Exercises for
Pattern Analysis Programming
Lina Felsner, Mathias Seuret, Dalia Rodriguez-Salas
Task 5: May 20 – May 27

Pattern
Recognition
Lab

Clustering: K-Means versus Mean Shift

Let us have a closer look at the inner workings of K-Means and Mean Shift. Mean Shift clustering assigns each point to its associated mode in the density. K-Means clustering assigns each point to the nearest mean vector. The purpose of this exercise is to understand under which circumstances these two approaches lead to different results.

Exercise 1 For this exercise, we will resort to a synthetic distribution that consists of random samples from three 2-D Gaussian distributions. Use K-Means and Mean Shift algorithms to cluster the samples.

We believe it is a good exercise if you implement these algorithms yourself. The implementation is not difficult, and it helps to deepen your understanding of the algorithms. However, it is also not forbidden to use already implemented versions, such as `sklearn.cluster.MeanShift` or `sklearn.cluster.KMeans`.

For details on the K-Means algorithm, please have a look at the book by Hastie/Tibshirani/Friedman Sec. 14.3.6 or in the book by Bishop Sec. 9.1. Details for the Mean Shift algorithm can be found in the paper by Comaniciu/Meer (uploaded to studOn Sec. A), together with the lecture video and slides 02b.

You can quickly test if both algorithms work by having the center of each of the three Gaussians well separated from each other. An example is shown in Fig. 1.
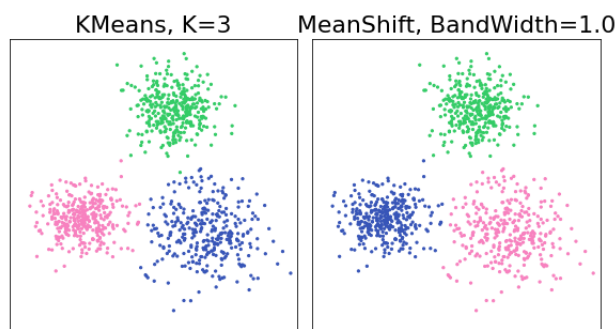

Figure 1: Example with three Gaussian distributions. The clustering result for K-Means (left) and Mean Shift (right) are color coded.

Exercise 2 From the results of the first part of the exercise (e.g. Fig. 1), the difference between the algorithms is not clear. Both give the same clustering result. However, from the theory we can expect that there is a difference. Think about and design your own density, where the two algorithms give different results. Any density is valid, and there are certainly different possibilities and approaches to create such a density. Maybe think about a density that is particularly simple, or that realizes the difference between the methods in a particularly instructive way; or

just report an unexpected case. There is an infinite amount of different possible densities you can try, so have fun exploring!

Please post a figure of your density with the respective results of K-Means and Mean Shift clustering on the forum and add a short text describing what made you choose that specific density.

Comments:
We ask for only <u>one figure per group</u>. Please also state your group number. Bring your code to the joint meeting on May 27 or May 28 for a potential little extra experiment.