FUZZY CLUSTERING

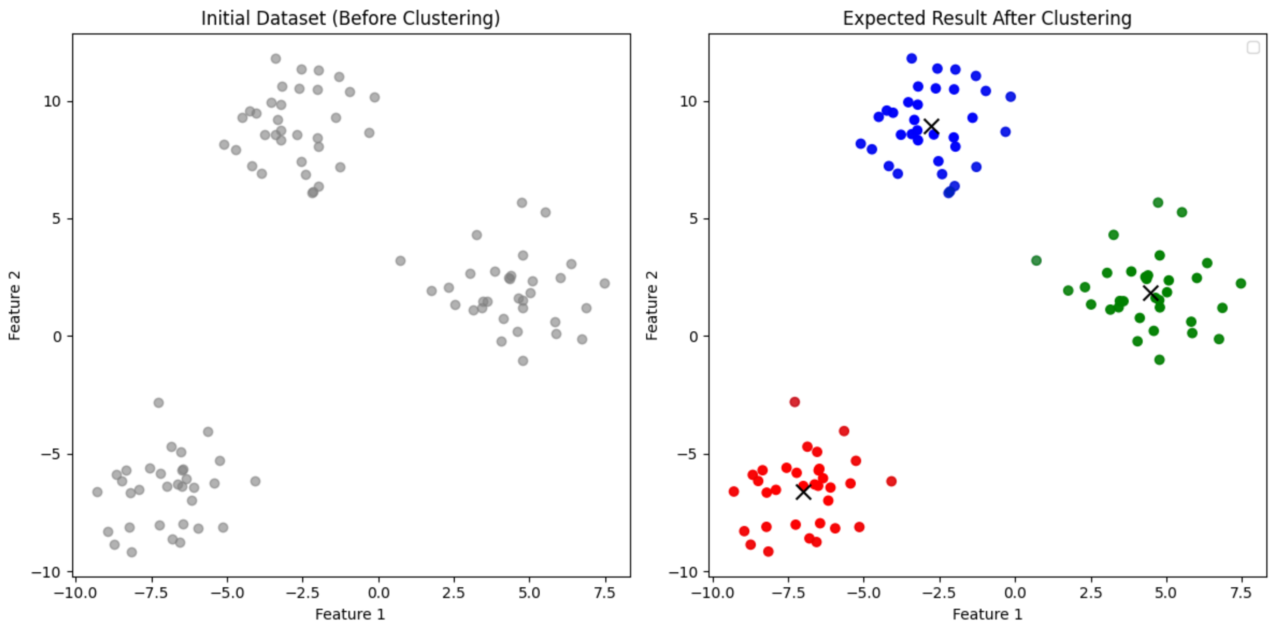
By Alessandro Gentili and Jakub Swistak

Content

- Clustering
- K-Means Clustering
- Fuzzy C-Means Clustering
- Possibilistic Fuzzy C-Means Clustering
- KNN Clustering
- Fuzzy KNN Clustering
- Fuzzy Scaling Process

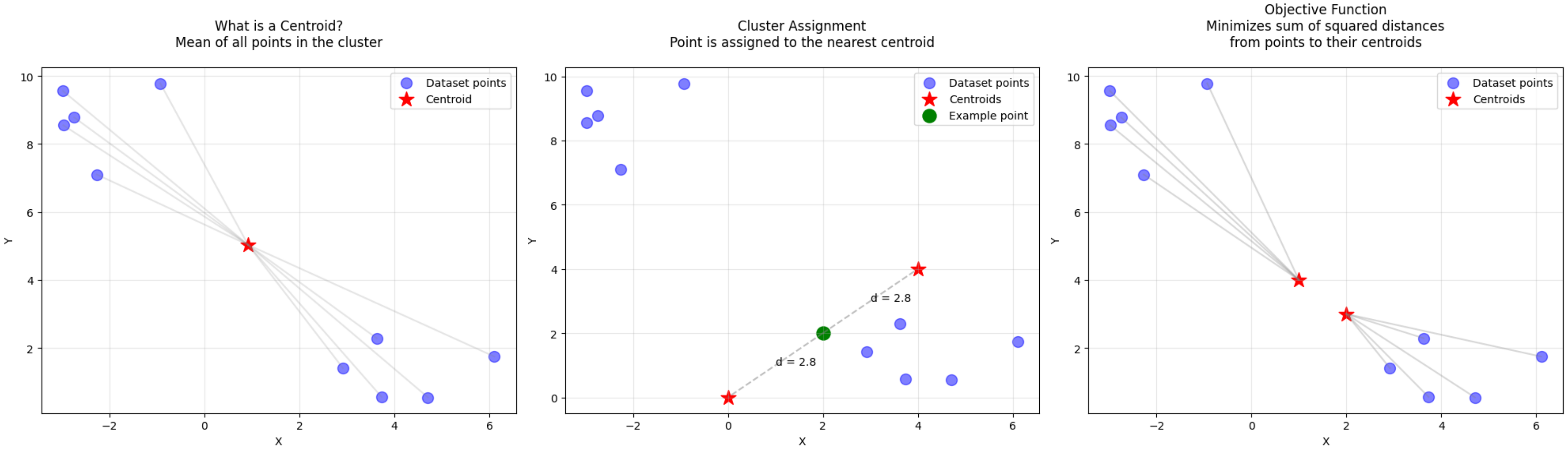
What is clustering?

- Clustering is the task of partitioning a set of elements into groups (clusters) such that the elements belonging to the same group are similar (in some way)
- Elements are represented as **vectors**
- We can compare elements using **distance**



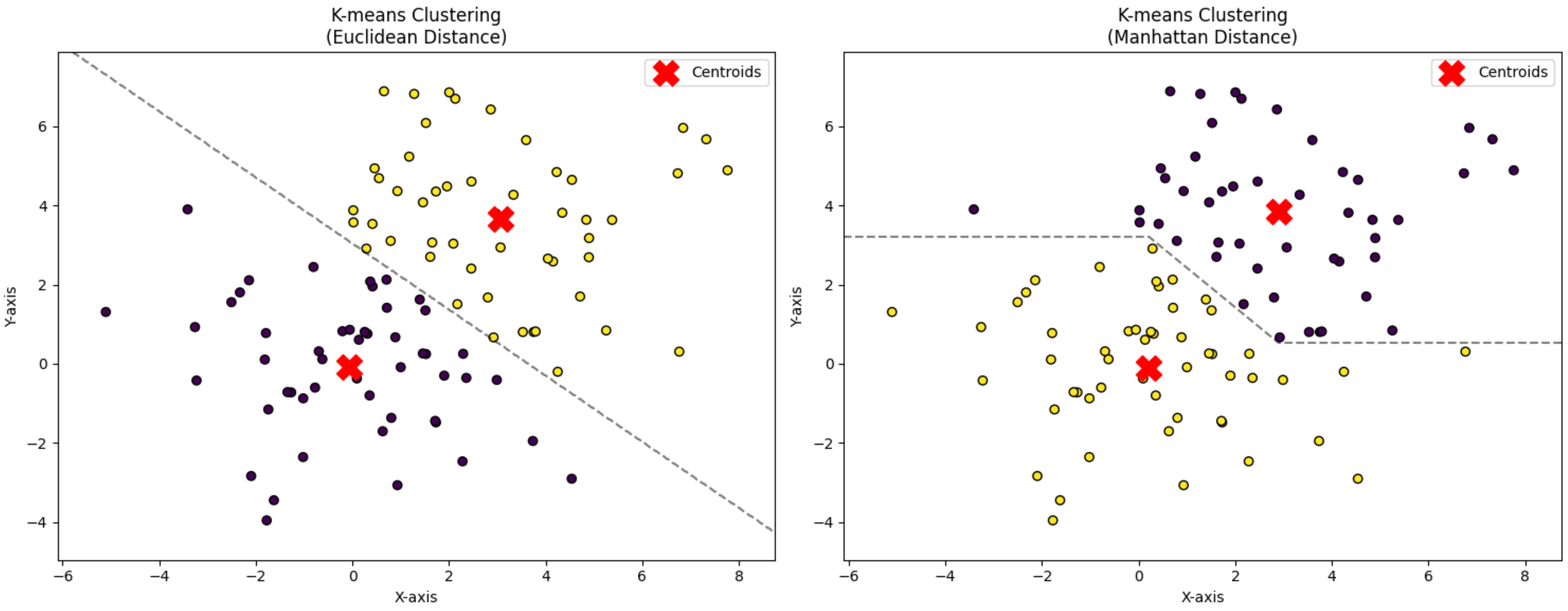
Preliminaries

- Set of vectors: $x = \{x_1, x_2, \ldots, x_n\}$
- Centroid: $\mathbf{c} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$
- Distance: $d_{Euclidean}(\mathbf{a},\mathbf{b}) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2}$ $d_{Manhattan}(\mathbf{a},\mathbf{b}) = \sum_{i=1}^{n}|a_i - b_i|$
- Loss function: $\mathcal{L}(x,c) = \sum_{k=1}^{K}\sum_{i=1}^{n}\|x_i - c_k\|^2$

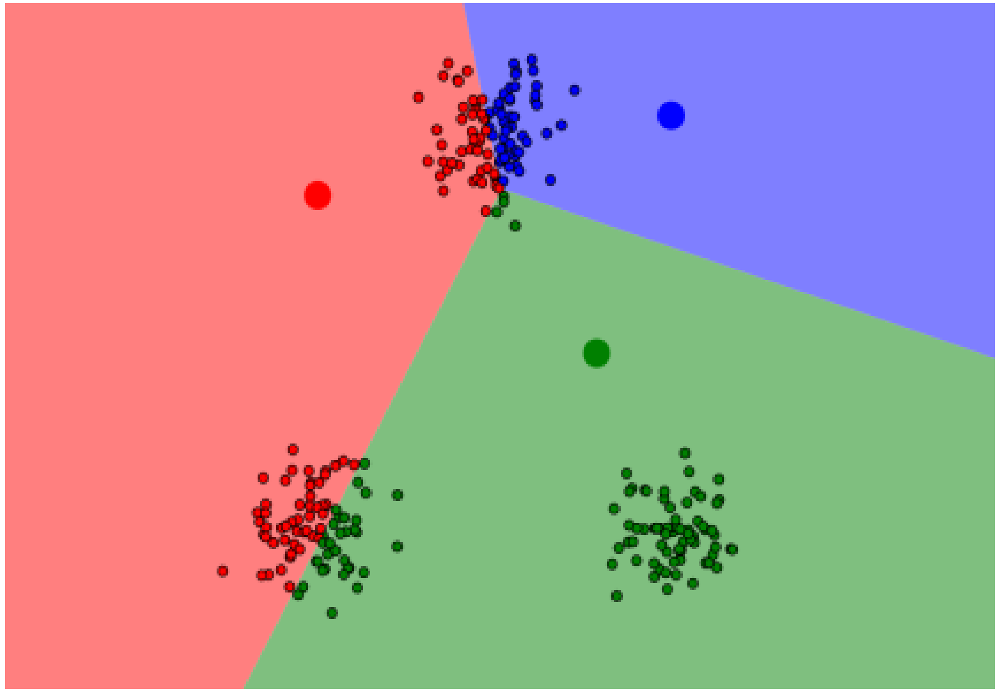


Distance matter

If we change the distance we change the **decision boundaries**, therefore we get different clustering results

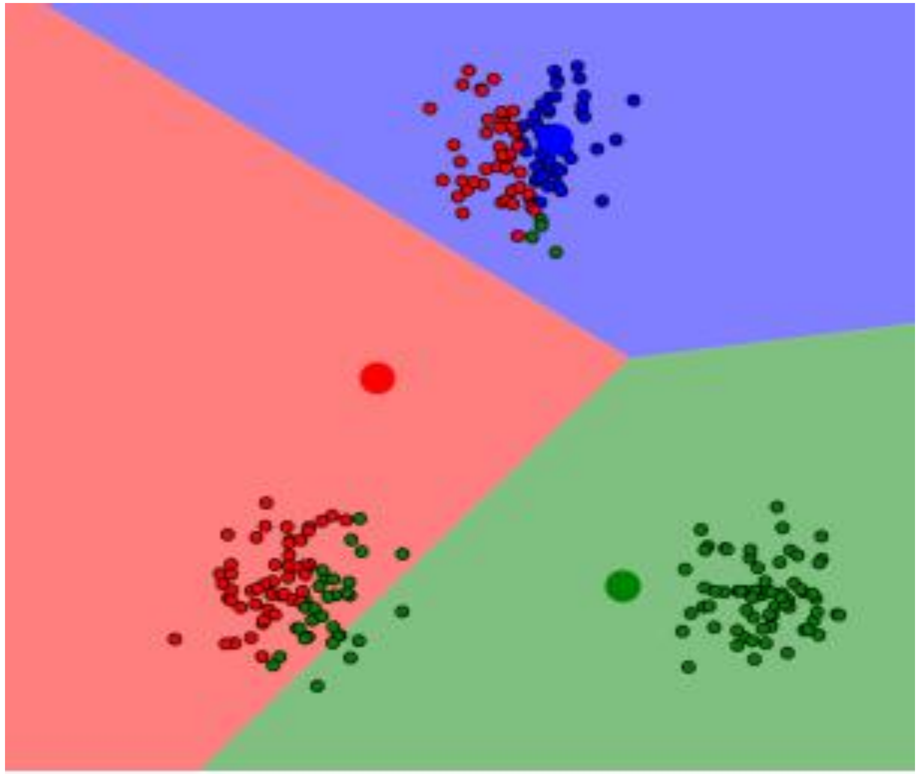


K-Means Algorithm – step 1



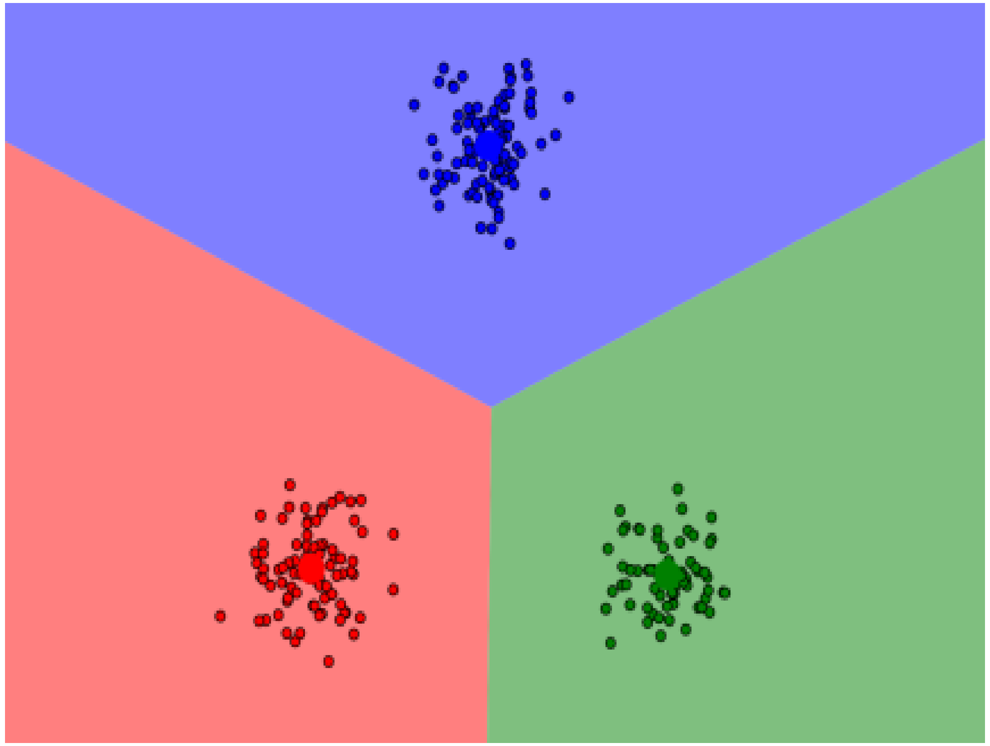
- Select k the number of clusters
- Assign k random centroids
- Assign each point to the closest centroid and create the clusters

K-Means Algorithm – step 2



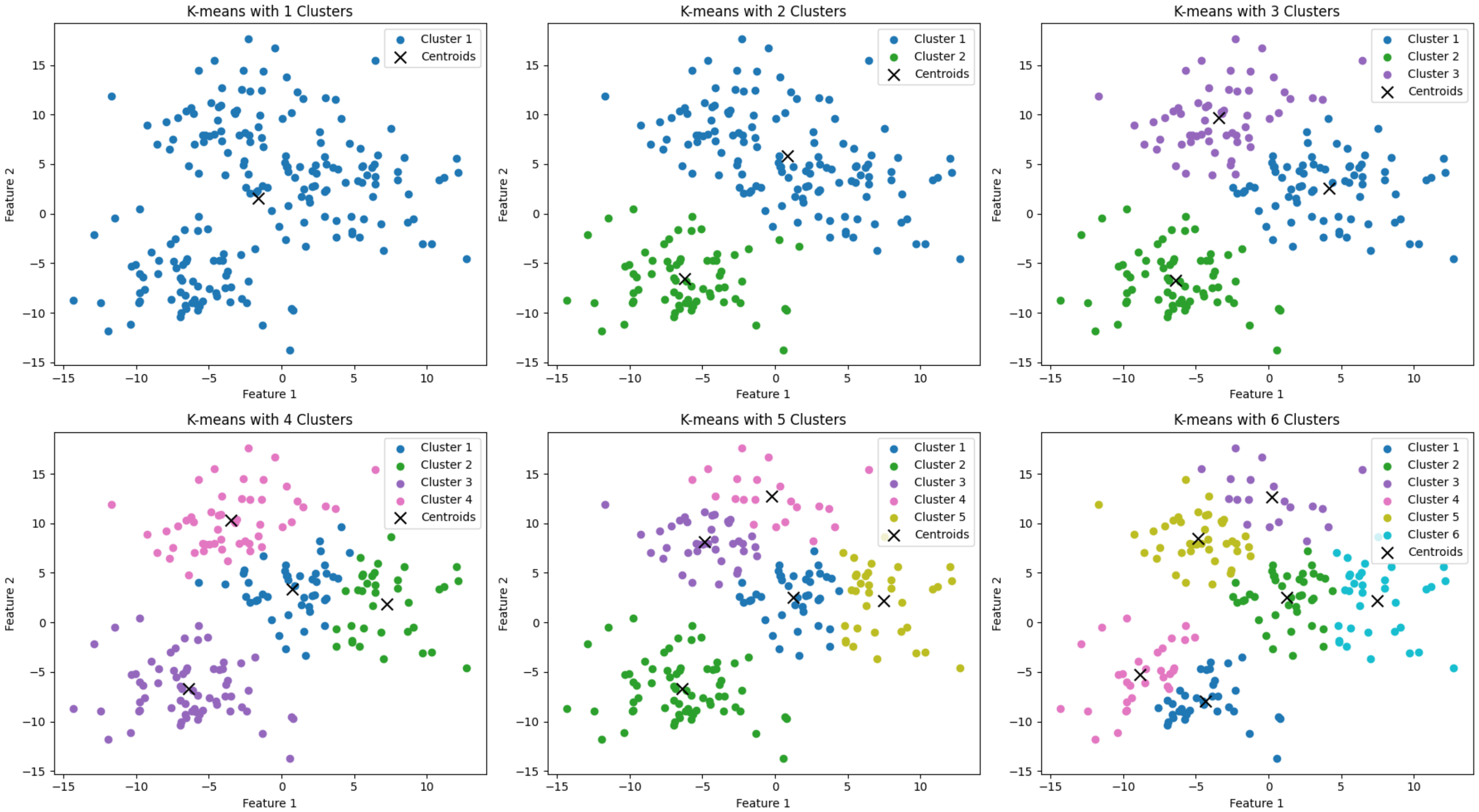
- Compute centroids for the new clusters
- Reassign each point to the closest centroid

K-Means Algorithm – step 3

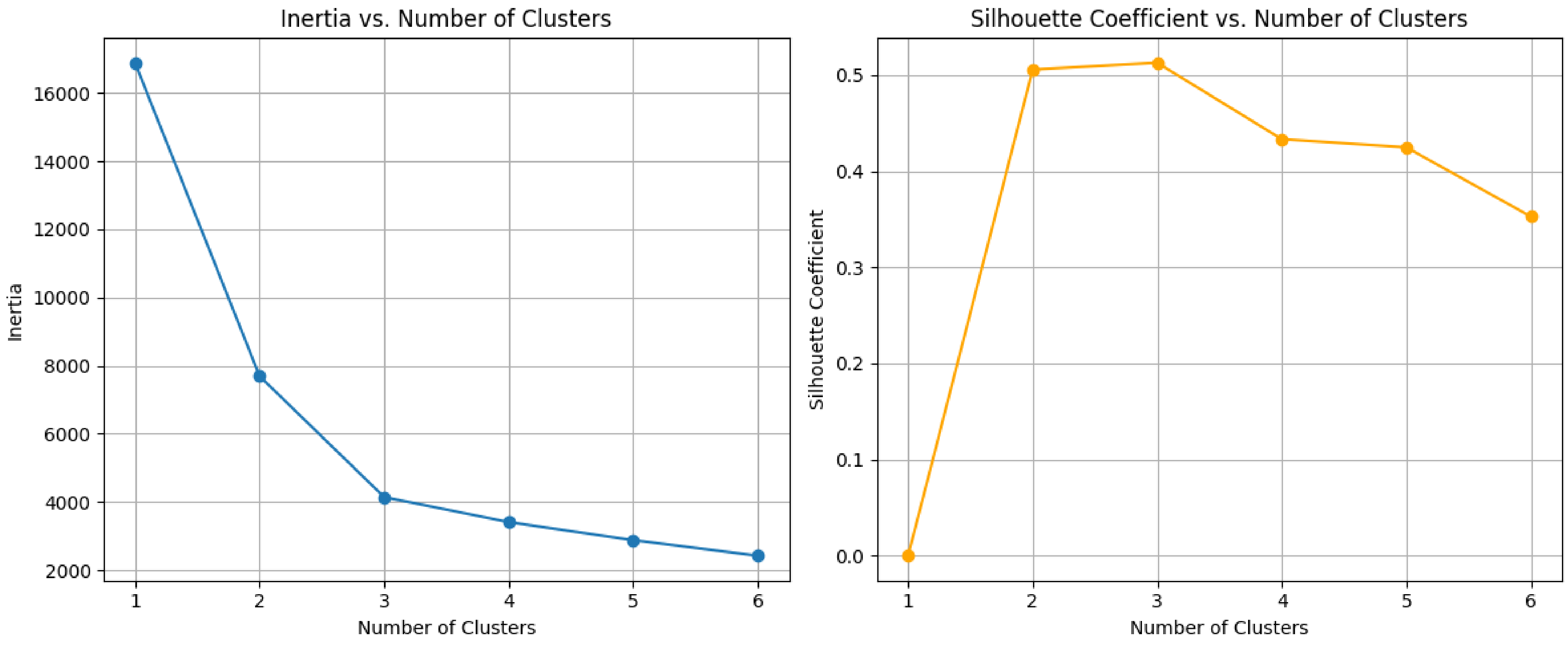


› Repeat the process until **convergence**, meaning that the loss does not improve significantly any more

K-Means Algorithm – parameter tuning



K-Means Algorithm – parameter tuning



Fuzzy C-Means – additional elements

- We need a new data structure:

$$U = [u_{ij}] \qquad 0 \leq u_{ij} \leq 1, \quad \sum_{j=1}^{k} u_{ij} = 1, \quad \forall i$$

- For each cluster the centroid is computed as follows:

$$c_j = \frac{\sum_{i=1}^{N} u_{i,j}^{m} x_i}{\sum_{i=1}^{N} u_{i,j}^{m}}$$

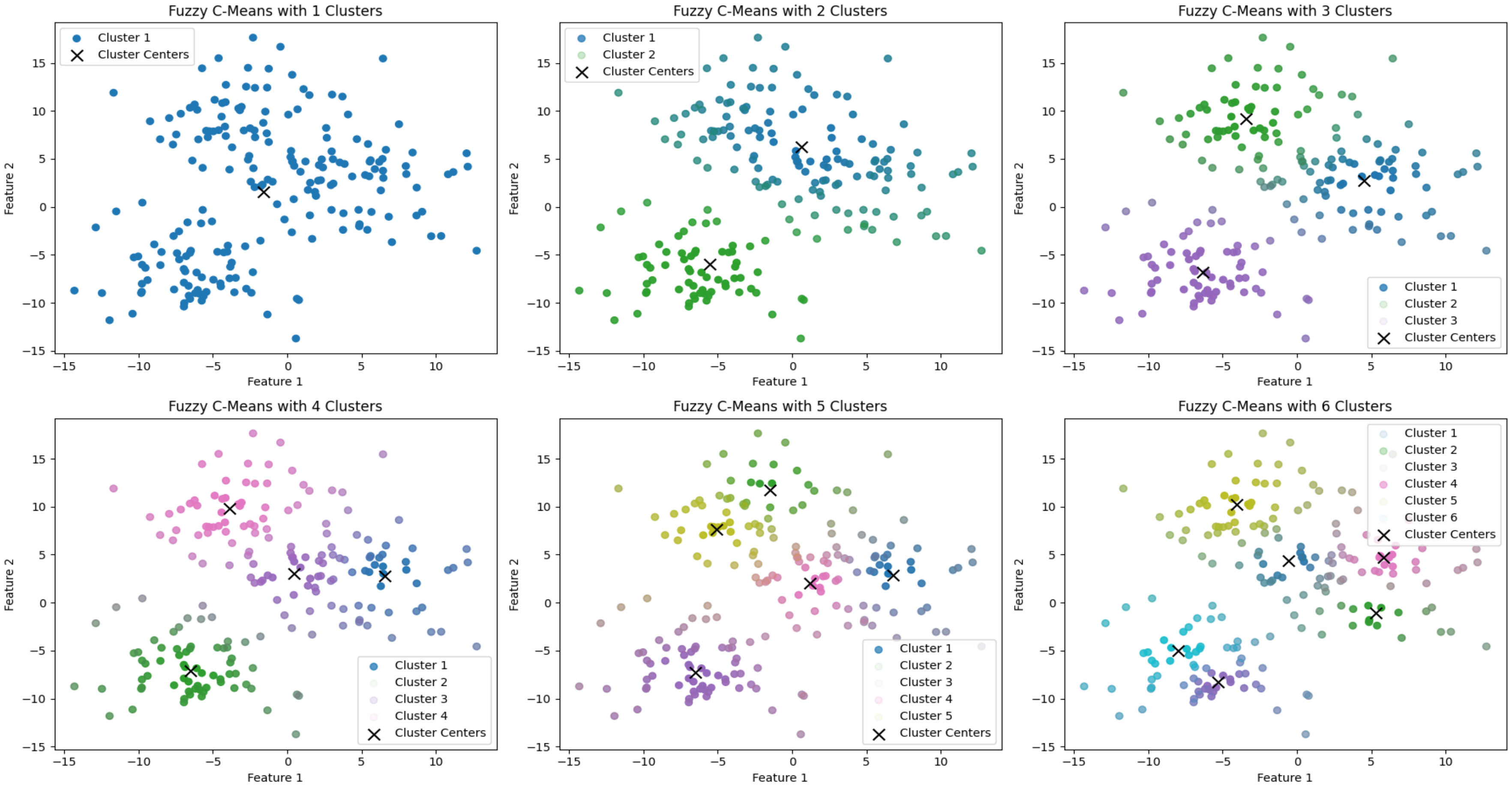
- For each element the degree of membership is updated as follows:

$$u_{i,j} = \left(\sum_{k=1}^{c} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}} \right)^{-1}$$

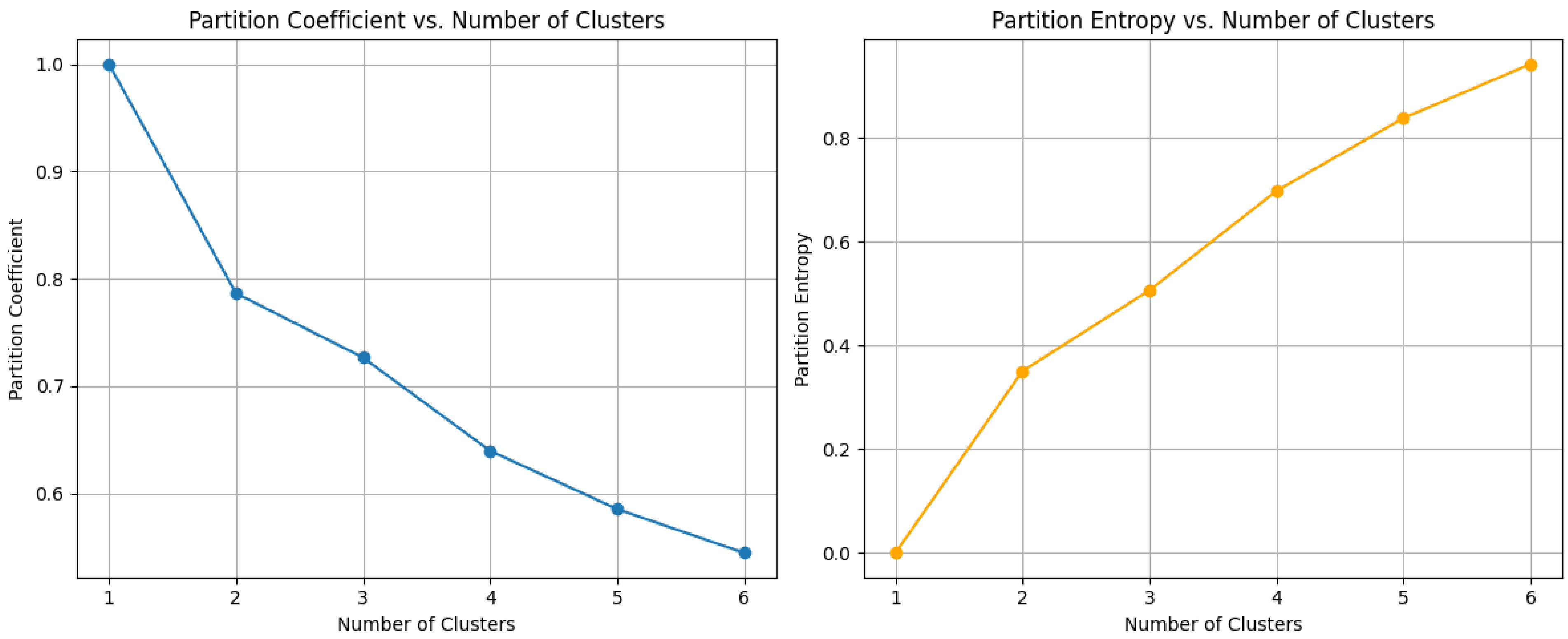
Fuzzy C-Means - algorithm

- Select the number of clusters to detect
- Assign centroids randomly
- Assign to each point a **degree of membership** to each cluster
- Repeat untill convergence:
  - Compute the new centroids
  - Assign each point a new degree of membership to each cluster

Fuzzy C-Means – parameter tuning



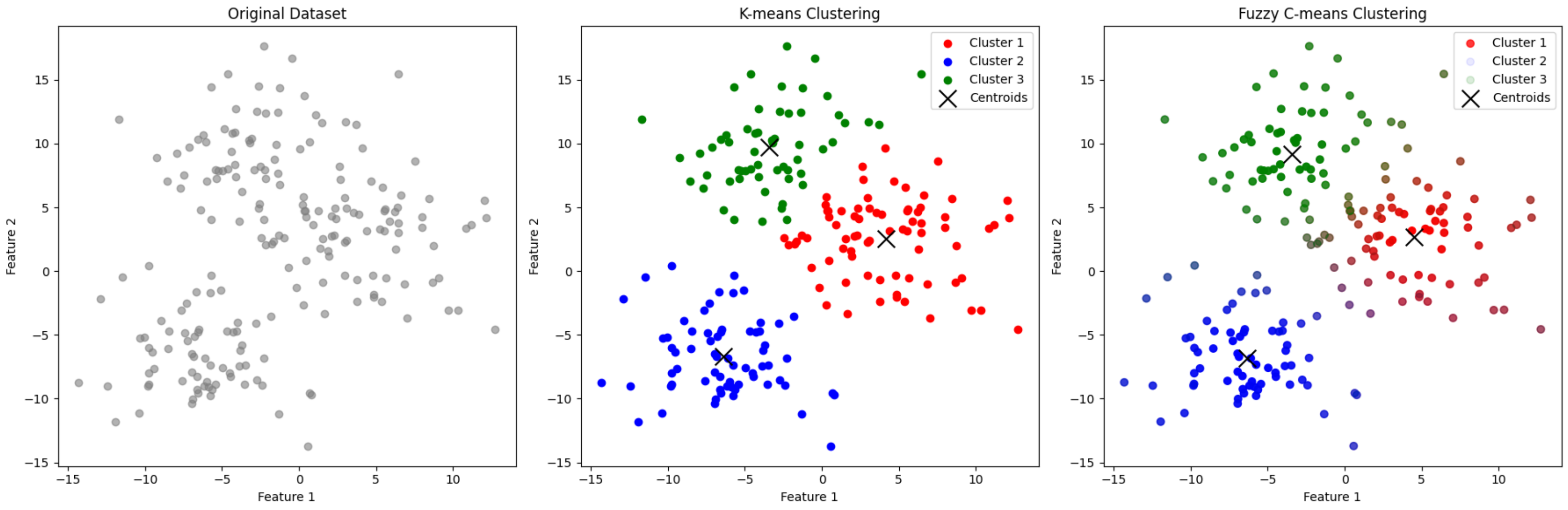
Fuzzy C-Means – parameter tuning



Fuzzy C-Means – pros and cons

- Computationally more expensive
- Choosing the right number of clusters is hard
- Difficulty in handling outliers

- Possible overlapping clusters



Possibilistic Fuzzy C-Means – additionalities

- We introduce a new parameter $t_{i,j}$, called **typicality**, that measure how much each elements fits each cluster, without consider the others (it is not require that them sum to 1).
- **a** and **b** are additional hyperparameters. They define the relative importance of fuzzy membership and typicality.
- An additional hyperparameters defined for each clusters: δ_j

Possibilistic Fuzzy C-Means – steps

- Compute centroids:

$$c_j = \frac{\sum_{i=1}^{N}(au_{i,j}^m + bt_{i,j}^\eta)x_i}{\sum_{i=1}^{N}(au_{i,j}^m + bt_{i,j}^\eta)}$$

- Compute membership values based on new clusters:

$$u_{i,j} = \left(\sum_{k=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}\right)^{-1}$$

- Compute typicality values based on new centroids:

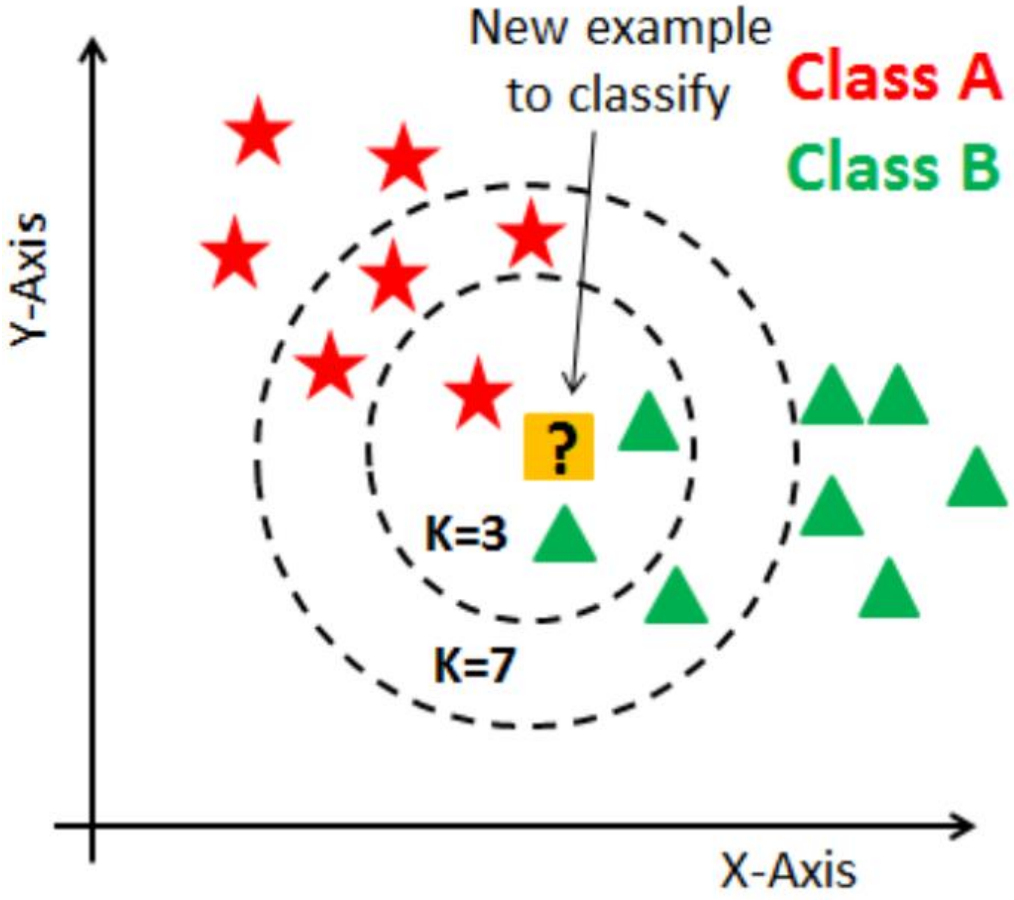
$$t_{i,j} = \left(1 + \left(\frac{b\|x_i - c_j\|^2}{\delta_j}\right)^{\frac{1}{\eta-1}}\right)^{-1}$$

Possibilistic Fuzzy C-Means – Pros and Cons

- Requires carefully choosing more hyperparameters
- Able to detect more patterns in the data
- Handles outliers better than the fuzzy c-means clustering

kNN clustering

- Red stars and green triangles are classified
- Calculate distances and find k nearest neighbours
- Assign label to the unknown points based on k-closest labels



kNN clustering – How to Compute

Point	X	Y	Class
A	1.0	1.0	Red
B	2.0	1.0	Red
C	4.0	4.0	Red
D	5.0	5.0	Blue
E	4.5	3.5	Blue
New Point	3.0	2.0	?

- $\sqrt{(3.0-1.0)^2+(2.0-1.0)^2} = 2.24$
- $\sqrt{(3.0-2.0)^2+(2.0-1.0)^2} = 1.41$
- $\sqrt{(3.0-4.0)^2+(2.0-4.0)^2} = 2.24$
- $\sqrt{(3.0-5.0)^2+(2.0-5.0)^2} = 3.61$
- $\sqrt{(3.0-4.5)^2+(2.0-3.5)^2} = 2.12$

Fuzzy kNN (K = 3)

Point	X	Y	Class
A	0.8	0.8	A
B	0.8	1.2	A
C	3.8	2.8	B
D	4.2	3.2	B
E	4.5	3.5	A
New Point	3.0	2.0	?

$$\sqrt{(3.0-0.8)^2+(2.0-0.8)^2} = 2.50$$
$$\sqrt{(3.0-0.8)^2+(2.0-1.2)^2} = 1.23$$
$$\sqrt{(3.0-3.8)^2+(2.0-2.8)^2} = 2.16$$
$$\sqrt{(3.0-4.2)^2+(2.0-3.2)^2} = 2.34$$
$$\sqrt{(3.0-4.5)^2+(2.0-3.5)^2} = 1.96$$

Fuzzy kNN (K = 3, m=2)

Point	X	Y	Class	
A	0.8	0.8	A	$\sqrt{(3.0-0.8)^2+(2.0-0.8)^2} = 2.50$
B	0.8	1.2	A	$\sqrt{(3.0-0.8)^2+(2.0-1.2)^2} = 1.23$
C	3.8	2.8	B	$\sqrt{(3.0-3.8)^2+(2.0-2.8)^2} = 2.16$
D	4.2	3.2	B	$\sqrt{(3.0-4.2)^2+(2.0-3.2)^2} = 2.34$
E	4.5	3.5	A	$\sqrt{(3.0-4.5)^2+(2.0-3.5)^2} = 1.96$
New Point	3.0	2.0	?	

Fuzzy kNN (K = 3, m=2)

$$\mu_i(P) = \frac{\sum_{j=1}^{k} \mu_{ij} \left(\frac{1}{d(P_i,X_j)^{\frac{2}{m-1}}} \right)}{\sum_{j=1}^{k} \left(\frac{1}{d(P_i,X_j)^{\frac{2}{m-1}}} \right)}$$

Point	X	Y	Class
A	0.8	0.8	A
B	0.8	1.2	A
C	3.8	2.8	B
D	4.2	3.2	B
E	4.5	3.5	A
New Point	3.0	2.0	?

$\sqrt{(3.0-0.8)^2+(2.0-0.8)^2} = 2.50$

$\sqrt{(3.0-0.8)^2+(2.0-1.2)^2} = 1.23$

$\sqrt{(3.0-3.8)^2+(2.0-2.8)^2} = 2.16$

$\sqrt{(3.0-4.2)^2+(2.0-3.2)^2} = 2.34$

$\sqrt{(3.0-4.5)^2+(2.0-3.5)^2} = 1.96$

Fuzzy kNN (K = 3, m=2)

$$\mu_i(P) = \frac{\sum_{j=1}^{k} \mu_{ij} \left(\frac{1}{d(P_i,X_j)^{\frac{2}{m-1}}}\right)}{\sum_{j=1}^{k} \left(\frac{1}{d(P_i,X_j)^{\frac{2}{m-1}}}\right)}$$

Point	X	Y	Class	
A	0.8	0.8	A	$\sqrt{(3.0-0.8)^2+(2.0-0.8)^2} = 2.50$
B	0.8	1.2	A	$\sqrt{(3.0-0.8)^2+(2.0-1.2)^2} = 1.23$
C	3.8	2.8	B	$\sqrt{(3.0-3.8)^2+(2.0-2.8)^2} = 2.16$
D	4.2	3.2	B	$\sqrt{(3.0-4.2)^2+(2.0-3.2)^2} = 2.34$
E	4.5	3.5	A	$\sqrt{(3.0-4.5)^2+(2.0-3.5)^2} = 1.96$
New Point	3.0	2.0	?	

$$\mu_A(P) = \frac{\frac{1}{1.23^2} + \frac{0}{2.16^2} + \frac{1}{1.96^2}}{\frac{1}{1.23^2} + \frac{1}{2.16^2} + \frac{1}{1.96^2}} \approx 0.811$$

$$\mu_B(P) = \frac{\frac{0}{1.23^2} + \frac{1}{2.16^2} + \frac{0}{1.96^2}}{\frac{1}{1.23^2} + \frac{1}{2.16^2} + \frac{1}{1.96^2}} \approx 0.189$$

Fuzzy kNN (K = 3, m=2)

$$\mu_i(P) = \frac{\sum_{j=1}^{k} \mu_{ij} \left(\frac{1}{d(P_i, X_j)^{\frac{2}{m-1}}} \right)}{\sum_{j=1}^{k} \left(\frac{1}{d(P_i, X_j)^{\frac{2}{m-1}}} \right)}$$

Point	X	Y	Class	
A	0.8	0.8	A	$\sqrt{(3.0-0.8)^2+(2.0-0.8)^2} = 2.50$
B	0.8	1.2	A	$\sqrt{(3.0-0.8)^2+(2.0-1.2)^2} = 1.23$
C	3.8	2.8	B	$\sqrt{(3.0-3.8)^2+(2.0-2.8)^2} = 2.16$
D	4.2	3.2	B	$\sqrt{(3.0-4.2)^2+(2.0-3.2)^2} = 2.34$
E	4.5	3.5	A	$\sqrt{(3.0-4.5)^2+(2.0-3.5)^2} = 1.96$
New Point	3.0	2.0	?	

$$\mu_A(P) = \frac{\frac{1}{1.23^2} + \frac{0}{2.16^2} + \frac{1}{1.96^2}}{\frac{1}{1.23^2} + \frac{1}{2.16^2} + \frac{1}{1.96^2}} \approx 0.811$$

$$\mu_B(P) = \frac{\frac{0}{1.23^2} + \frac{1}{2.16^2} + \frac{0}{1.96^2}}{\frac{1}{1.23^2} + \frac{1}{2.16^2} + \frac{1}{1.96^2}} \approx 0.189$$

Fuzzy Scaling Process – Main Idea

- The most basic algorithm for fuzzy clustering:
- For each pair $x_i,\ x_j\ determine\ their\ similarity\ s_{i,j}$
- Define a matrix $S\ =\ \left(s_{i,j}\ \right)_{n\times n}$
- Calculate the transitive closure $R = tc(S)$
- $Calculate$ the $\alpha - cut\ R_\alpha$
- If i—th and j—th rows of R_α are equal x_i and x_j are in the same cluster

Fuzzy Scaling Process – Example

- Let's take 3 colors sky blue, light blue and green

Fuzzy Scaling Process – Example

- Let's take 3 colors sky blue, light blue and green
- Let's pick some distance metric which will return the distance between two colors

Fuzzy Scaling Process – Example

- Let's take 3 colors sky blue, light blue and green
- Let's pick some distance metric which will return the distance between two colors
- Let's fill the matrix with the distances between colors

$$S = \begin{bmatrix} 1 & 0.9 & 0.2 \\ 0.9 & 1 & 0.3 \\ 0.2 & 0.3 & 1 \end{bmatrix}$$

Fuzzy Scaling Process – Pros and Cons

- We have to obtain R

$$R = \begin{bmatrix} 1.0 & 0.9 & 0.3 \\ 0.9 & 1.0 & 0.3 \\ 0.3 & 0.3 & 1.0 \end{bmatrix}$$

Fuzzy Scaling Process – Pros and Cons

- We have to obtain R

$$R = \begin{bmatrix} 1.0 & 0.9 & 0.3 \\ 0.9 & 1.0 & 0.3 \\ 0.3 & 0.3 & 1.0 \end{bmatrix}$$

- We have to perform alpha-cut (let's take 0.5)

Fuzzy Scaling Process – Pros and Cons

- We have to obtain R

$$R = \begin{bmatrix} 1.0 & 0.9 & 0.3 \\ 0.9 & 1.0 & 0.3 \\ 0.3 & 0.3 & 1.0 \end{bmatrix}$$

- We have to perform alpha-cut (let's take 0.5)

$$R_\alpha = \begin{bmatrix} 1.0 & 0.9 & 0 \\ 0.9 & 1.0 & 0 \\ 0 & 0 & 1.0 \end{bmatrix}$$

THANK YOU FOR YOUR ATTENTION!