Gath - Geva Clustering

First thing first, let's use the generator tools from previous works

Classical fuzzy clustering: FCM

the algorithm implementation

Fuzzy C-means (FCM) is a clustering algorithm that is used to partition data points $\{x_i\}_1^n$ into a predefined number of clusters c. The algorithm is a 'soft' clustering method, which allows each data point to belong to multiple clusters to a certain degree.

The FCM algorithm can be summarized as follows:

1. Initialize the cluster centers.
2. Calculate the membership degrees of each data point to each cluster.

$$p(w_j|x_i) \leftarrow \frac{(\frac{1}{d_{ij}^2})^{1/b-1}}{\sum_{i=1}^{N}(\frac{1}{d_{ij}^2})^{1/b-1}}$$

3. Update the cluster centers based on the membership degrees.

$$\mu_{cx} \leftarrow \frac{\sum_{i=1}^{N}\hat{p}\left(w_i|x_j\right)^b \cdot x_j}{\sum_{i=1}^{N}\hat{p}\left(w_i|x_j\right)^b}$$

Repeat steps 2-3 until the cluster centers converge.

where $d_{ij} = d(x_i, c_j)$ is the distance between the data point x and the cluster center c_i, b is a fuzzifier parameter, and C is the number of clusters.

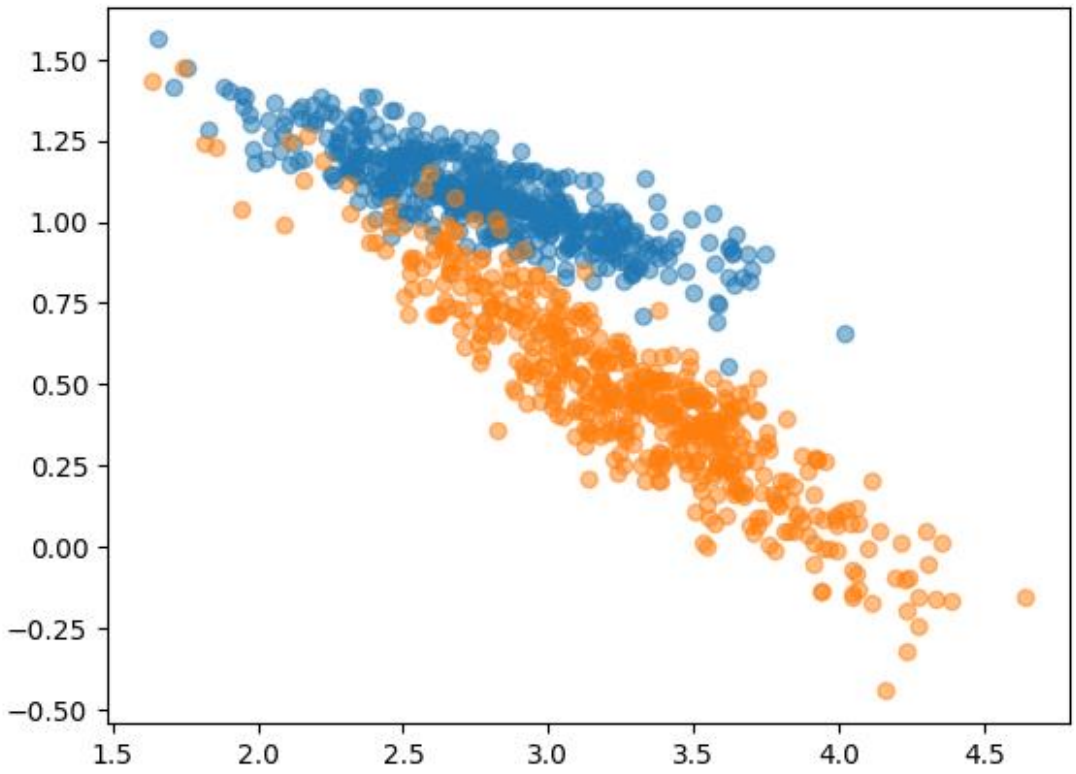
The FCM algorithm is a versatile clustering algorithm that can be used to cluster data of different types. It is also relatively robust to noise and outliers.

generate the data

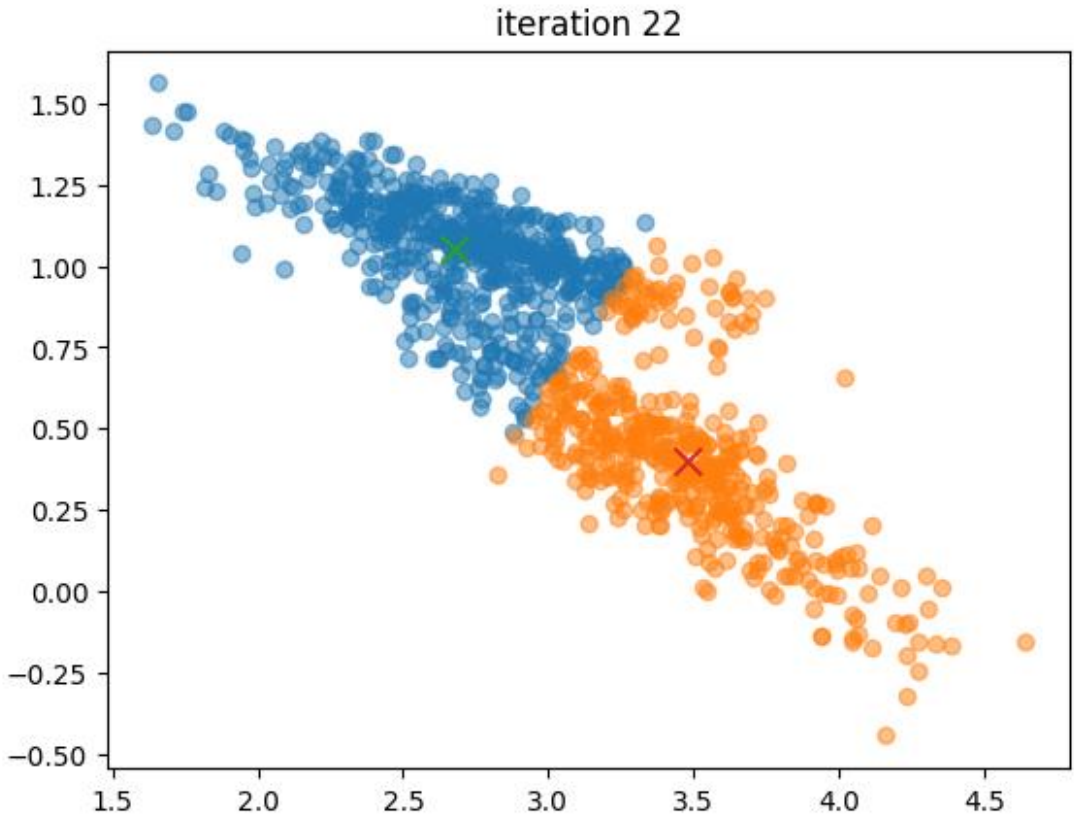
```
K = 2
d = 2
N = 1000
```

plot the data

```
plot_clusters(clusters)
```


1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
1.5
2.0
2.5
3.0
3.5
4.0
4.5

run the algorithm



Gath-Geva algorithm

The WUOFC algorithm:

The WUOFC algorithm also known as the GG clustering algorithm is a variant of the FCM algorithm that is used to cluster data points $\{x_i\}_1^n$ when the initialization centers $\{c_j\}_1^n$ are not known, but also when the number of clusters K is not known. the algorithm, prorposed in the Gath-Geva algorithm paprt, is a six-step algorithm:

1. Initialize the cluster centers. Initialize $K \leftarrow 1$, choose a single initial center c_1 at the center of mass of the data,
$$c_1 = \frac{1}{N}\sum_{i=1}^{N} x_i$$
2. while $K < K_{max}$ do:
   - Calculate the new partition:
     - Call FCM with K clusters and the current cluster centers $c_1, \dots, c_K$. The output of the FCM algorithm is the membership matrix $p(w_j|x_i)$ and the cluster centers μ_{cx}.
     - use the centroids μ_{cx} as the new cluster centers $c_1, \dots, c_K$. as the initial cluster centers for the K-Means with exponential kernel algorithm.
   - calculate the cluster validity criteria V for the partition.
   - add a new centroid, c_{K+1} equally distant from all the other centroids, to the set of cluster centers. set $K \leftarrow K + 1$.
3. return the partition with the highest validity criteria V.

K-Means with exponential kernel

The K-Means with exponential kernel algorithm is a variant of the K-Means algorithm that is used to cluster data points $\{x_i\}_1^n$ when the initialization centers $\{c_j\}_1^n$ are not known, but with known number of clusters K .

Validity criteria

Validity criteria play a crucial role in clustering algorithms by providing a quantitative measure of the quality and appropriateness of the clusters generated by the algorithm. These criteria help assess how well the algorithm has partitioned the data into meaningful and distinct groups. Validity criteria serve as objective measures to guide the selection of the best clustering solution among multiple alternatives. Here's how validity criteria contribute to the algorithm:

The trace criterion

the trace criterion is a cluster validity criteria that is used to evaluate the quality of a clustering partition by estimating the sum of variances of the distribution by summing the diagonal elements. the trace criterion is defined as:

$$V = trace(S_W) = \sum_{k=1}^{K} t\,race(S_i) = \sum_{k=1}^{K} \frac{1}{N_k} \sum_{i=1}^{N} p\,(w_k|x_i) \cdot ||x_i - c_k||^2$$

where N_k is the number of data points in cluster k.

The Fuzzy Hyper-Volume criterion

the determinant criterion is a cluster validity criteria that is used to evaluate the quality of a clustering partition by estimating the volume of the distribution. the determinant criterion is defined as:

$$V = det\left(\frac{1}{N_k} \sum_{i=1}^{N}{}_i\,|(x_i - c_k) \cdot (x_i - c_k)^t\right)$$

the fuzzy hypervolume is defined as:

$$V_{hv} = \sum_{k=1}^{K} h_k$$

$$= \sum_{k=1}^{K} \sqrt{det(F_k)}$$

F_k is the fuzzy covariance matrix of cluster k.

$$F_k = \frac{\sum_{i=1}^{N} w_i \cdot p(w_k|x_i) \cdot ||x_i - c_k||^2}{\sum_{i=1}^{N} w_i \cdot p(w_k|x_i)}$$

the Partition Density criterion

the partition density criterion is a validity criteria that evaluates the quality of a partition by estimating the density of the distribution. the partition density criterion is defined as:

$$V_{pd} = \frac{\sum_{k=1}^{K} b_k}{\sum_{k=1}^{K} h_k}$$

where

$$b_k = \sum_{i \in I_k} w_i \cdot p(w_k|x_j) \cdot w_j$$

and could be thought of as the weighted sum of the datapoints that have strong membership to the partition's clusters, weighted by how strong their membership is, all of that divided by the hyper-volume of the partition

the Average Partition Density criterion

The average density criteria is a validity criterion for clustering algorithms that measures the average density of the clusters. A high average density indicates that the clusters are well-separated and compact. the average partition density criterion is defined as:

$$V_{ad} = \frac{1}{k} \sum_{k=1}^{K} \frac{b_k}{h_k}$$

where

$$c_k = \sum_{i \in I_k} w_i \cdot p(w_k|x_j)$$

and could be thought of as the weighted sum of the datapoints that have strong membership to the partition's clusters, weighted by how strong their membership is, all of that divided by the hyper-volume of the partition where N_k is the number of data points in cluster k. The average density criteria is a simple and intuitive validity criterion. However, it is not without its drawbacks. One drawback is that it can be sensitive to the choice of distance metric. Another drawback is that it can be computationally expensive for large data sets.

the Maximum Average Partition Density criterion

the maximum average partition density criterion is a validity criteria that evaluates the quality of a partition by estimating the average density of the clusters in this partition. the maximum average partition density criterion is defined as:

$$V_{ad} = \frac{1}{k} \sum_{k=1}^{K} \frac{m_k}{h_k}$$

where

$$ m_k = \sum_{i \in J_k} w_i \cdot p(w_k|x_j) }$$

note that J_k is different then I_k from the previous criteria. J_k the set of data points that satisfy

$$J_K = max_{j \in 1 \ldots K} \left(p\left(w_k | x_j\right) \right)$$

meaning that J_k is the set of data points that have the strongest membership to cluster k.

the normalized partition coefficient criterion

the normalized partition coefficient criterion is a validity criteria that evaluates the quality of a partition in a very straight-forward way: by evaluating the loss function of the partition algorithm itself: the exponential distance function value for the distances of datapoints in the partition and using it to calculate a weighted sum. we normalize the loss value J_q by the k parameter itself, it is obvious that a larger number of cluster will provide us with a smaller loss, so with the normalization $K \cdot J_q^{(k)}$ we create a drawback to the selection of a larger k . the normalized partition coefficient criterion is defined as:

$$V_{npc} = K \cdot J_q^{(k)} = K \cdot \sum_{k=1}^{K} \sum_{i=1}^{N} w_i \cdot p\left(w_k | x_j\right)^q \cdot d^2(c_k, X_i)$$

the Invariant criterion

the invariant criterion is defined as:

$$V_I = tr(S_W^{-1} S_B) = \sum_{i=1}^{d} \lambda_i$$

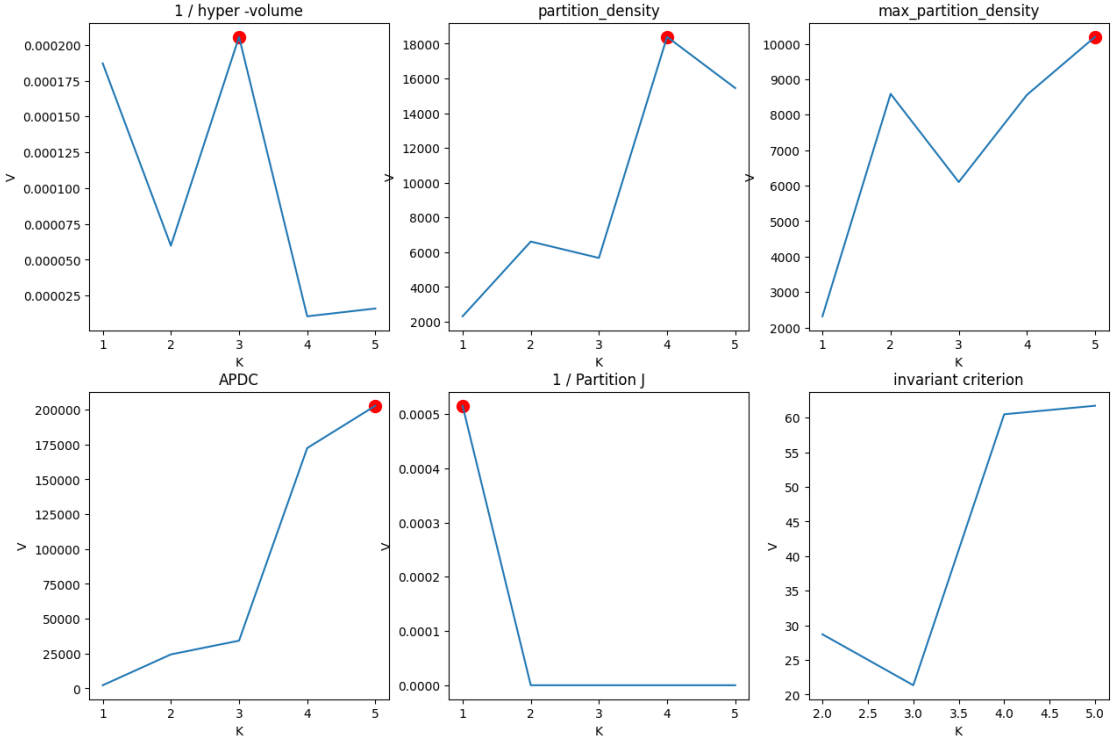
where $ d^2(c_k,X_i) $ is the exponential distance function

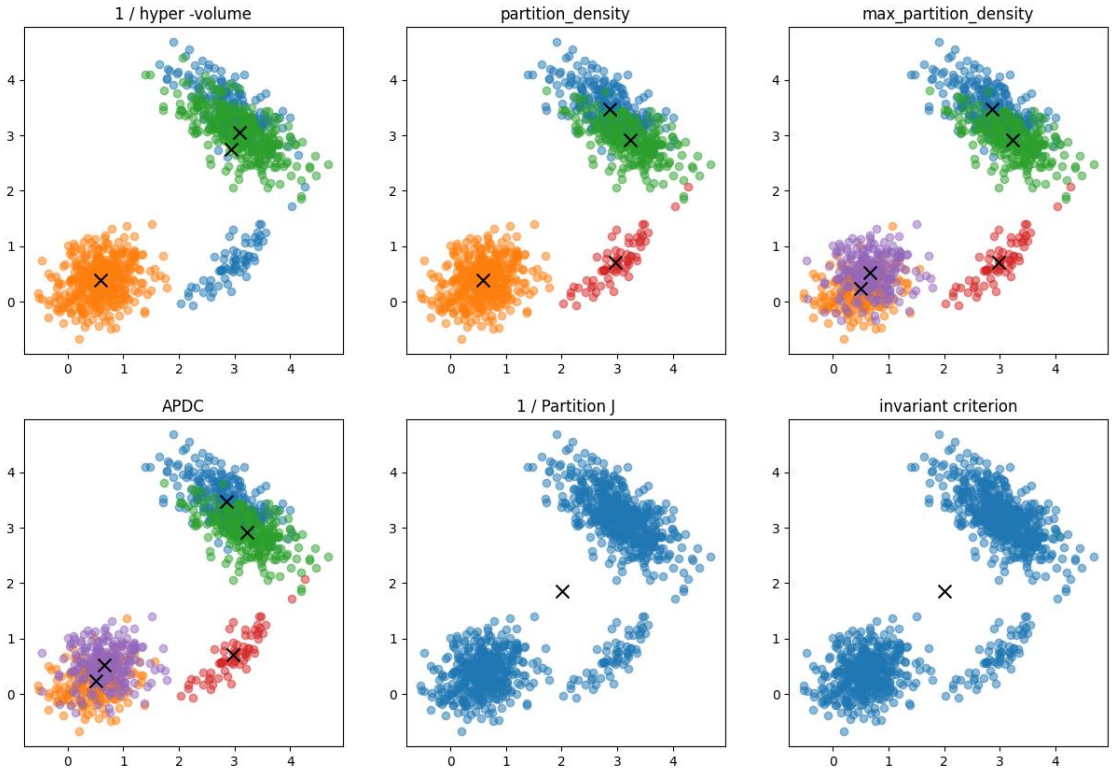
the invariant criterion takes into account the fact that clustering algorithms are often sensitive to transformations of the data, such as translations, rotations, and scalings. The invariant criterion measures how well the clustering algorithm is able to identify the same clusters after the data has been transformed.

The invariant criterion is a useful validity criterion for clustering algorithms that are used to analyze data that is likely to be transformed. For example, the invariant criterion can be used to evaluate the quality of clusterings that are created from images or time series data.

compare the different criteria

for each dataset we will conduct an experiment in which we will compare the different criteria and determine which one is the fittest to the characteristics of the data, in the following format:



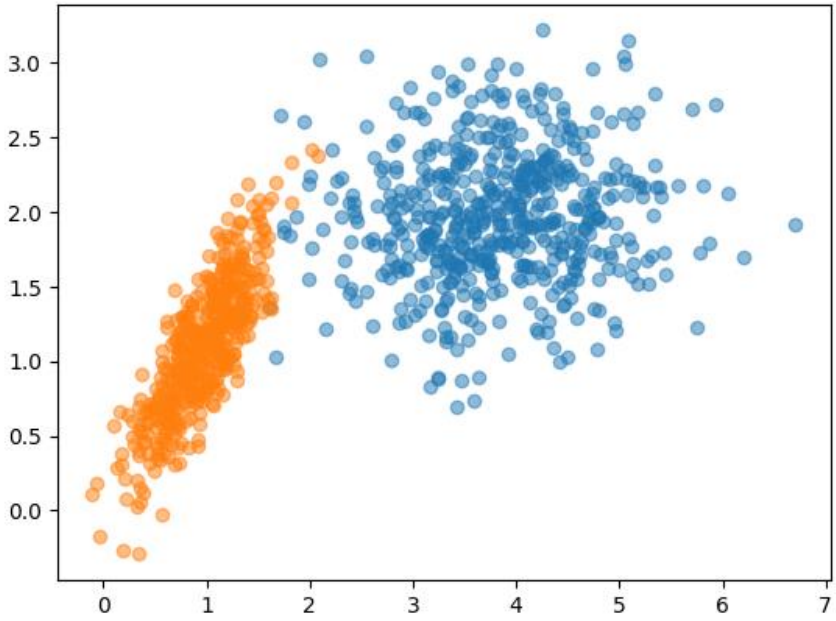


experiment : simple gaussian mixture data

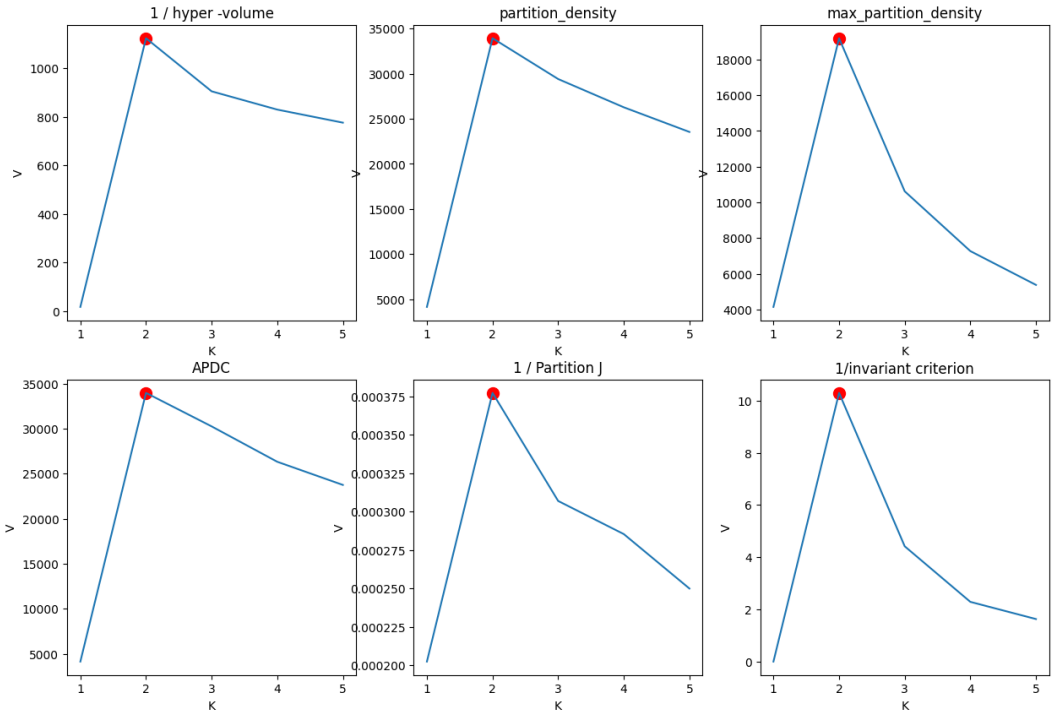
we will generate a simple, easy to cluster dataset with 2 clusters, one with diagonal covariance matrices, the other with a full covariance matrix, the two clusters are slightly tuching each other, we will examin if all the criteria manage to correctly cope with these basic challenges.

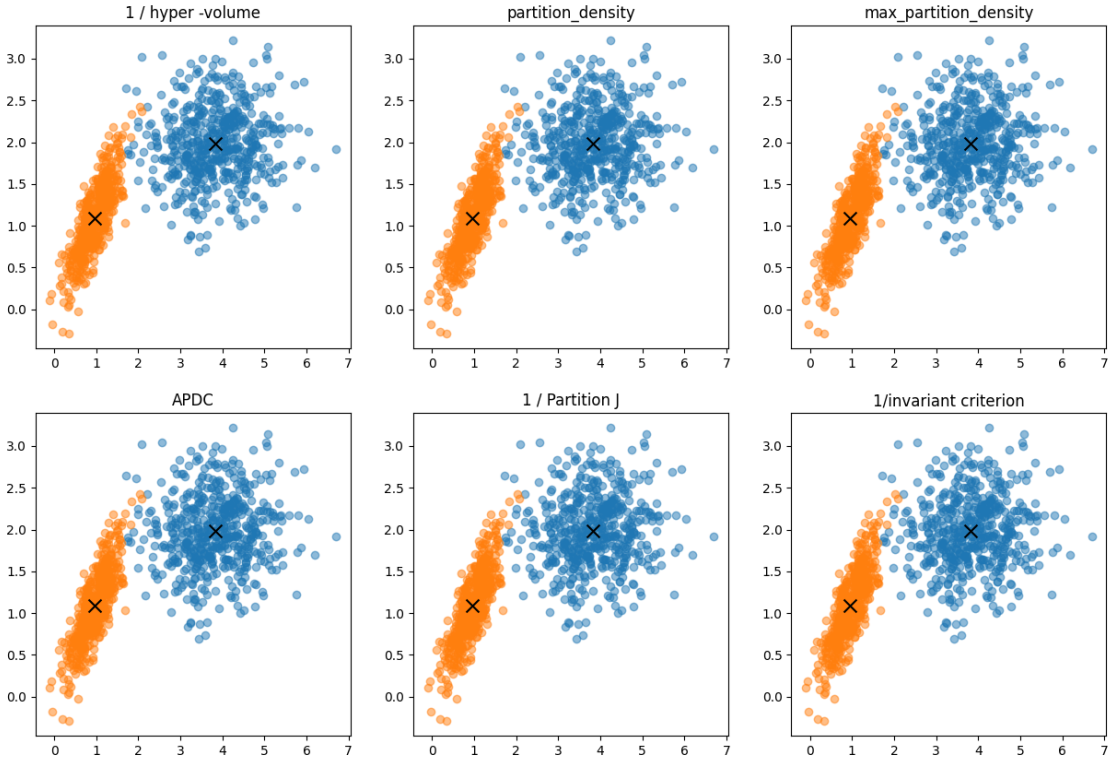
generate data

```
K = 2
d = 2
N = 1000
```



run WUOFC





results

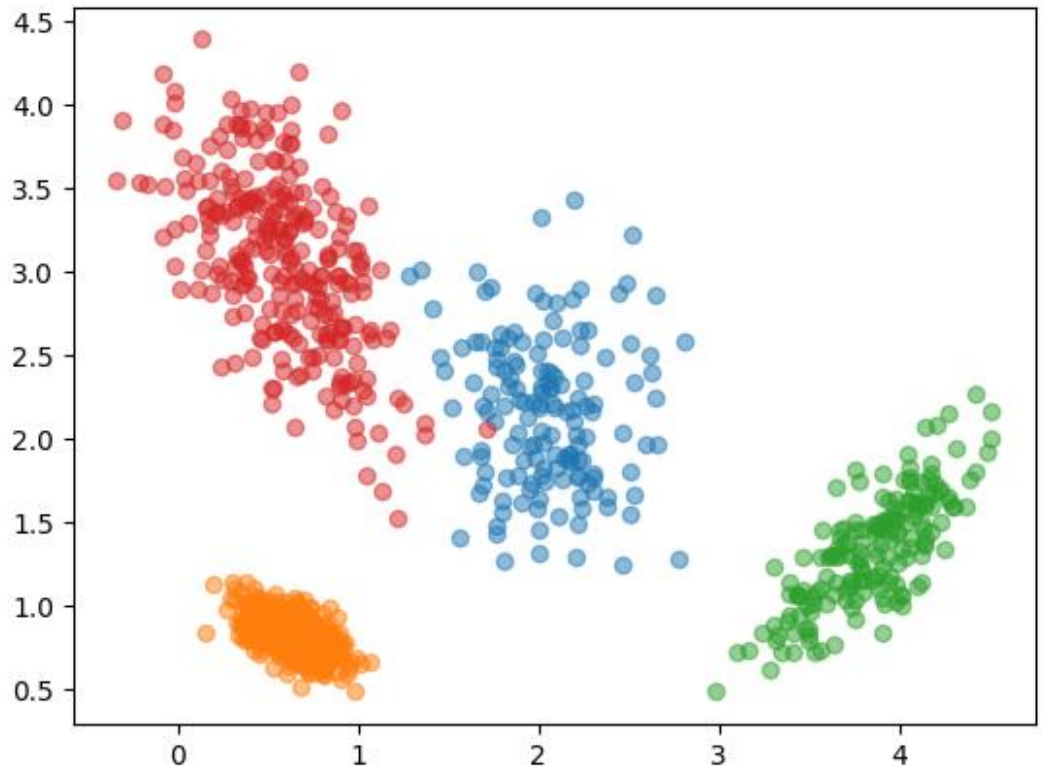
as we can see, all the criteria are able to find the correct number of clusters, and the correct cluster partition. this won't be the case for more complex datasets, but it is reassuring to see that the criteria are able to find the correct number of clusters in this simple case. we

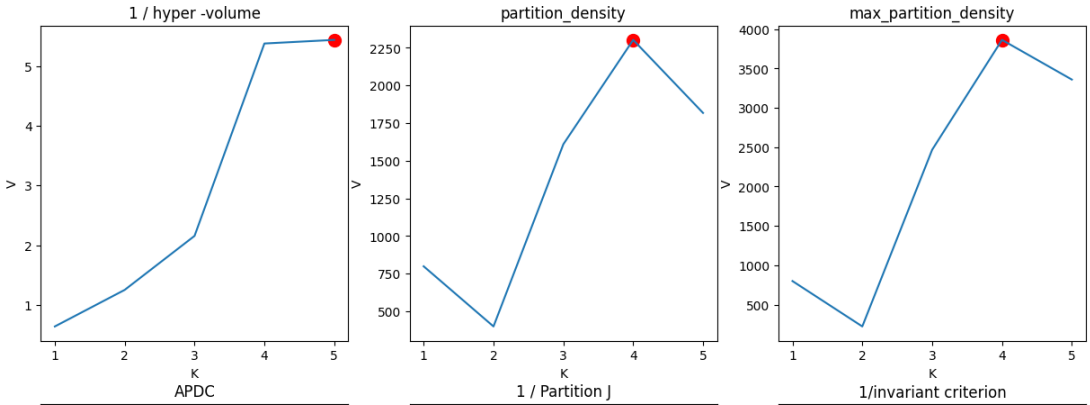
can see that the max partition density criterion and the invariant criterion were the most decisive. maybe it is because further cluster seperation would decreace S_B and increase S_W resulting in a lower validity value.

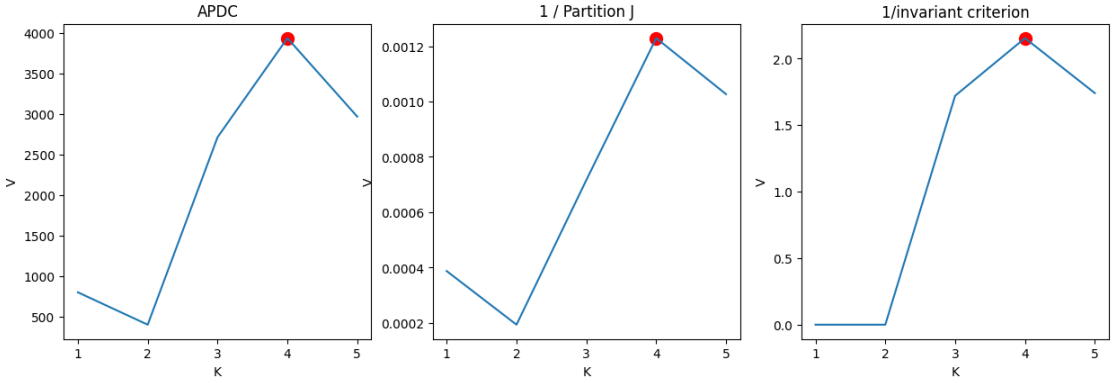
experiment part 2: 4 clusters, full covariance matrix gaussian mixture data

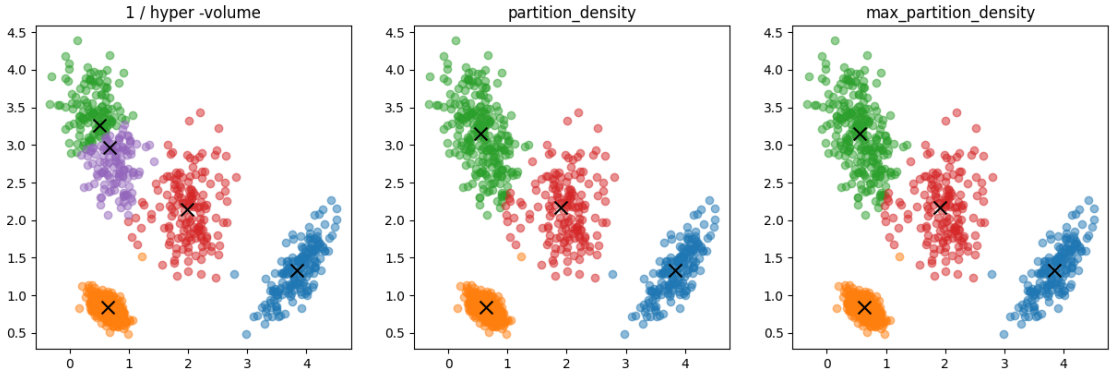
now let us generate a dataset with 4 clusters, and full covariance matrices,more complex than the previous one. again two of the clusters are tuching each other, this time, the other two are far away from each other. We will see that the fuzzy hyper volume criterion is not able to find the correct number of clusters, while the other criteria are.

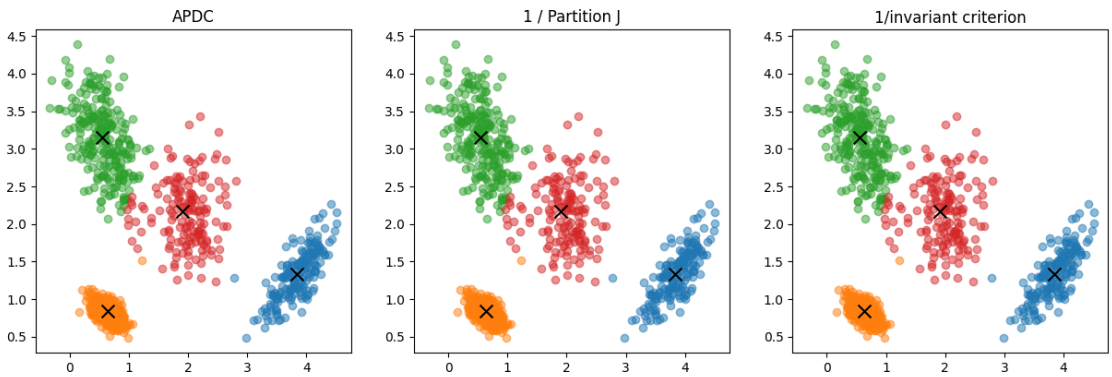
```
K = 4
d = 2
N = 1000
distance = 4
max_iters = 50
```




1 / hyper -volume
partition_density
max_partition_density
V
K


APDC
1 / Partition J
1/invariant criterion
V
K


1 / hyper -volume
partition_density
max_partition_density

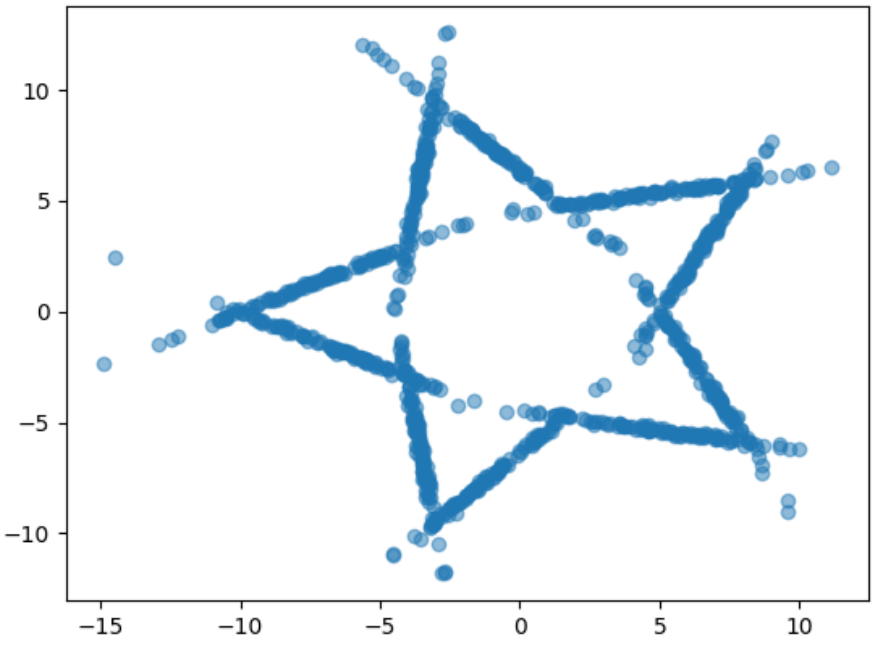

APDC
1 / Partition J
1/invariant criterion

results

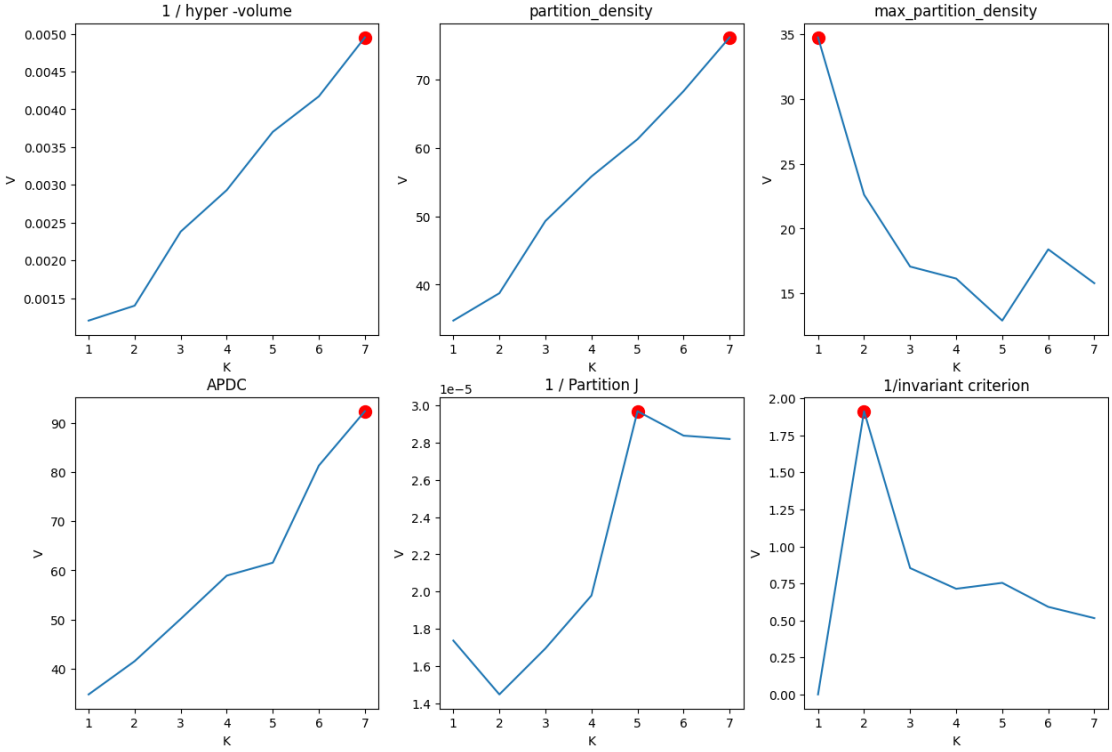
as we can see, most of the criteria are able to find the correct number of clusters, except for the fuzzy hyper volume criterion. the algorithm managed to "shrink" the total hypervolume of the cluster marked in green by splitting it into two, thus getting slightly better score in the fuzzy hyper volume criterion. however, the difference between the two scores for $k = 4$ and $k = 5$ is not significant enough to mark thi as a faliure of the criterion.

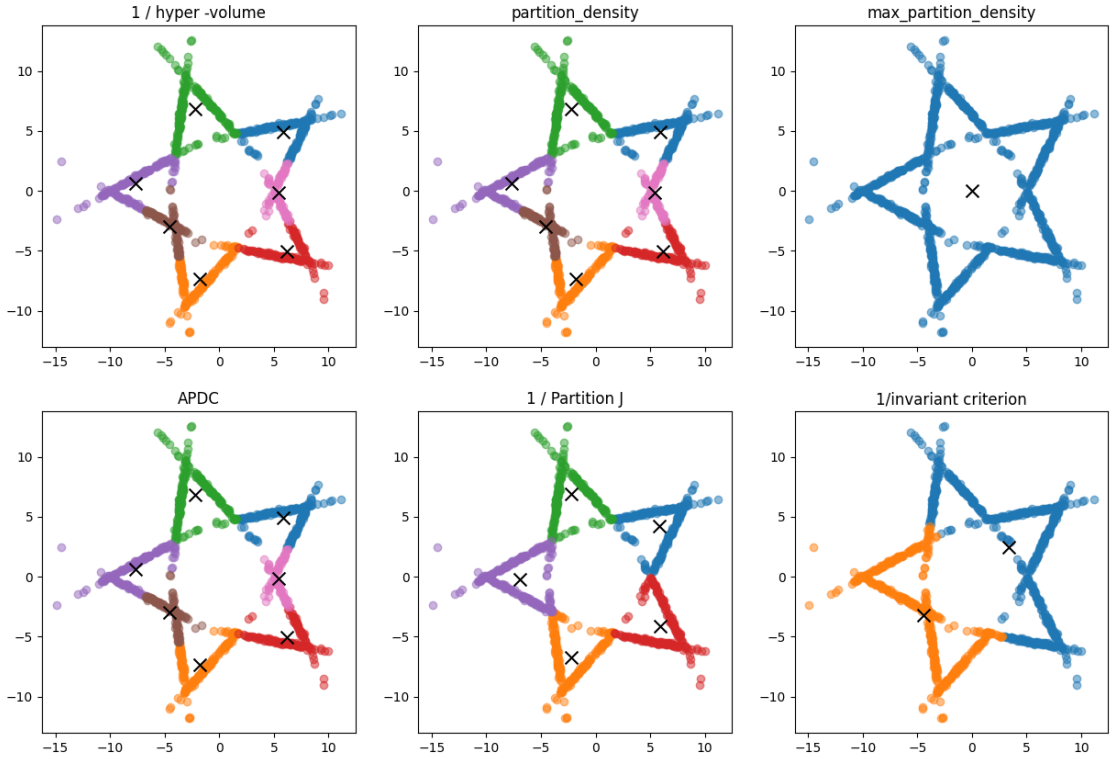
2nd experiment : star shaped data

we will generate a star shaped dataset, and see if the criteria are able to find the correct number of clusters. we will also see if the algorithm is able to find the correct cluster partition. will any of the criteria be able to detect the star 'arms' as seperate clusters like we saw in the Agglomerative clustering algorithm?



run WUOFC


1 / hyper -volume
partition_density
max_partition_density
APDC
1e−5
1 / Partition J
1/invariant criterion
V
K

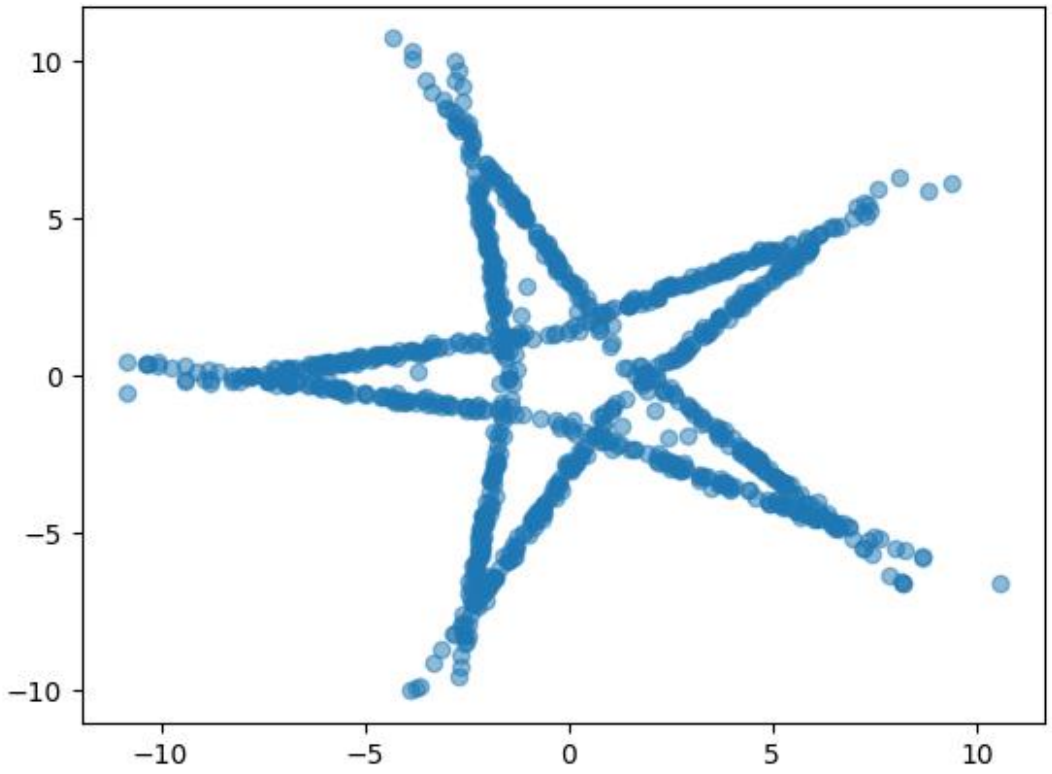

1 / hyper -volume
partition_density
max_partition_density
APDC
1 / Partition J
1/invariant criterion

results

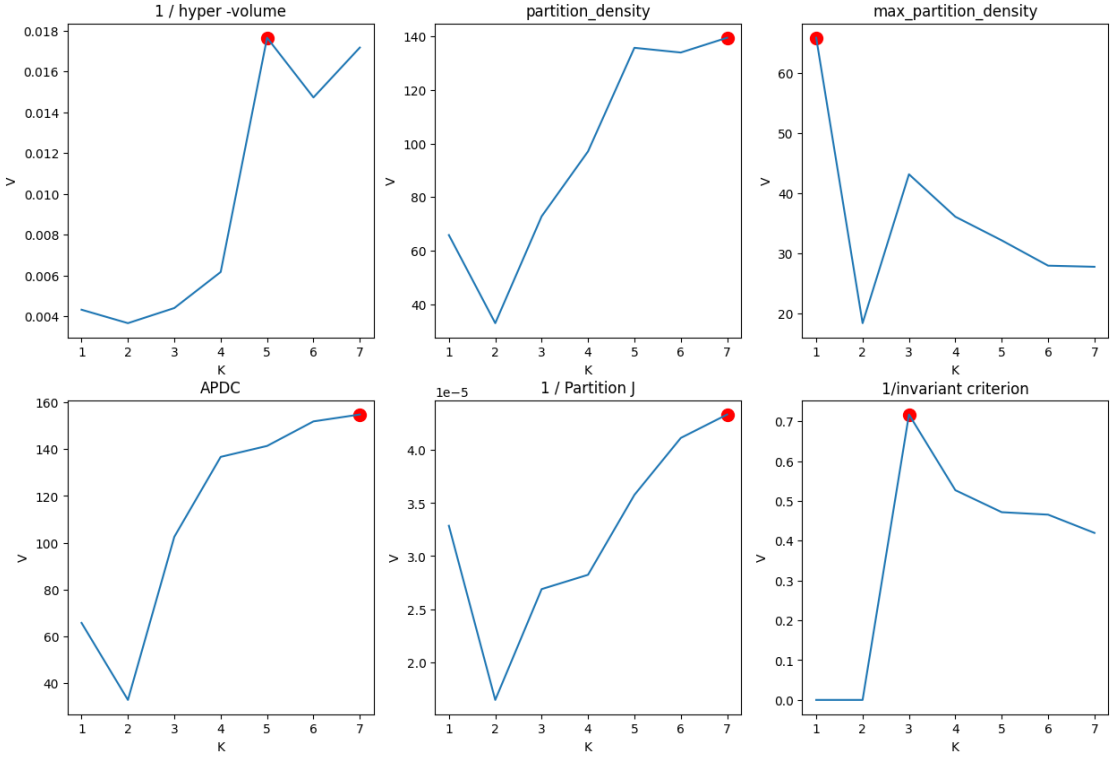
it seems that the partition J criterion actually separated the star to its arms, a visually logical thing to expect, yet a remarkable result. the other criteria returned other values of k, as we would expect, for example, the fuzzy hyper volume criterion, the partition density criterion and the average partition density criterion returned $k = 7$, as it maximizes the density of clusters that converge mostly to the straight lines of the shape.

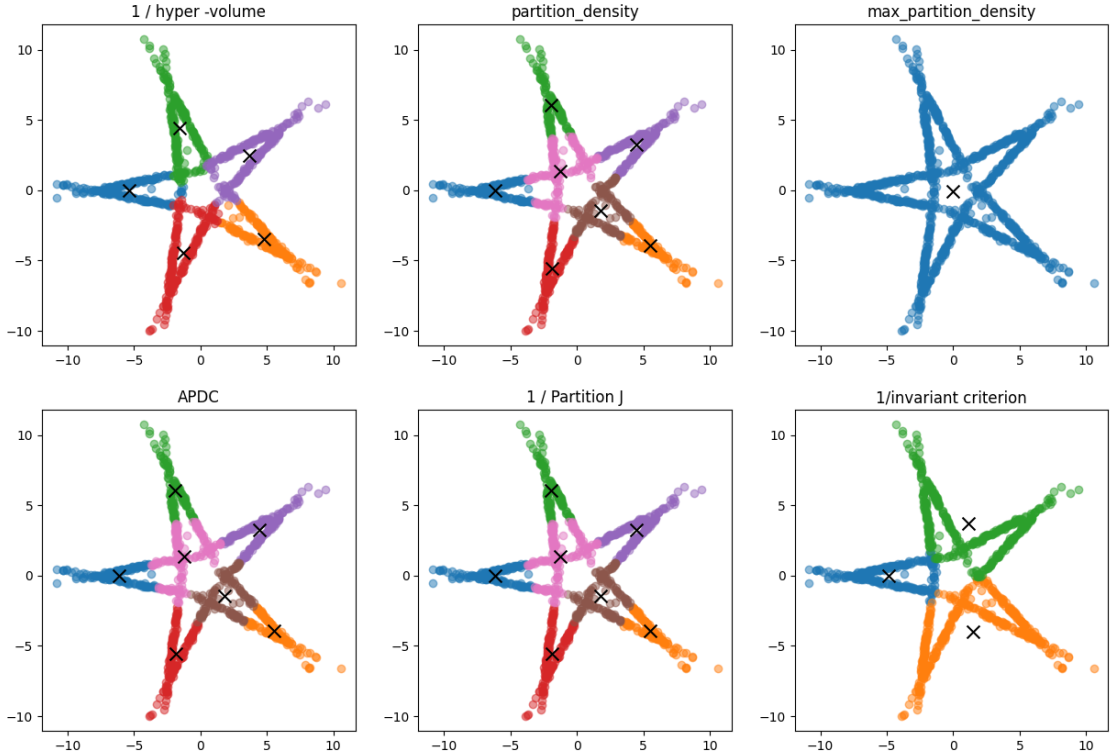
3nd experiment : second attempt to cluster a star

we will generate a star shaped dataset, this time with very thin hands, this way the datapoints responsible to each arm are relatively close together and see if the criteria are able to identify them as clusters. we will also see if the algorithm is even able to find a reasonable partition k. will any of the criteria be able to detect the star 'arms' other than partition J?



run WUOFC

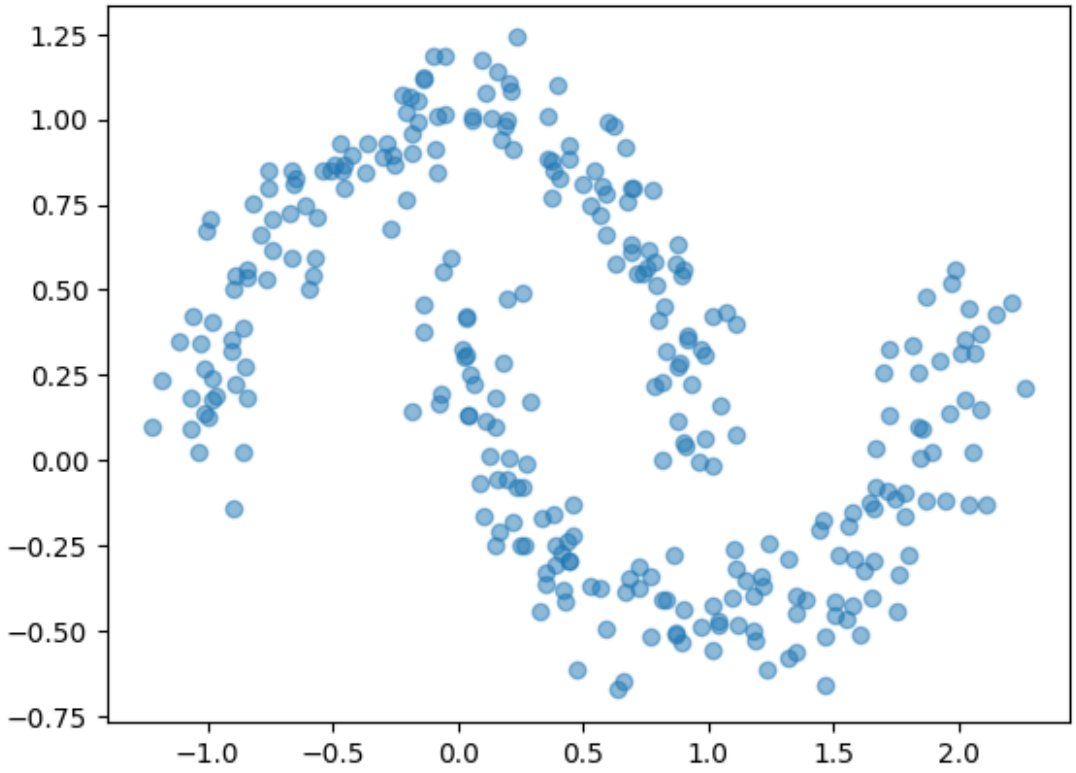

1 / hyper -volume
partition_density
max_partition_density
APDC
1e−5
1 / Partition J
1/invariant criterion
V
K

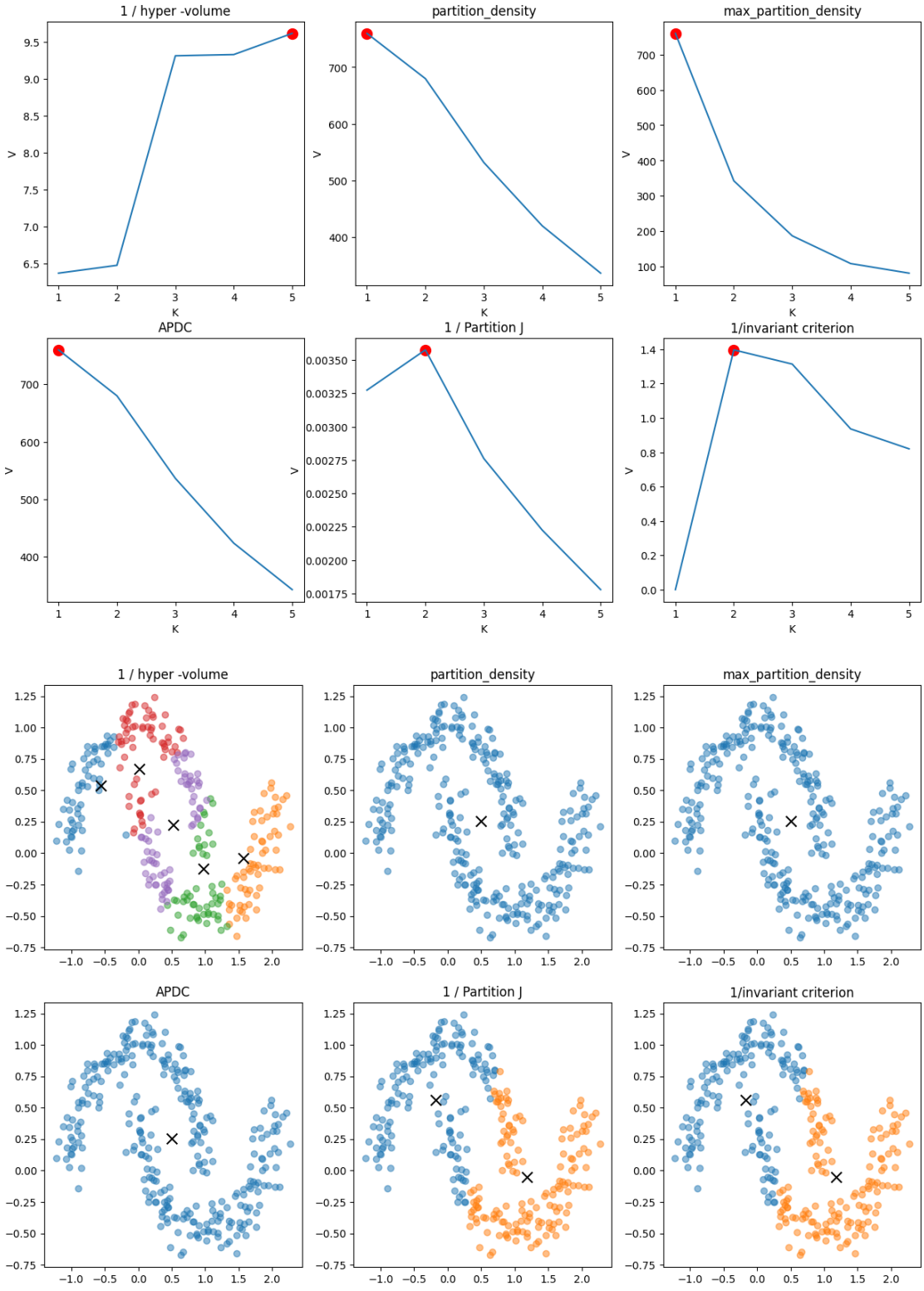

1 / hyper -volume
partition_density
max_partition_density
APDC
1 / Partition J
1/invariant criterion

results

an interesting result emerges - the hyper volume criteria was able to detect and seperate the arms of the star, as cluster while the other criteria were not. this is because the hyper volume criteria is not affected by the density of the clusters, only by their volume. the arms of the star are very thin, and the datapoints responsible for each arm are very close together, so the sum of the volumes of the clusters is very small. the criteria that are affected by the density of the clusters, such as the partition density criterion, the average partition density criterion and the fuzzy hyper volume criterion, returned $k = 7$, as it maximizes the density of clusters when dense areas in the middle of the dataset are clustered separately.

4th experiment : 2 'S' shaped clusters



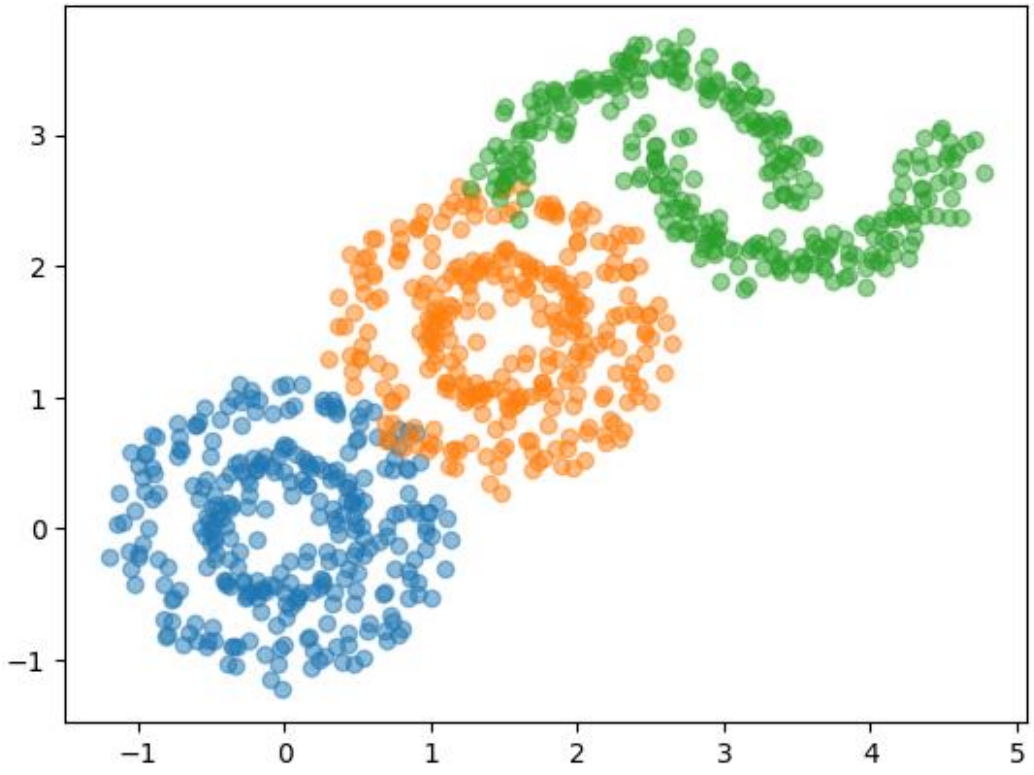


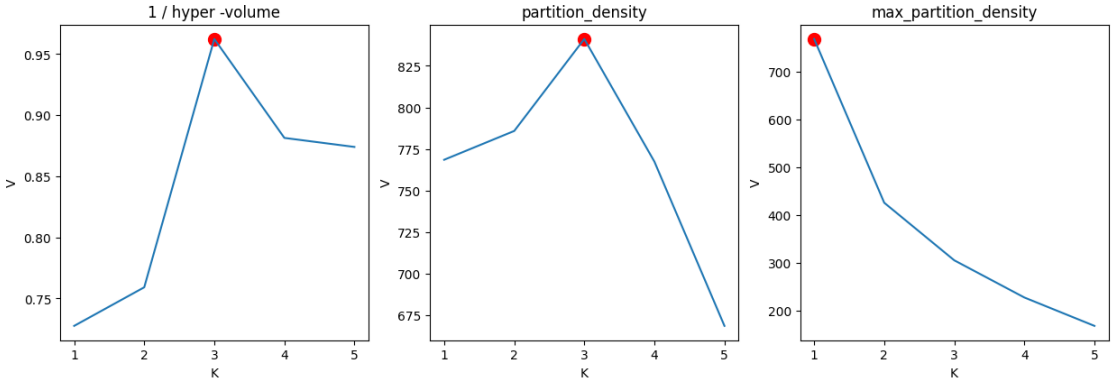
results

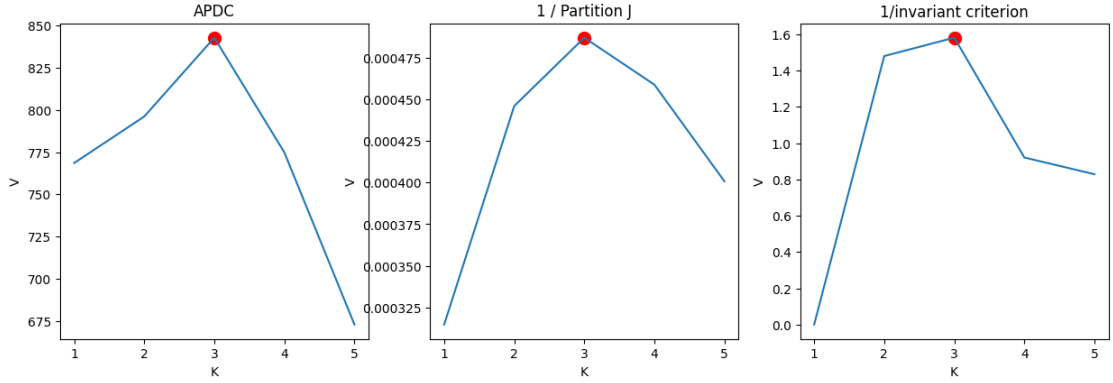
Only the partition J criterion and the invariant criterion were able to find the correct number of clusters. the other criteria were not able to find the correct number of clusters,

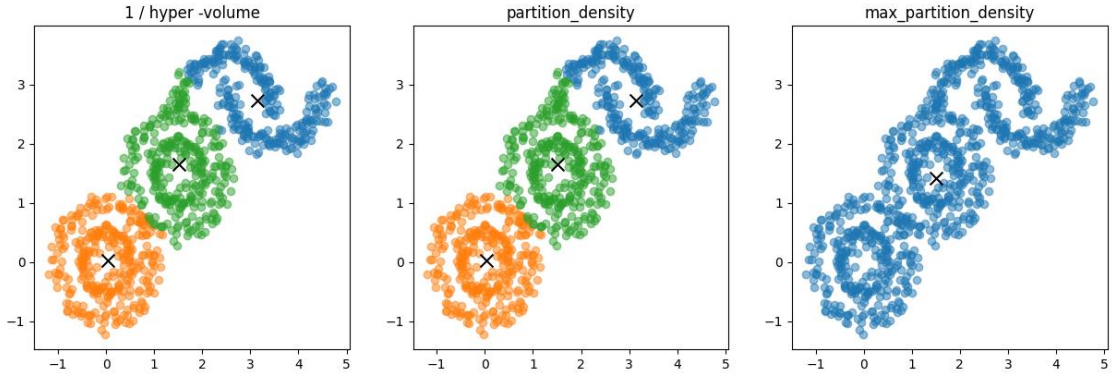
and returned ether $k = 1$ or $k = K_{max}$. the partition itself is not perfect but, knowing the right number of classes and the centers could be a good initializer to a other clustering methods.

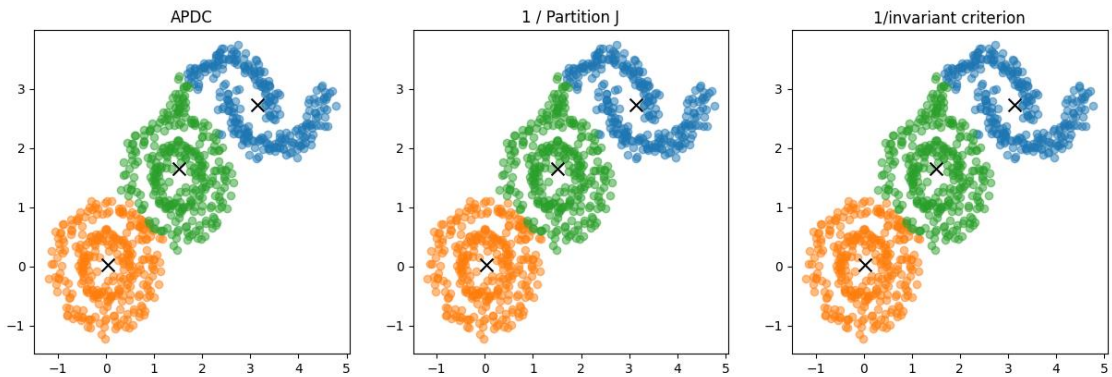
5th experiment: non - uniformly dense circles




1 / hyper -volume
partition_density
max_partition_density
V
K

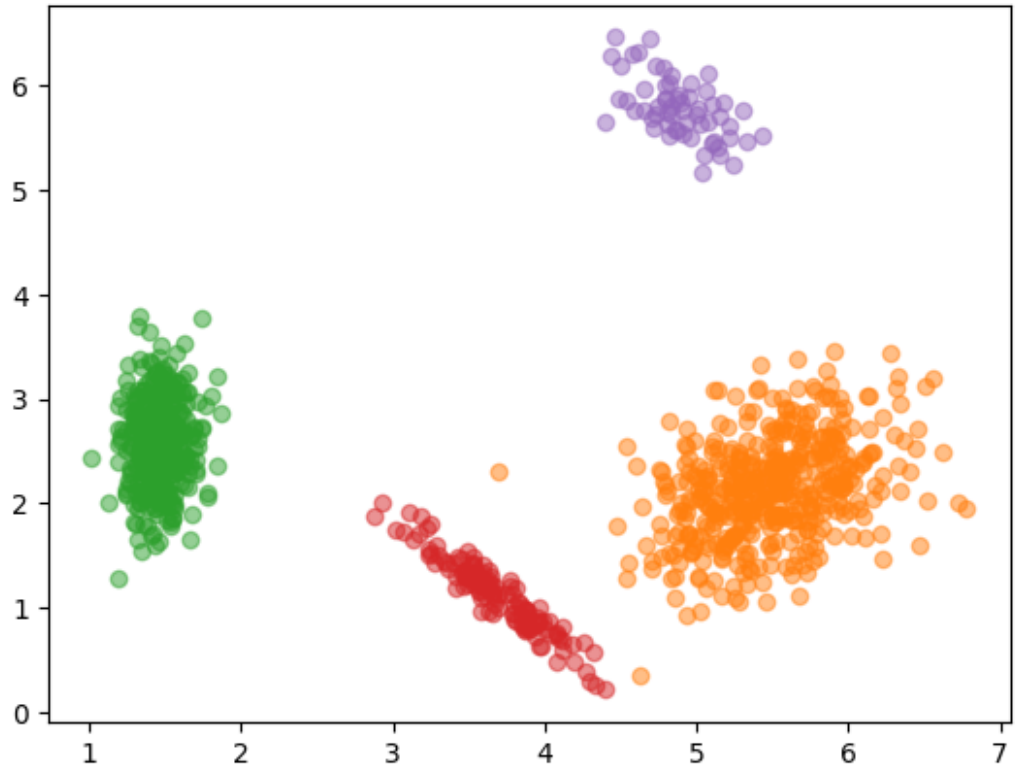

APDC
1 / Partition J
1/invariant criterion
V
K

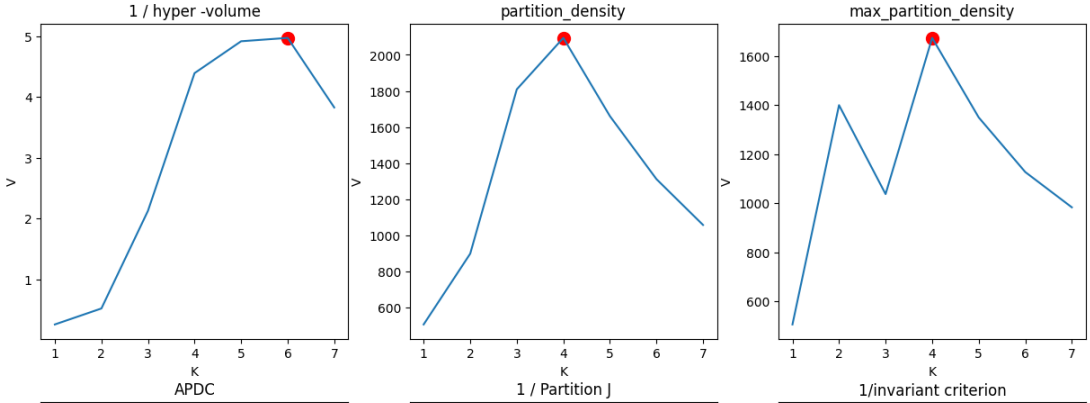

1 / hyper -volume
partition_density
max_partition_density

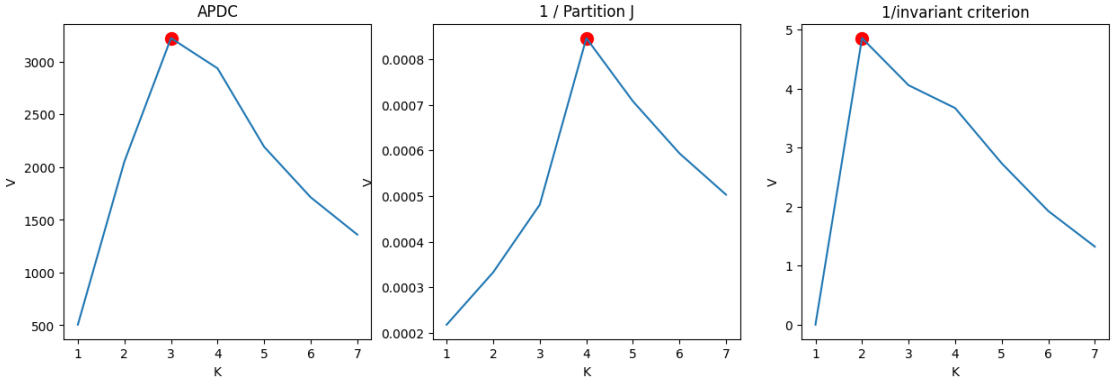

APDC
1 / Partition J
1/invariant criterion

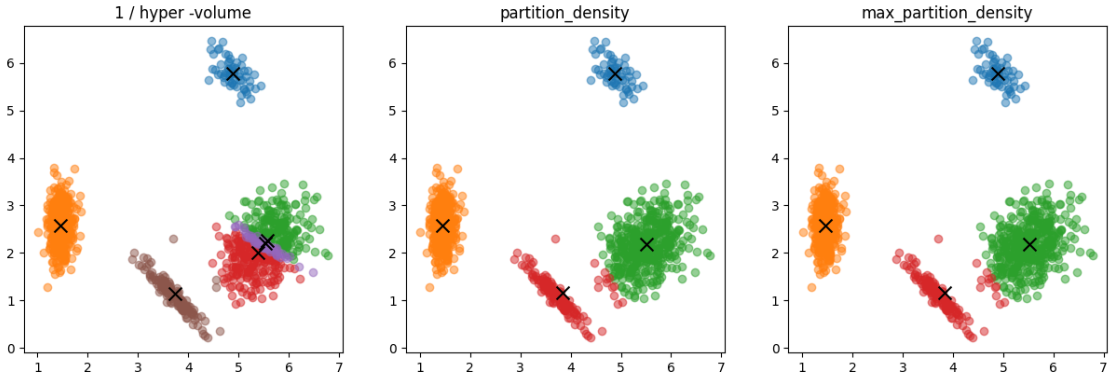
6th experiment:

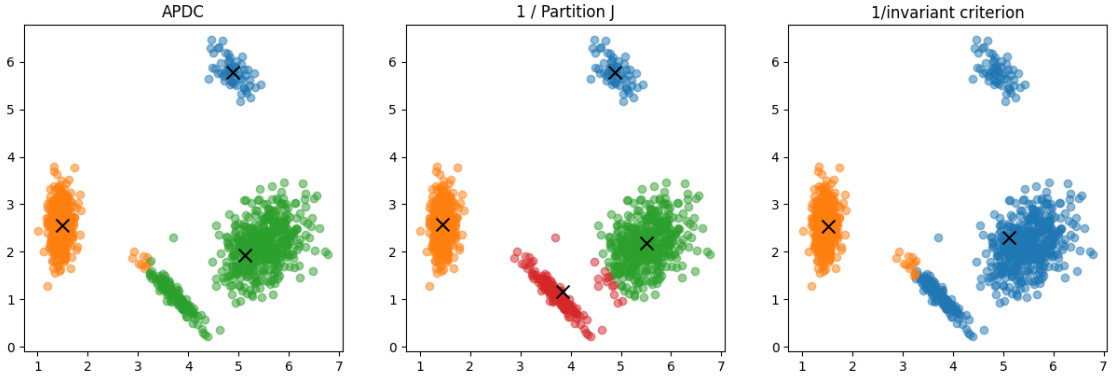
K = 4
d = 2




1 / hyper -volume
partition_density
max_partition_density
V
K

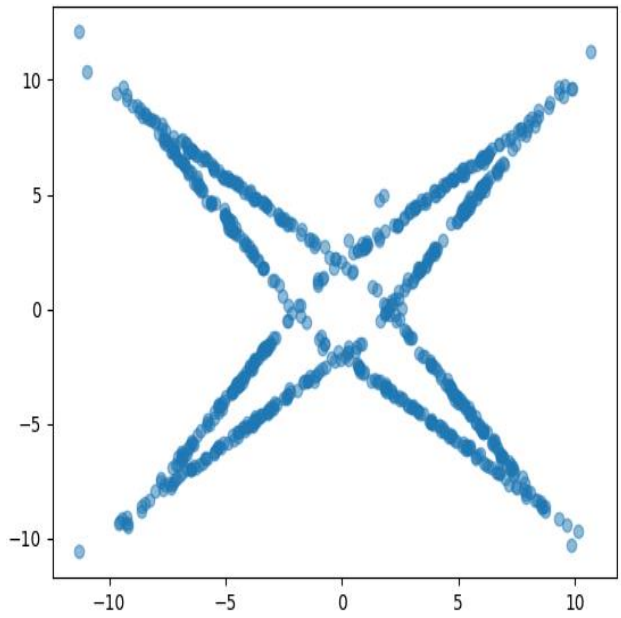

APDC
1 / Partition J
1/invariant criterion
V
K

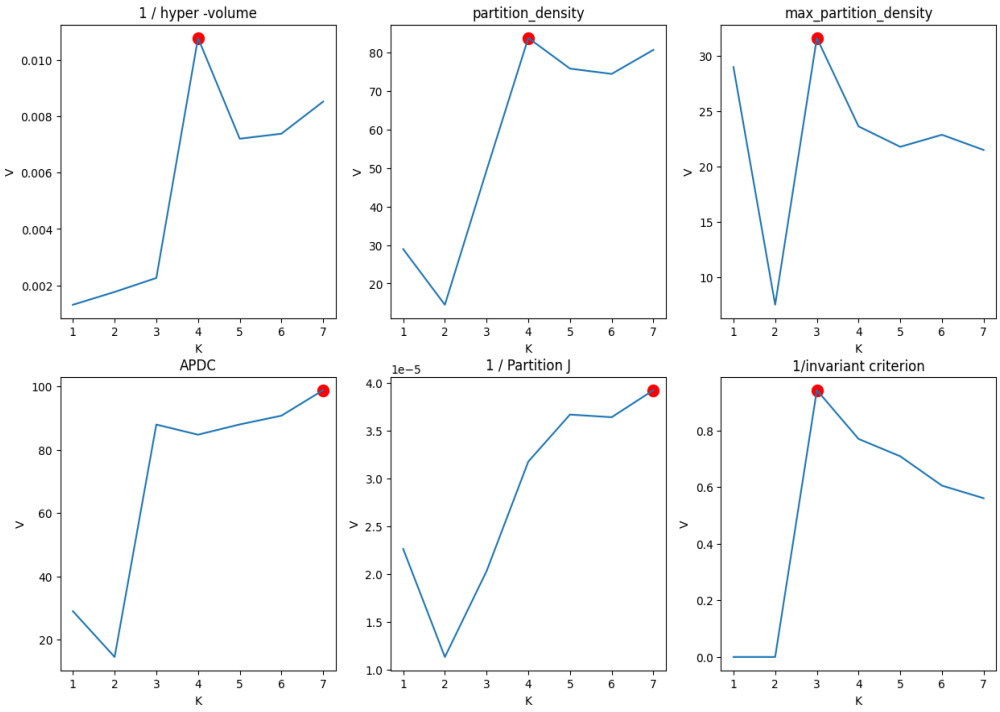

1 / hyper -volume
partition_density
max_partition_density

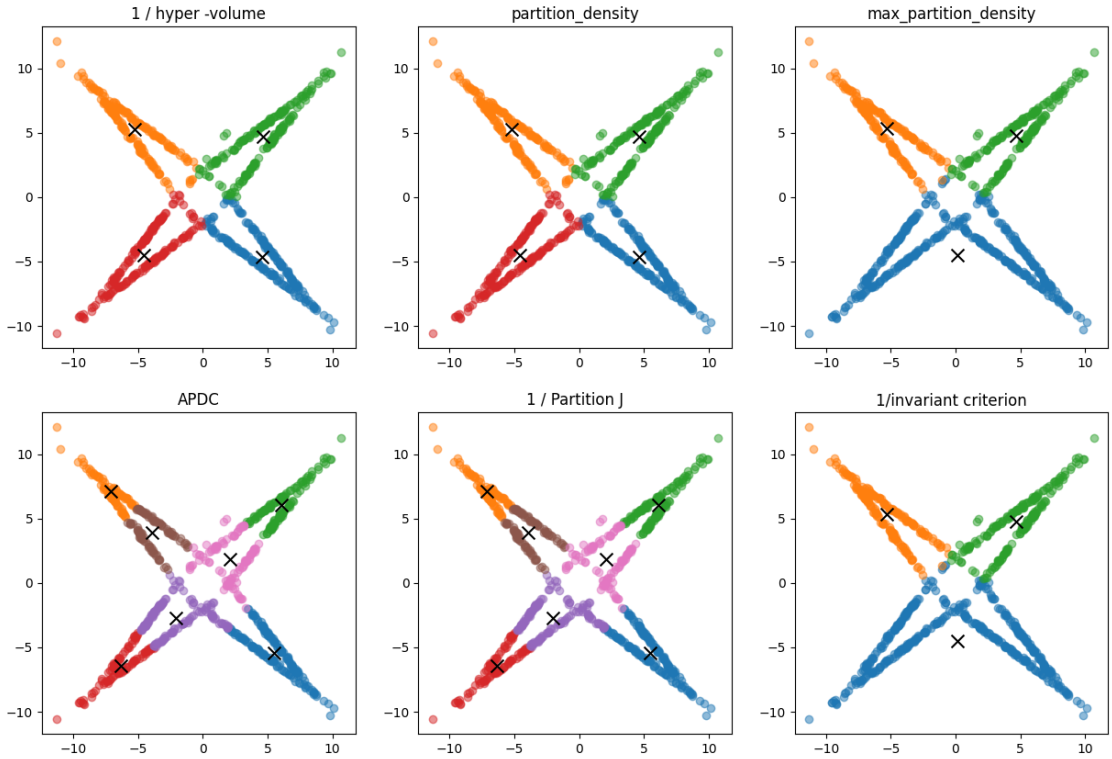

APDC
1 / Partition J
1/invariant criterion

7th experiment: back to the stars

it is logical to ecxpect that more criteria would return the correct number of star arms if they arefurther apart, so let us create gaussian dataset shaped a star with 4 arms, making the angular distance between them grater.







results

the results show that indeed the WUFOC was more successful on this dataset, as before, the hypervolume criterion cluster but also the partition density returned the number of arms, and the partition is accurately what we did expect, this is likely due to the enhanced distance between them in this dataset. more criteria got a close result of $k = 3$ which is indeed something to work with