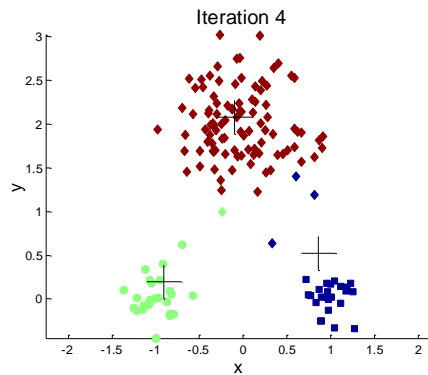
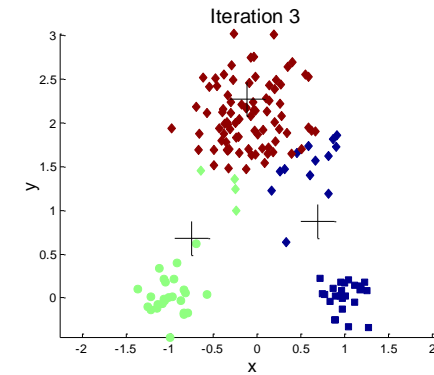
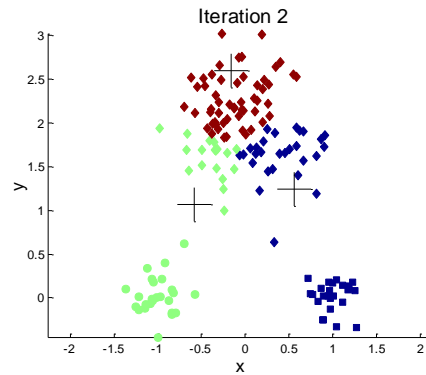
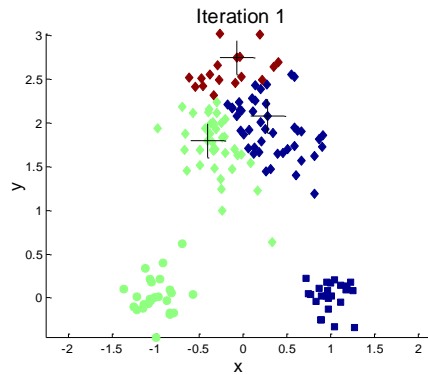


K-means Clustering

- Number of clusters, K , must be specified in advance.
- Each cluster is associated with a *centroid* (center point).
- Each point is assigned to the cluster with the closest centroid.
- The basic algorithm is very simple

-
- 1: Select K points as the initial centroids.
 - 2: **repeat**
 - 3: Form K clusters by assigning all points to the closest centroid.
 - 4: Recompute the centroid of each cluster.
 - 5: **until** The centroids don't change
-

Example of K-means Clustering



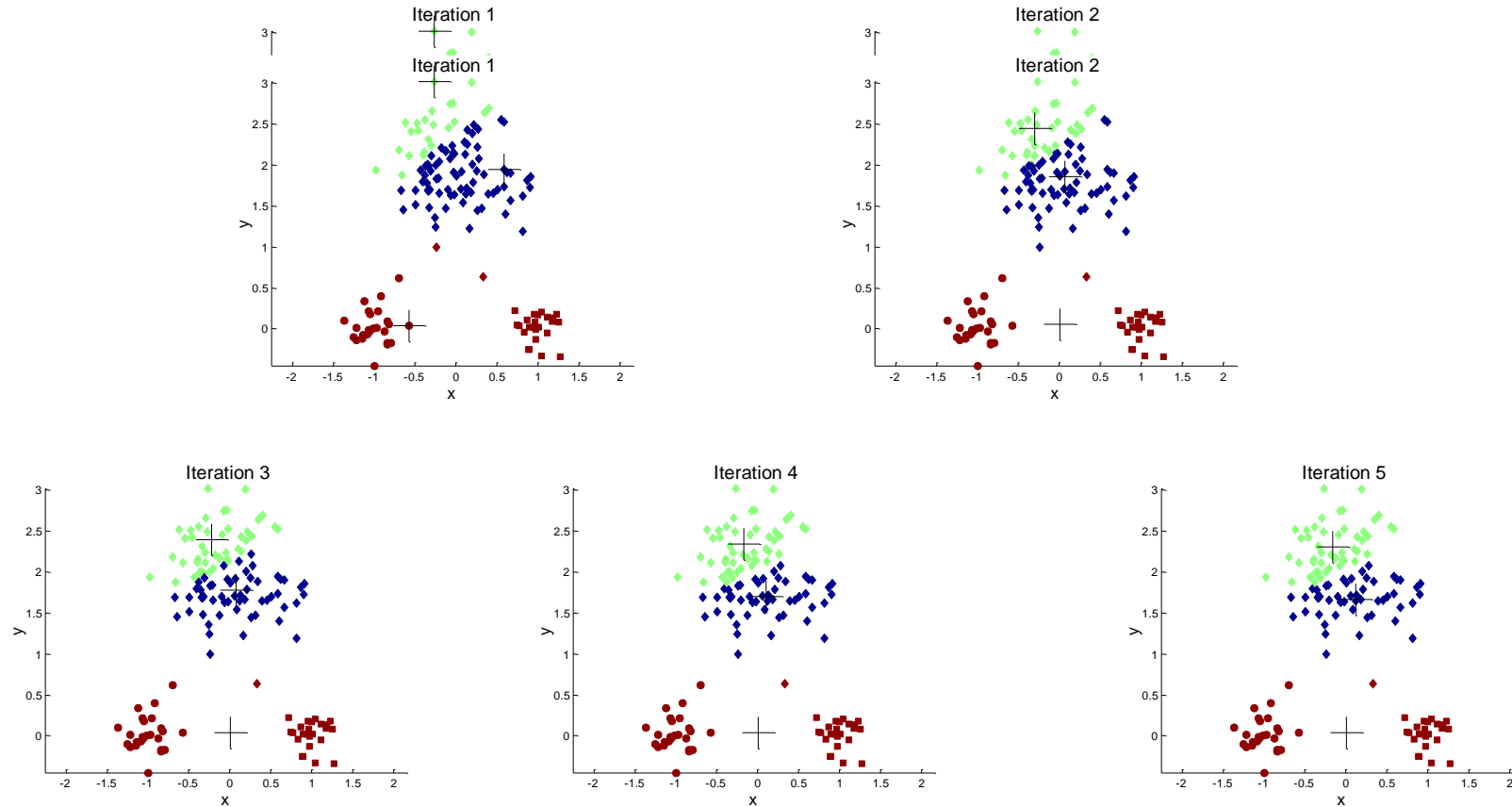
K-means Objective Function

- Euclidean Sum of Squared Error (SSE)
 - For each point, the error is the distance to the nearest cluster center.
 - To get SSE, we square these errors and sum them.

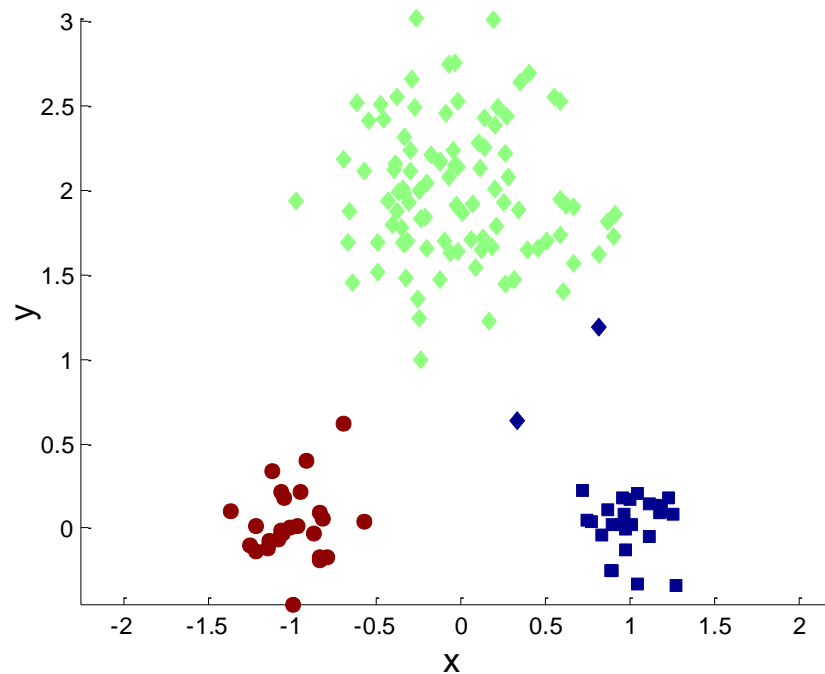
$$SSE = \sum_{i=1}^K \sum_{x \in C_i} dist^2(m_i, x)$$

- This is also called distortion or inertia.
- x is a data point in cluster C_i and m_i is the centroid (mean) for cluster C_i
- SSE improves in each iteration of K-means until it reaches a local or global minima.

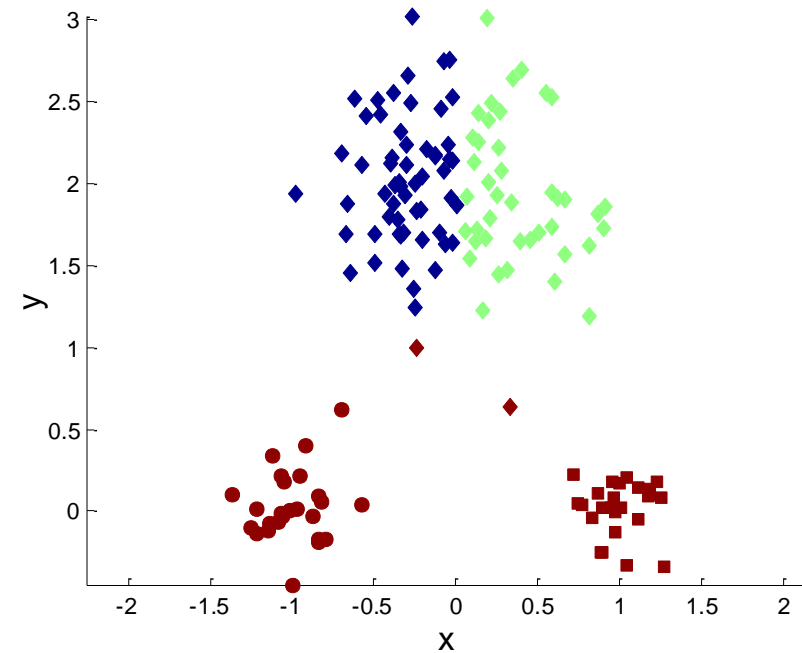
Effect of Random Initialization



Effect of Random Initialization



Optimal Clustering

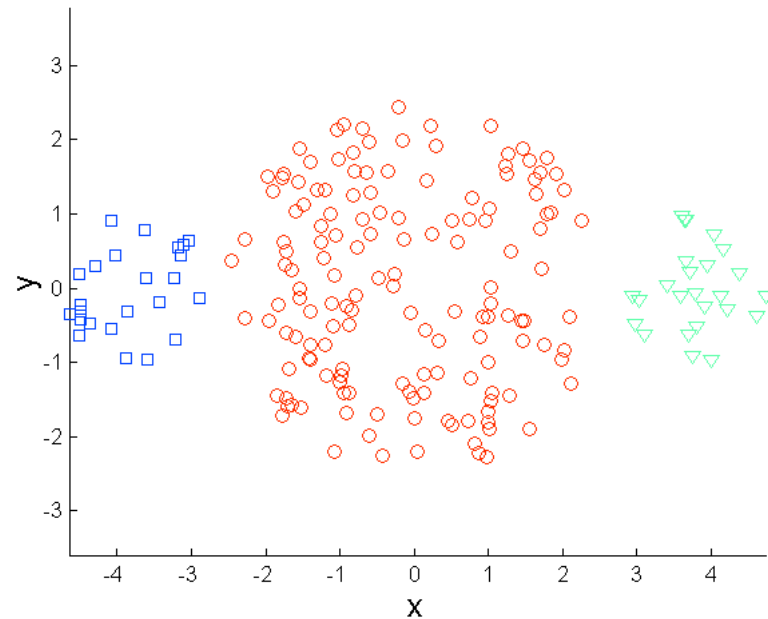


Sub-optimal Clustering

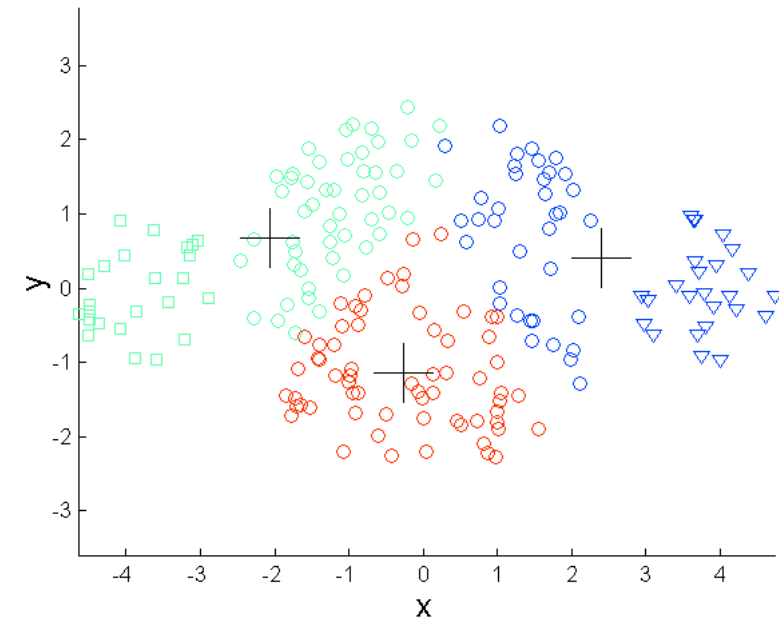
Limitations of K-means

- K-means has problems when clusters are of differing
 - Sizes
 - Densities
 - Non-globular shapes
- K-means is faster than hierarchical clustering.
- K-means is susceptible to suboptimal initialization.
- What do [boundaries](#) between clusters look like?
- K-means has problems when the data contains outliers.
 - One possible solution is to remove outliers before clustering

Limitations of K-means: Differing Sizes

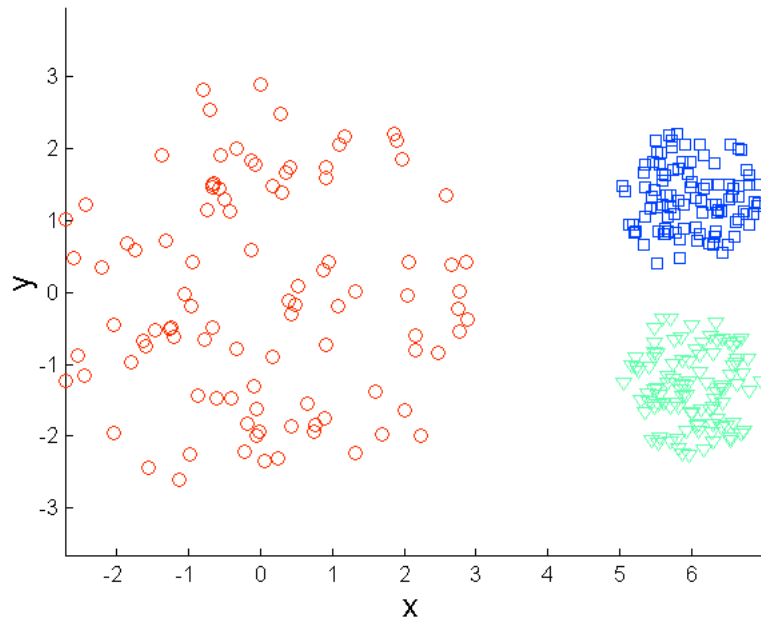


Original Points

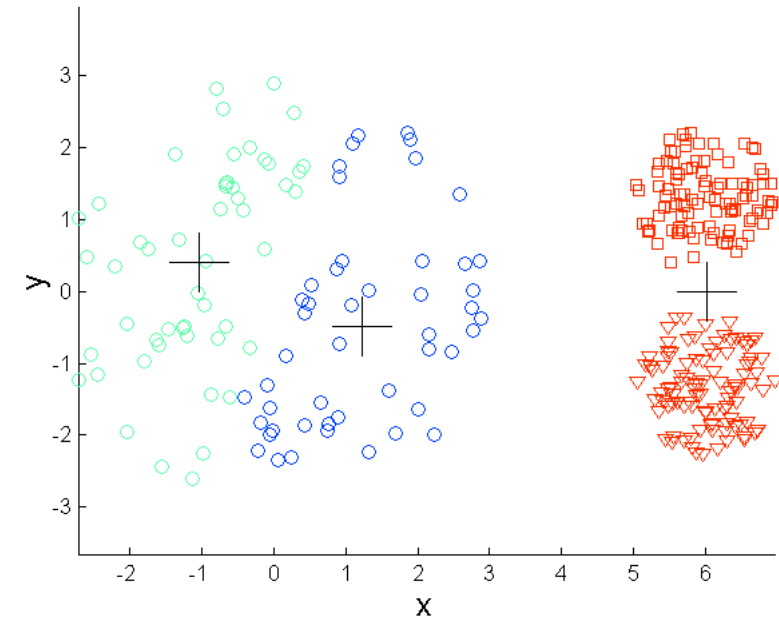


K-means (3 Clusters)

Limitations of K-means: Differing Density



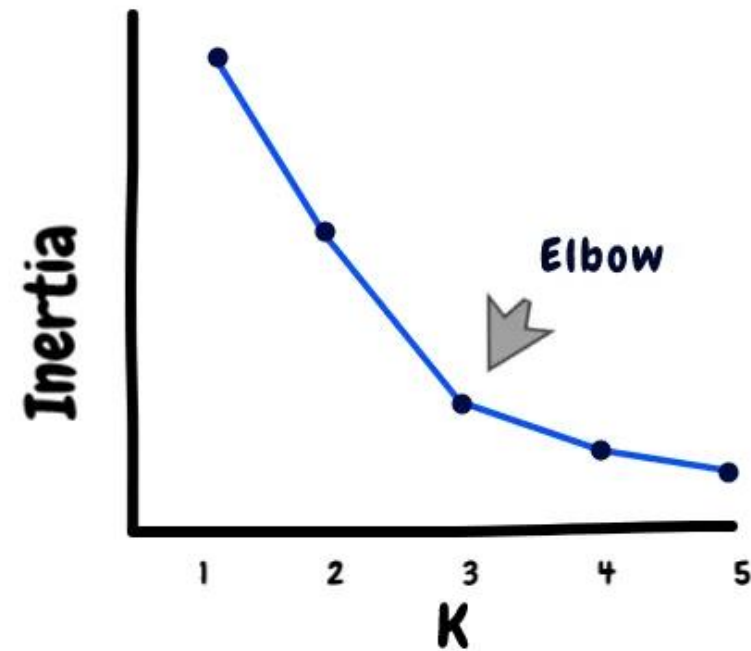
Original Points



K-means (3 Clusters)

Deciding the number of clusters

Inertia: Sum of squared distances of samples to their respective closest cluster centers.



towardsdatascience.com

Measuring Clustering Quality

Measuring Cluster Validity Via Correlation

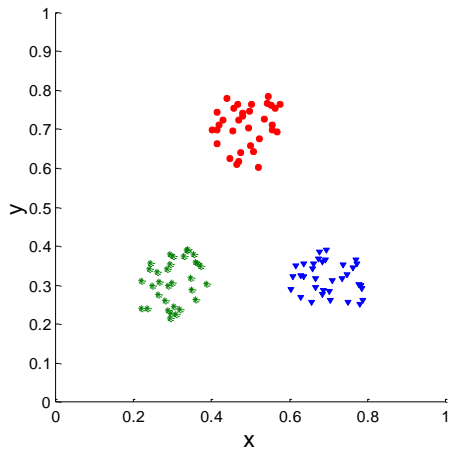
- Compute the correlation between the two matrices

Since the matrices are symmetric, only the correlation between $n(n-1) / 2$ entries needs to be calculated.

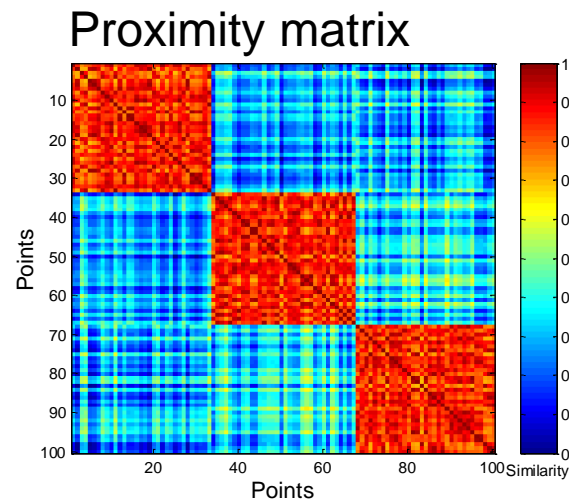
- High magnitude of correlation indicates that points that belong to the same cluster are close to each other.
- Not a good measure for hierarchical clustering.

Method 2: via Correlation

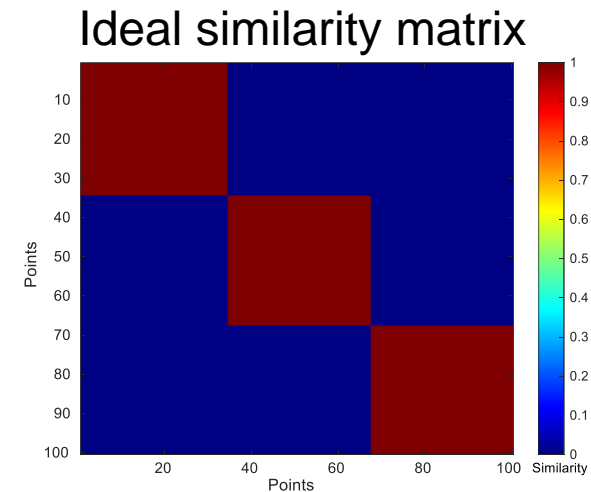
- Correlation of ideal similarity and proximity matrices for the K-means clustering



well-clustered data set

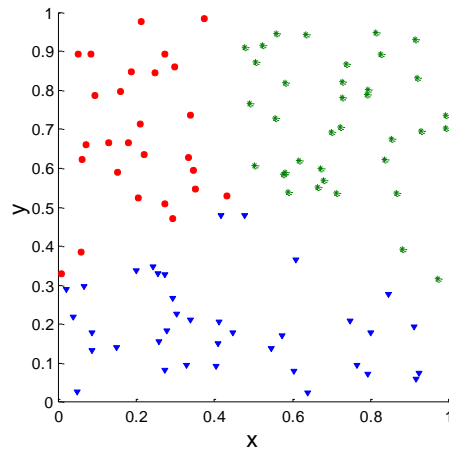


Corr = 0.9235

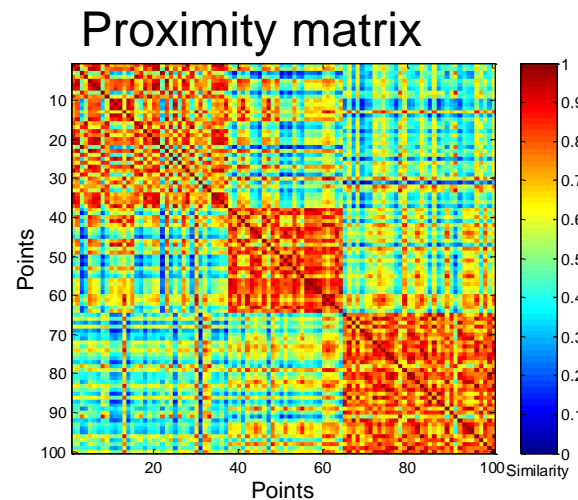


Method 2: via Correlation

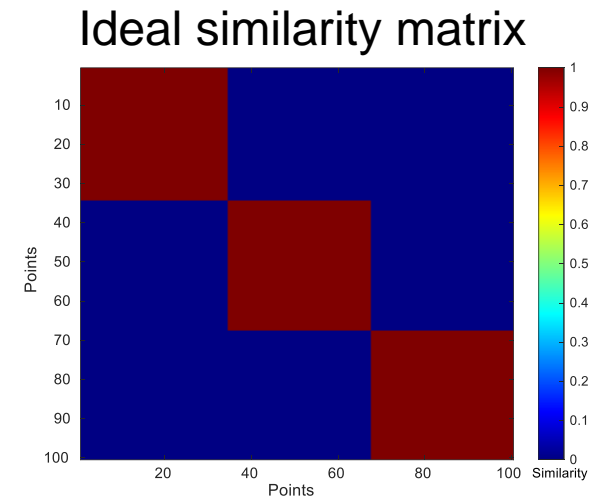
- Correlation of ideal similarity and proximity matrices for the K-means clustering



poorly-clustered data set



Corr = 0.5810



Method 1: via Labels

- Testing clustering using supervised learning.
- Calculate the error between ideal labelling and assigned cluster label.