Assignment 07: Gaussian mixture models

Abstract

MLE cannot be used and becomes difficult to apply when the data comes from more than 1 gaussian distribution. So we utilize GMM along with latent variable assumption $z_{ij} = 1 \in$ if comes from a gaussian and $z_{ij} = 0$ if not from a gaussian.
The goal is to maximize the log likelihood of:

$$l(\theta) = \sum_{i=1}^{m} \log p(x^{(i)}; \theta)$$

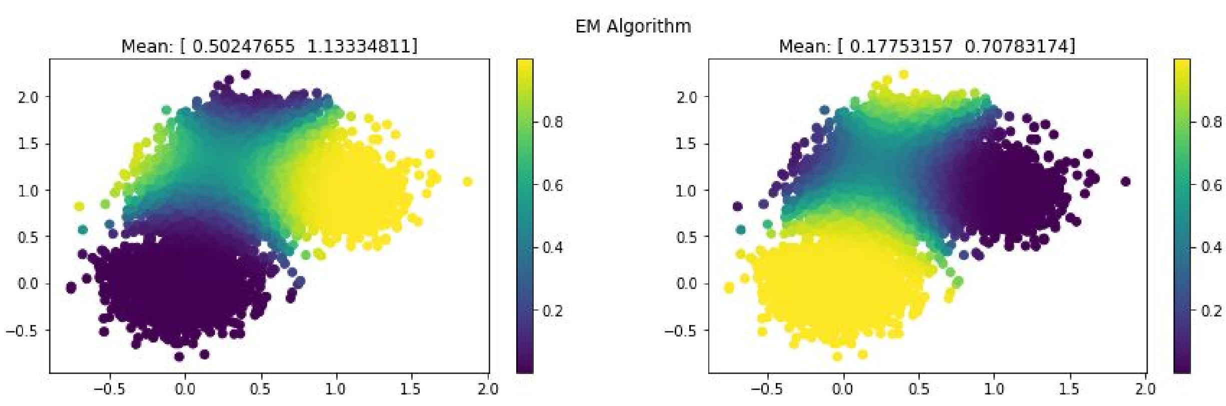
This consists of two steps, the E-step (estimation step), where we methodically guess values for the variance, mean, and probability of belonging for each point (pi) and compute the responsabilities for the variables involved. The M-step (maximization step), where we calculate the formal values for the mean, variance, and probabilty. Thus, the E and M steps are iterated until convergence is found or a maximum bound of iterations are approached.

Question

1. Does EM find global optima?
   No, it is very difficult for an alternating optimization problem to find the global optima because it is very dependent on this initialization values.

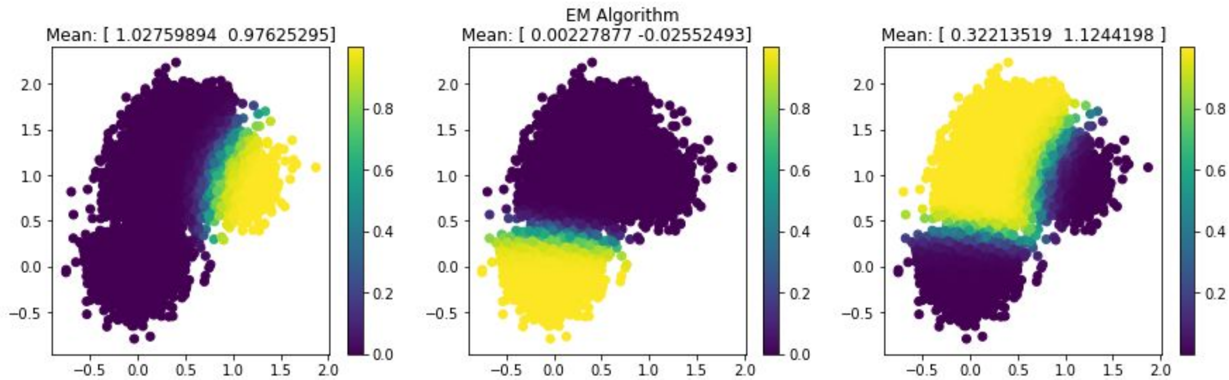
Figures

Code running for K = 2



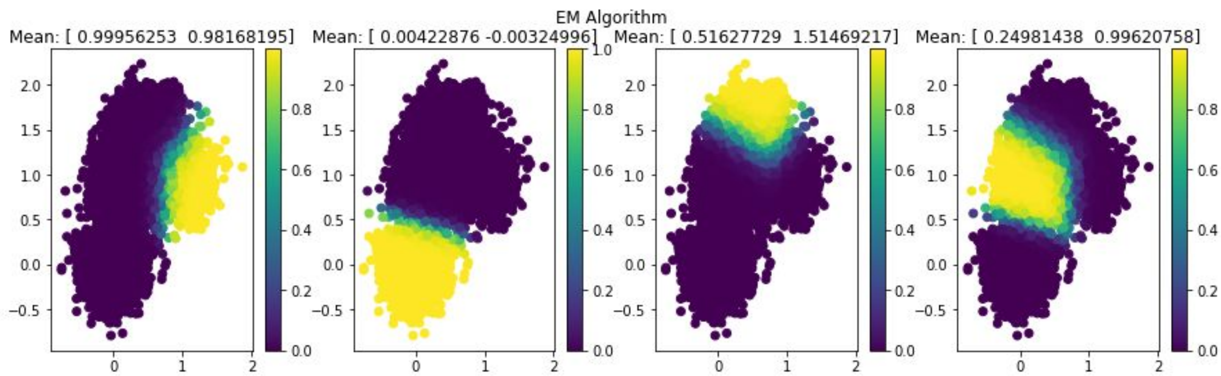
Is clearly defining two clusters and the middle area is inconclusive because the data blends in two different direction when comparing figure 1 and figure 2

Code running for K = 3



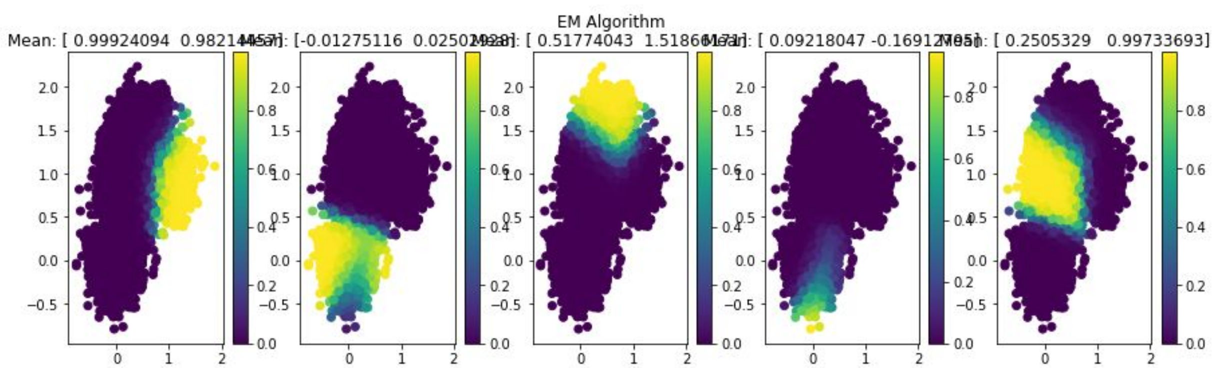
For this execution the regions are very defined and isolated in each plot with a probabilty gradient separating each cluster from the next.

Code running for K = 4

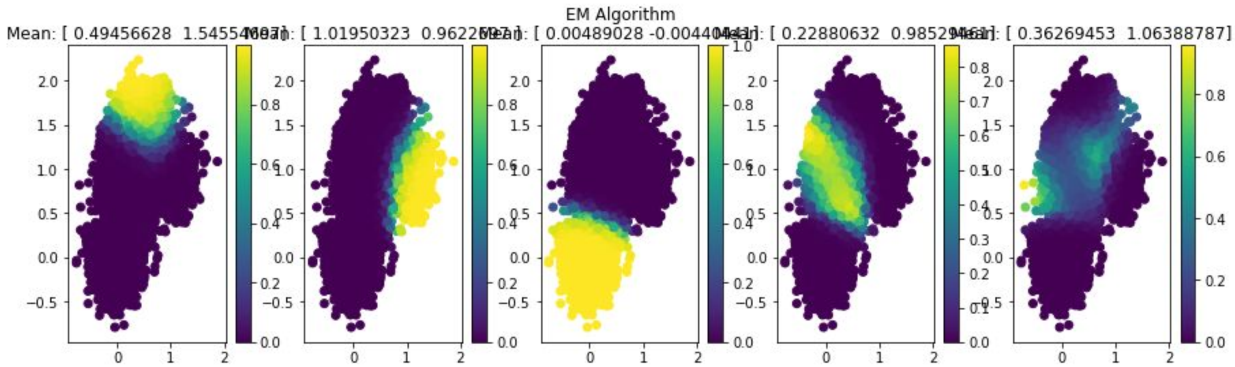


For K cluster = 4 there is clear definition but some look very similar to the ones in the K = 3, also some of the areas where the probablilties are around 0.5 (cyan-green) indicates too much competition for the same region in space

Code running for K = 5



Code running for K = 5 (second trial)



There are some differences between both iterations of K = 5, this alone tells us that it cannot be K = 5 because the EM algorithm is deterministic. Another thing to note is that, if two similar shaped plots occur in K =5 we could use that to initialize our data and could optimize our code and output plots

Discussion

How many mixture components?

3

How was this determined?

Trial and error, after plotting the different K's the one with the best clustering shapes is k = 3. Also the convergence rate for K = 3 was the quickest in determining the pi probabilities in regards to size. When NumComponents is determined apriori, the data fits the best.

How was the data initialized?

The each gaussian was mean centered and had a isotropic covariance matrix, thus the data was not affected is a point was chosen at random which is what we did in this python script.