

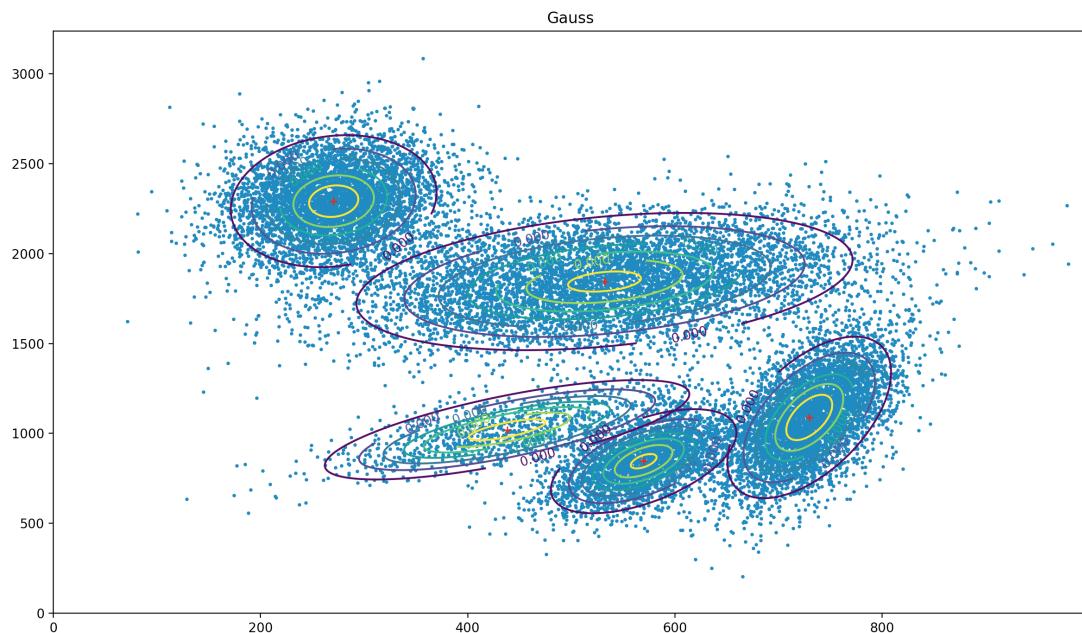
Assignment 5

Computational Intelligence SEW, SS2017

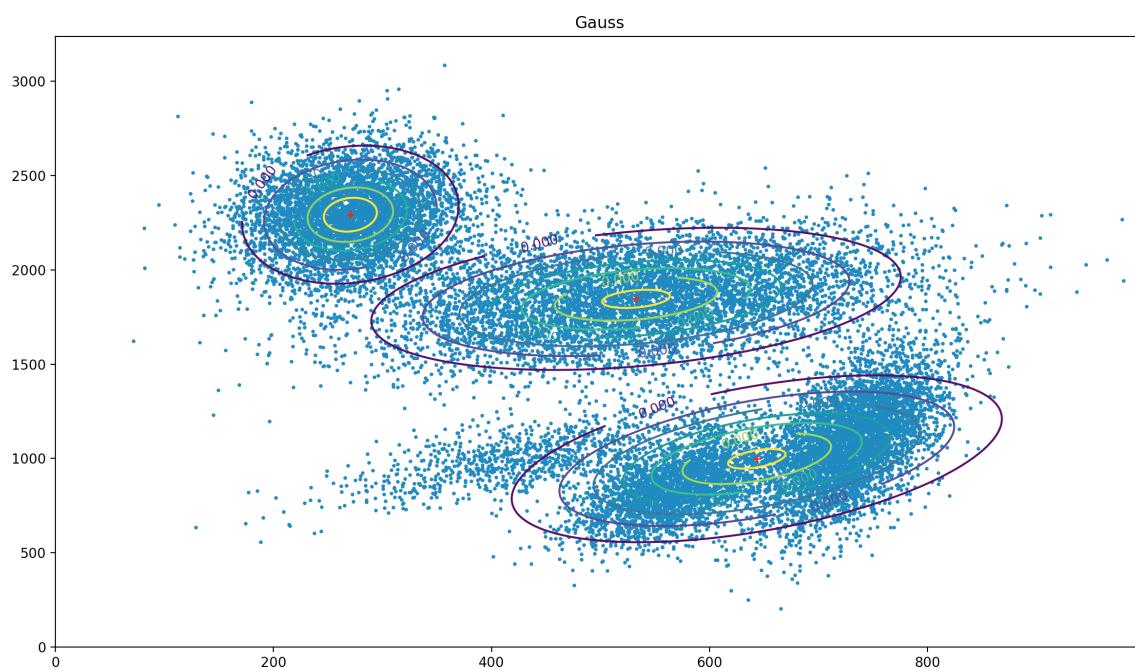
Team Members		
Last name	First name	Matriculation Number
Papst	Stefan	1430868
Guggi	Simon	1430534
Perkonigg	Michelle	1430153

1 Expectation Maximization Algorithm

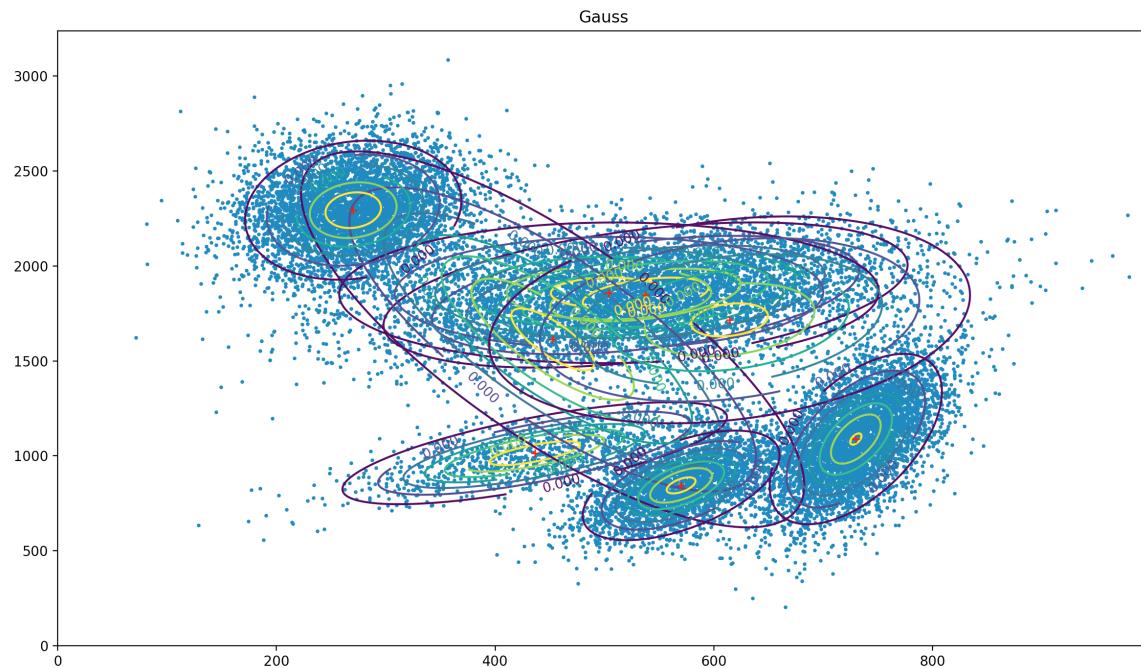
Scatterplot with Gauss mixture model over the plot with $M = 5$:



Plot with $M = 3$



Plot with M = 8



How do you choose your initialization Θ_0 ?

The init values are chosen like in the script is described.

$$\alpha_m^0 = \frac{1}{M}$$

$$\mu = \frac{1}{N} \sum_{n=1}^N x_n.$$

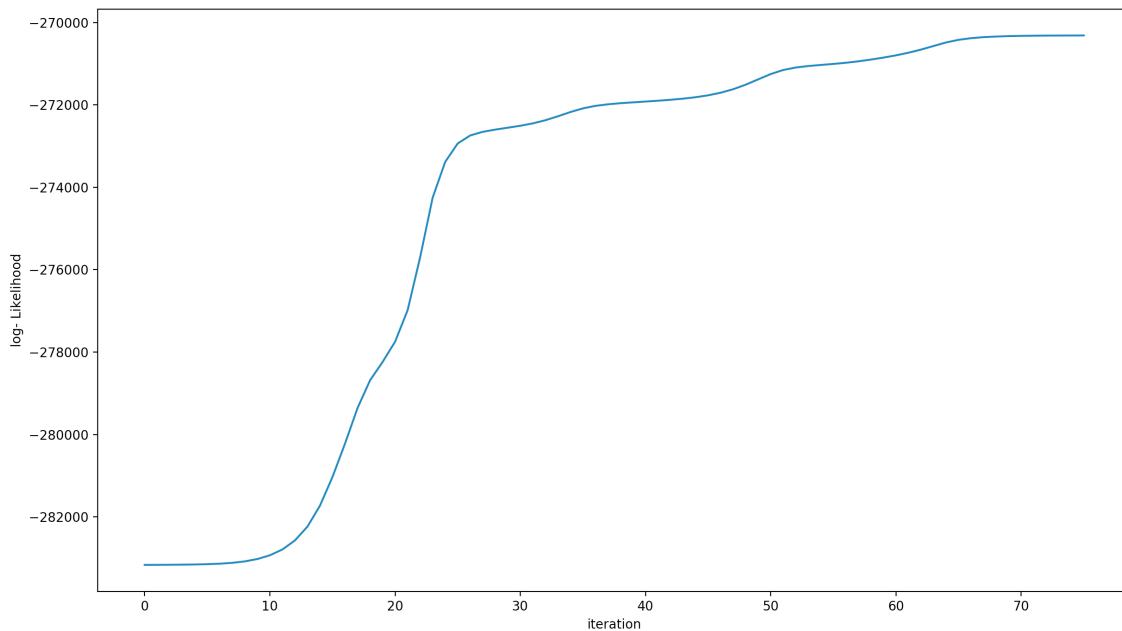
$$\Sigma = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \mu)(\mathbf{x}_n - \mu)^T$$

N is the number of sample we have chosen for initialisation, which was in our case 1000.

Does this choice have an influence on the result?

Yes, because these values are used for further calculations, so they influence the results. A possible issue would be if some μ' are close to each other, so their GMM is overlapping.

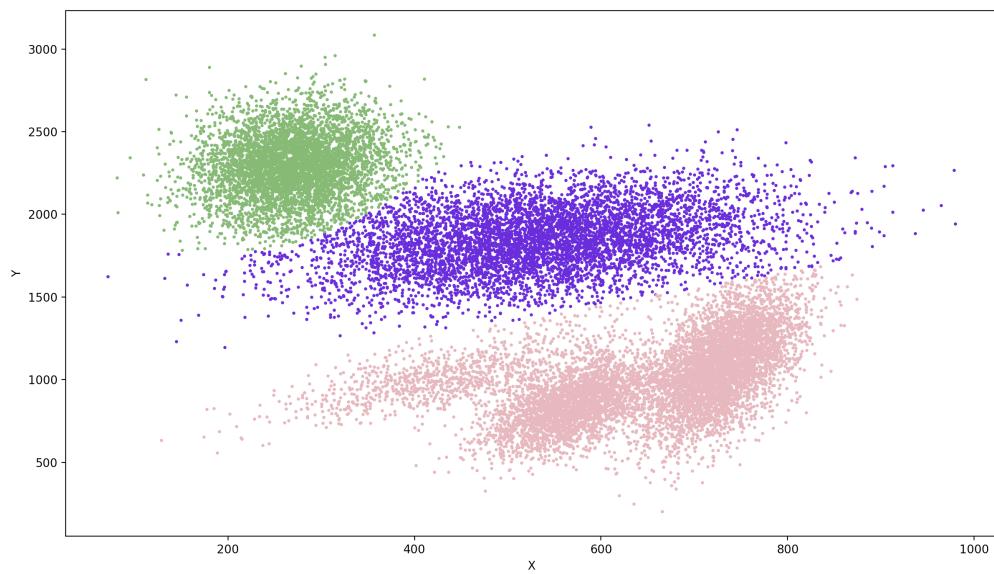
Log Likelihood over iterations:

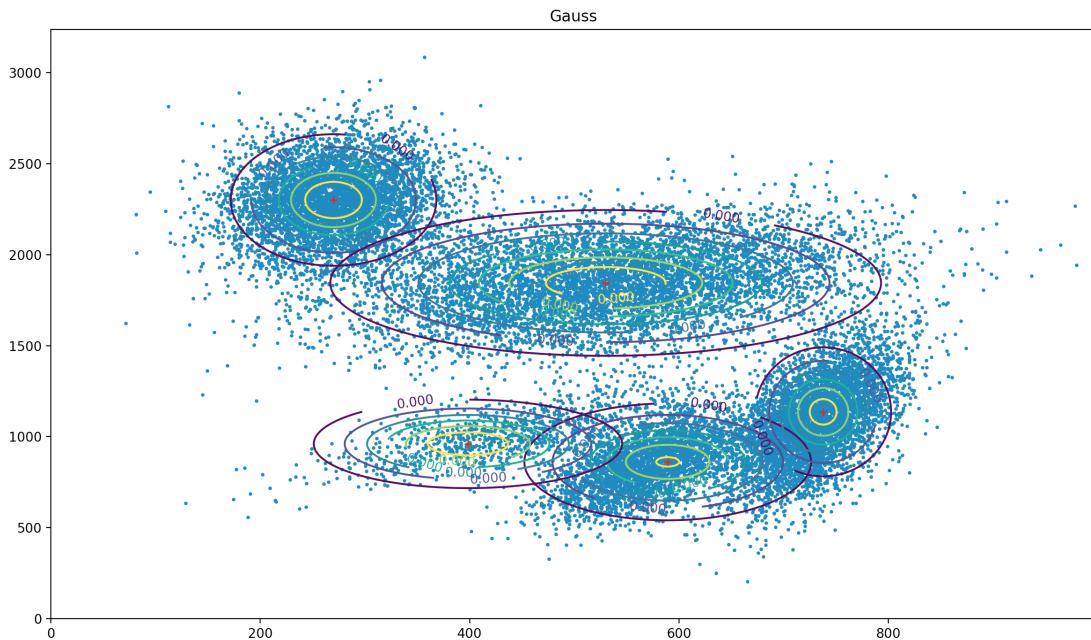


You can see that the log is converging to a certain maximum, which means that the right parameters are found and no significant changes are possible on more iterations.

Within your EM-function, confine the structure of the covariance matrices to diagonal matrices!

M = 5



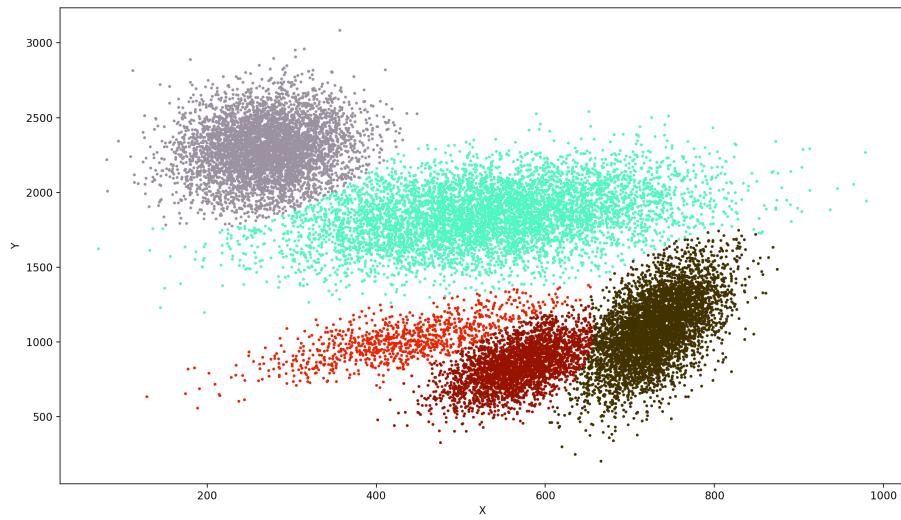


What is the influence on the result?

A diagonal matrix as covariance matrix means that the gauss curves are circle or elliptic, but they are not rotated.

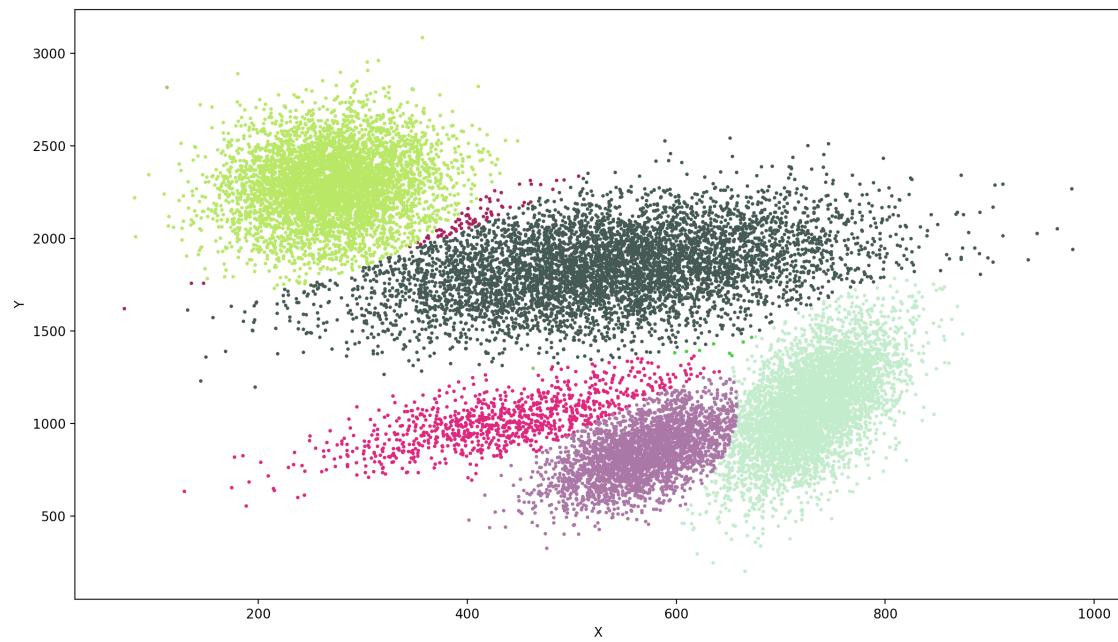
Soft Classification

$M = 5$



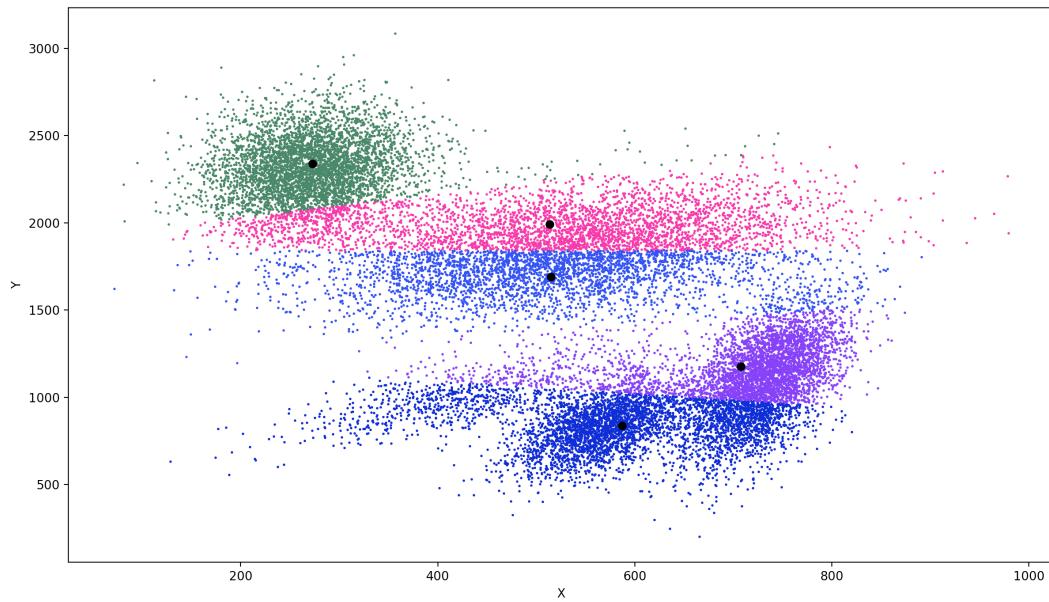
$M = 3$

M = 8

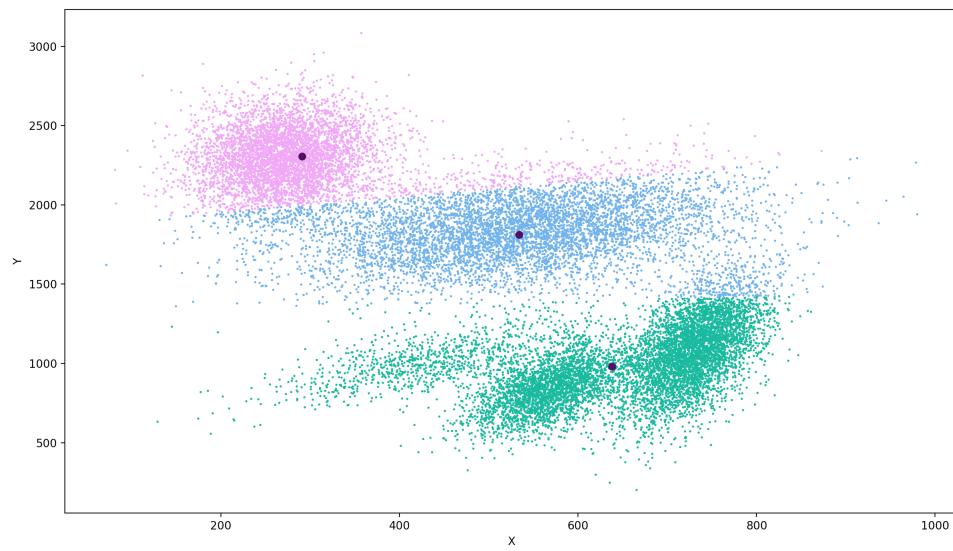


2 K - Means Algorithm

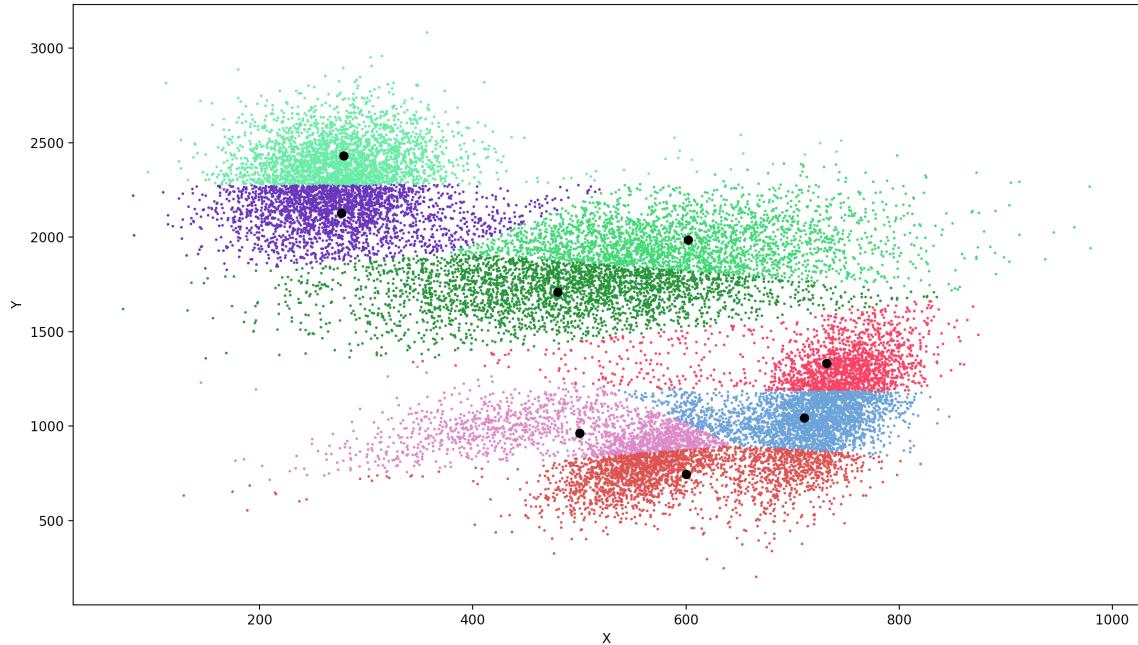
M = 5



M = 3



$M = 8$

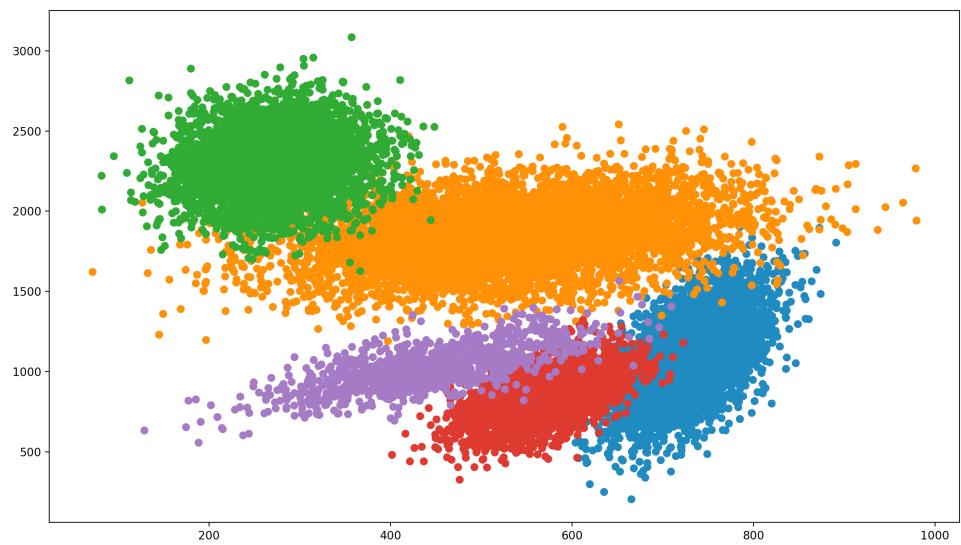


As you can see, the results are similar to the ones of EM. If the M is correct, the classification works very well, but on the one hand if there are too less, the clusters are to wide and on the other hand if the M value is set to high, the k-means splits the data set in too many clusters.

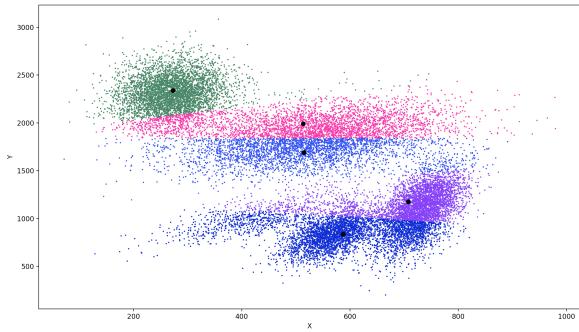
What is the nature of the boundaries between the classes?

The boundaries are very straight instead of the soft classification of EM.

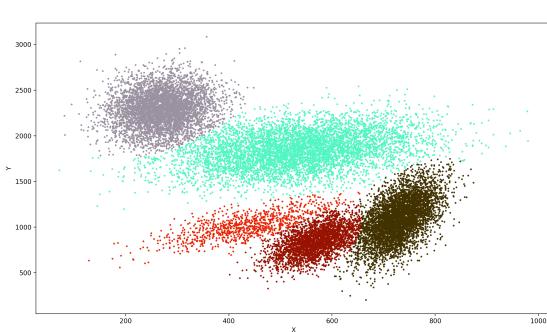
Compared to the labelled data, the soft classification is better structured than the k-mean results.



labelled data



k- means



soft classification