

Lecture 16 - 3: Likelihood of Gaussian Mixture Model

Tabaré Pérez

May 6, 2020

$$\theta; \mathbb{P}_1 \dots \mathbb{P}_K, \mu^{(1)} \dots \mu^{(K)}, \sigma_1^2 \dots \sigma_K^2 \quad (1)$$

So as you see, there's a whole family of these parameters that we need to determine. But before, we're going to start having a conversation of how, actually, to compute these θ s I just want to show you yet another piece.

How can we compute the likelihood of a particular point to be generated given these parameters?

And if you look here, I ask this question in the case of a single Gaussian.

When we have our parameters and we ask: what is the likelihood of x ?

So here, remember that when I was writing x and θ there is actually the whole family of these parameters, How can we do it?

And now you remember that the point may be generated from various different clusters. We don't know which one contributed to it. So what we will do, we will now marginalize over cluster assignment, meaning we would look at the likelihood of it being generated by all possible clusters, and then we will sum it up:

$$\mathbb{P}(x|\theta) = \sum_{j=1}^K \mathbb{P}_j \mathcal{N}(x, \mu^{(j)}, \sigma_j^2) \quad (2)$$

So in this case, we are going to go from 1 to K and then look at the sum of the likelihoods of the point being generated for each one of the clusters.

So now we have everything we need to think about a mixture model.

And the next question for us is how we can actually find all these different paths, all these different parameters, given a specific assignment of points on the plane.

And you can see that there is a lot of dependency here because, you know, in general case, we don't really know from where all these points come. We

don't know to which class they belong. So that makes the matter much more complex.

So let me first write you down the likelihood. We will do exactly the same thing as we've done before. We said OK, I'm giving you some set of points, S_n :

$$\mathbb{P}(S_n|\theta) = \prod_{i=1}^n \sum_{j=1}^K \mathbb{P}_j \mathcal{N}(x, \mu^{(j)}, \sigma_j^2 I) \quad (3)$$

It has n points with parameter θ . And what, in fact, it means? That we need to go and multiply, because all those points are independent.

We're going to go from point 1 to point n and multiply its likelihood to be generated by this Gaussian.

So now given this expression, I want to take the derivative of it with respect to my parameters, make it equal to 0, and find the parameters.

It turns out, it's actually a pretty complex task. So what we will do now, we will start with an easy case.

The first case that I will consider with you is the case which is called observed. I would assume that somebody gave me this point, but they actually gave me the hard assignment. They said, this point belongs to the first cluster, and this point belongs to the second cluster. You're going to measure them. Then, you know, if we would have here red and blue, that some of them are red and some of them are blue. So we have no uncertainty. We know your belonging. And I will demonstrate to you how we can solve that piece. And if we understand it well, we can very easily translate it into the case where the assignment is not observed. And that's where we are going to be talking about **E-M ALGORITHM**.

NOTES

- What is the difference between "mixture components" and Clusters ?
- What is the difference between "mixture weights" and probability?

You can consider them as synonyms although they are not quiet. Here is my understanding:

- The number of mixture components is the true number that underlies the generation of the data (the data is an output), while the number of clusters (in the K-means and K-medoids algorithms) is the size of the partition you want your algorithm to generate from the data (the data is an input). When analyzing the data, often we won't have prior knowledge of what the real number of mixture components is. We'll

have to either guess it or and try to estimate it by iterating over different numbers of clusters and comparing their outcomes.

- "Mixture weights" are probabilities if the number of components is finite. If we were to face a non-discrete set of mixture components, then we would need to use the probability density terminology.