**CMPE 257 :: Assignment4 :: Gaussian Mixture Model**

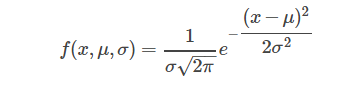
**Chaitrali Deshmukh (012501489)**

[Google Colab Link:](https://colab.research.google.com/drive/1SNAPL09Ls6P9wTx_qeLHffzR97AdJ1oj#scrollTo=FSUYiFg3uZoU)

**What is Gaussian Mixture Model?**

* Gaussian is a type of distribution which is a listing of outcomes of an experiment and the probability associated with each outcome.
* Its an even distribution and also used for density estimation.
* GMM is used for generative unsupervised clustering. It is also called as EM(Expectation-Maximization) clustering.
* We use GMM, when we know that our dataset contains multi-mode or bumps. Mode is the most common value in our dataset. It is useful for modeling more complex data, that has multiple peaks.
* GMM is the sum of weighted gaussians.

Gaussian Distribution in its 1-dimension form can be written as:

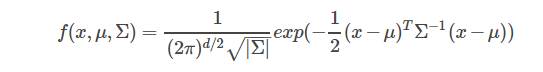


where, x is the data point,

μ is the mean of the data,

σ is the covariance

In two-dimensional case:



where, Σ is now the covariance matrix of the Gaussian.

μ controls the "center position" of the Gaussian,

Σ controls the "shape" the Gaussian.

Covariance is a measure of how changes in one variable are associated with changes in a second variable. Specifically, covariance measures the degree to which two variables are linearly associated.

**How it can be optimized?**

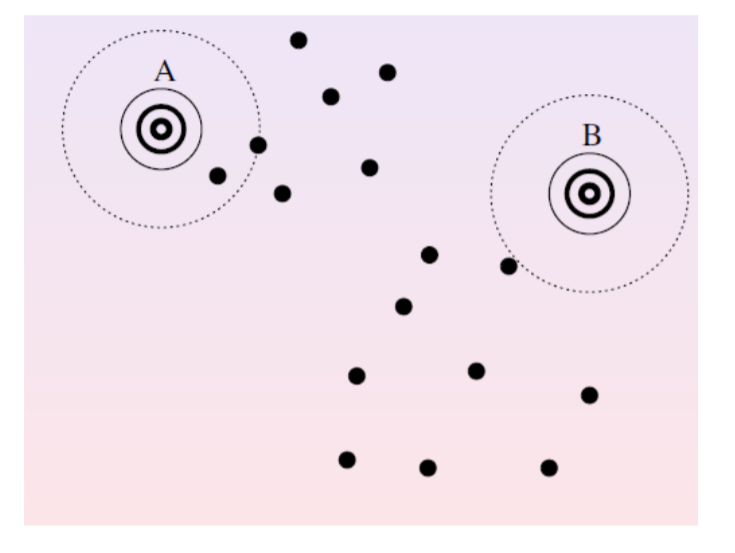
* Expectation maximization is the technique which is used to estimate the mixture model's parameters if we know the number of components K. Models are typically learned by using maximum likelihood estimation techniques, which seek to maximize the probability, or likelihood, of the observed data given the model parameters. Unfortunately, finding the maximum likelihood solution for mixture models by differentiating the log likelihood and solving for 00 is usually analytically impossible.
* Expectation maximization (EM) is a numerical technique for maximum likelihood estimation. It is an iterative algorithm and has the convenient property that the maximum likelihood of the data strictly increases with each subsequent iteration, meaning it is guaranteed to approach a local maximum or saddle point.
* It introduces a hidden variable so that its knowledge would simplify the maximization of likelihood (Probability).

**At each iteration:**

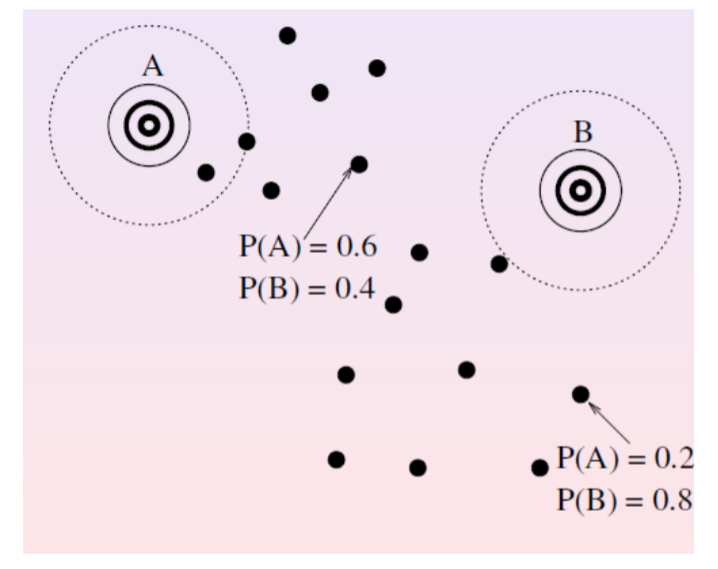
**E-step:** Estimate the distribution of the hidden variable given the data and the current value of the parameters.

**M-step:** Maximize the joint distribution of the data and the hidden variable.

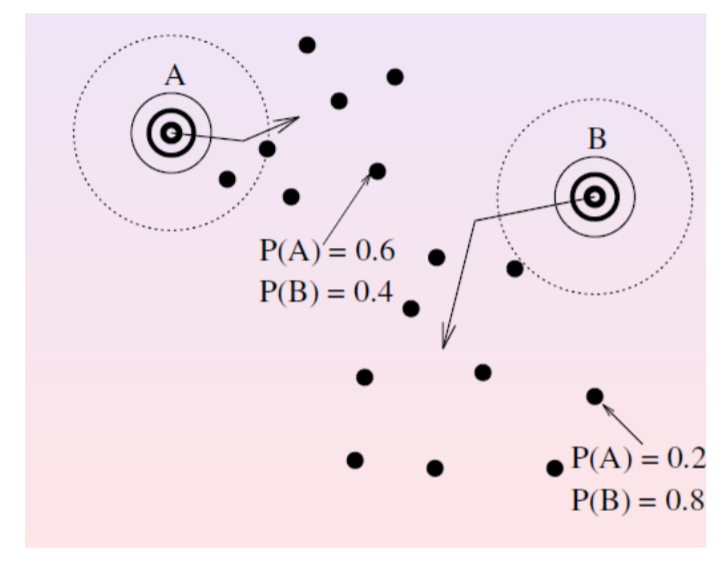
**Hidden variable at each point:**



**E-Step:** for each point, estimate the probability that each Gaussian generated it.



**M-Step:** modify the parameters according to the hidden variable to maximize the likelihood of the data (and the hidden variable).



**Comparison with K-means:**

1. One of the major drawbacks of k-means is its naive use of the mean value for calculating centroids. GMM is more flexible than k-means. The data points in GMM are Gaussian distributed and we have two parameters to specify the shape of clusters i.e. the mean and standard deviation. In a two-dimensional example, this means that clusters can take any kind of elliptical shape. On top of it, we can optimize these parameters with Expectation–Maximization algorithm.
2. Secondly, k-means uses hard clustering method which means that each point will be associated to one and only one cluster. There is no way to find out that what is the probability that the data point belongs to other clusters in the dataset given. On the other hand, GMM uses soft clustering where it uses probabilities which means that they can have multiple clusters for each data point. So, if a data point is in the middle of two overlapping clusters, we can describe its class by saying it belongs X% to class 1 and Y% to class 2. In short, GMMs support mixed membership.

**GMM Use Cases:**

1. GMM is commonly used in object tracking of multiple objects, where the number of mixture components and their means predict object locations at each frame in a video sequence. It is used to update component means over time as the video frames update, allowing object tracking.
2. It is also used in feature extraction from speech data for use in speech recognition systems.

Google Colab Link: <https://colab.research.google.com/drive/1SNAPL09Ls6P9wTx_qeLHffzR97AdJ1oj#scrollTo=FSUYiFg3uZoU>

Dataset Link:

<https://drive.google.com/open?id=1JQYwWgH-8iKHmeYSf0rT0Z0jp2KS4rAL>