Prof. Dr. Leif Kobbelt
Gregor Kobsik, Tim Elsner

# Theoretical Exercise 1

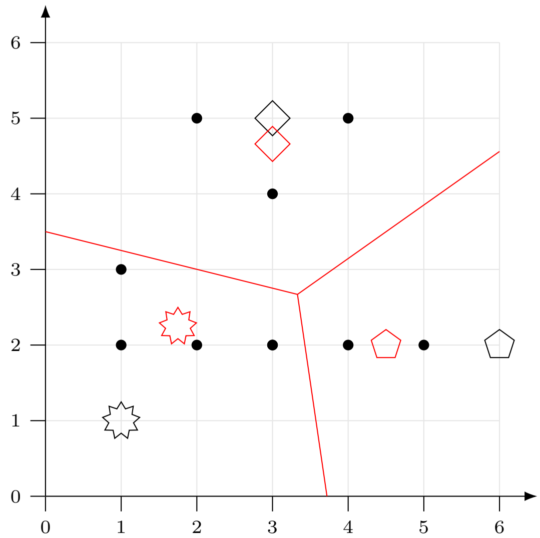
Date Published: April 20th 2023

# This is the sample solution.

## Exercise 1 Clustering

### (a) K-Means

Draw the underlying Voronoi Diagram to compute the assignment of each point to the corresponding cluster centers. Compute the new cluster centers after the assignment.




**Solution**

$$\mu_1 = \frac{1}{4}\left[\begin{pmatrix}1\\2\end{pmatrix} + \begin{pmatrix}1\\3\end{pmatrix} + \begin{pmatrix}2\\2\end{pmatrix} + \begin{pmatrix}3\\2\end{pmatrix}\right] = \frac{1}{4}\begin{pmatrix}7\\9\end{pmatrix} = \begin{pmatrix}1.75\\2.25\end{pmatrix}$$

$$\mu_2 = \frac{1}{2}\left[\begin{pmatrix}4\\2\end{pmatrix} + \begin{pmatrix}5\\2\end{pmatrix}\right] = \begin{pmatrix}4.5\\2\end{pmatrix}$$

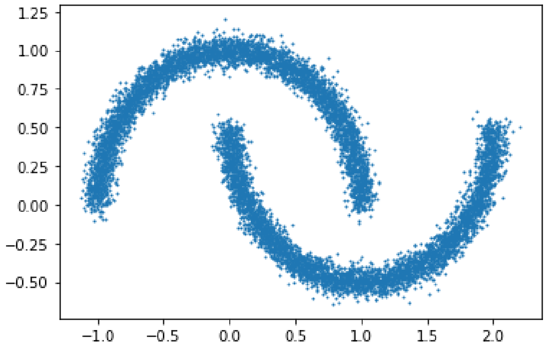
$$\mu_3 = \frac{1}{3}\left[\begin{pmatrix}2\\5\end{pmatrix} + \begin{pmatrix}3\\4\end{pmatrix} + \begin{pmatrix}4\\5\end{pmatrix}\right] = \frac{1}{3}\begin{pmatrix}9\\14\end{pmatrix} = \begin{pmatrix}3\\4.66\end{pmatrix}$$

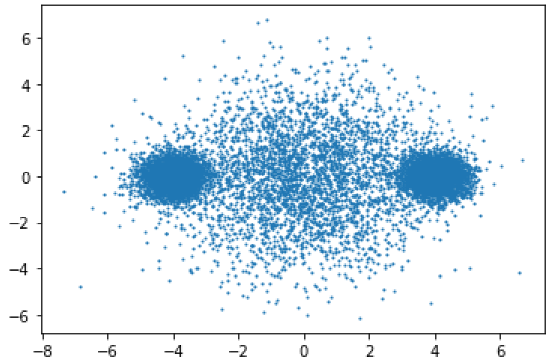
## (b) K-Means

Assume you could pick the perfect positions for 3 K-Means cluster centers. Draw a dataset (consisting of 3 clusters) where K-Means would not recover the ground truth clusters.

**Solution**

Anything with non-convex clusters or very uneven sizes.



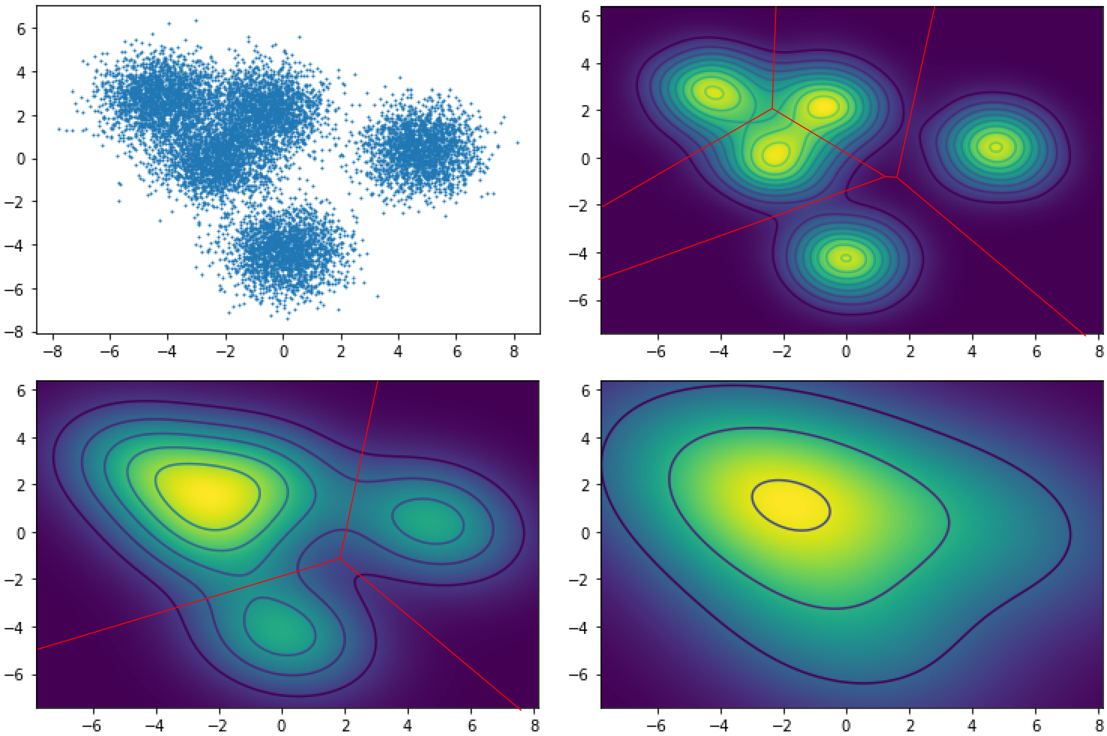

## (c) Clustering Methods

Categorize the different clustering methods according to the shapes can model.

| method | shape |
|---|---|
| K-Means | Voronoi Cell Shapes |
| GMM | Ellipsoids |
| Mean Shift | no particular shape |
| Hierarchical Clustering | no particular shape |
| Spectral Clustering | no particular shape |

## (d) Mean Shift Clustering

Given are the following sampling and density plots:



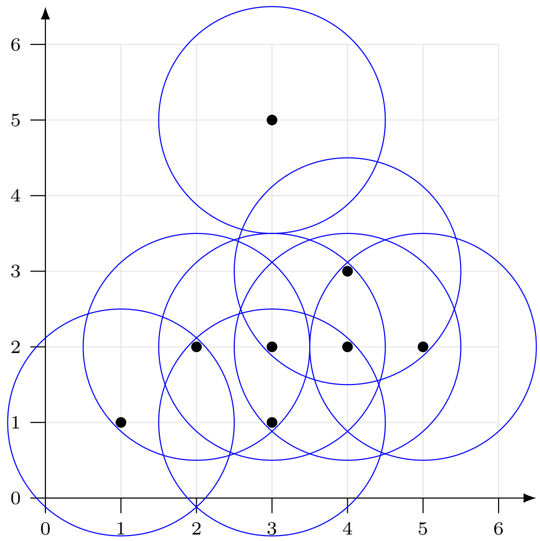
For each density plot draw in the cluster boundaries you would obtain with mean shift clustering. Which parameter of the algorithm needs to be varied, to obtain these different results?

**Solution**

In Mean-Shift clustering we do gradient ascent on the density function for each point. All points where this procedure converges to the same goal build a cluster together. This can easily be done graphically (at least approximately).
The bandwidth parameter of Mean-Shift Clustering can be seen as the area over which we smooth the density. So varying this parameter results in the different plots we see here.

## (e) DBSCAN

Classify each point according to the DBSCAN clustering method. The radius $\epsilon$ is visualized in blue. The minimum number of neighbours is 3.


0
1
2
3
4
5
6
0
1
2
3
4
5
6