8 Gaussian Mixture Models & EM

Intro

k-means :

- -a cluster is described only by its centroid
- -hard clustering technique

-Expectation Maximization, using Gaussian Mixture Models
-Each cluster : centroid, covariance and weight
-We now can calculate the probability of a point belonging to a cluster

8.1 The Gaussian distribution

$$\mathcal{N}(x|m,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-m)^2\right\}$$

For a D-dimensional vector X:

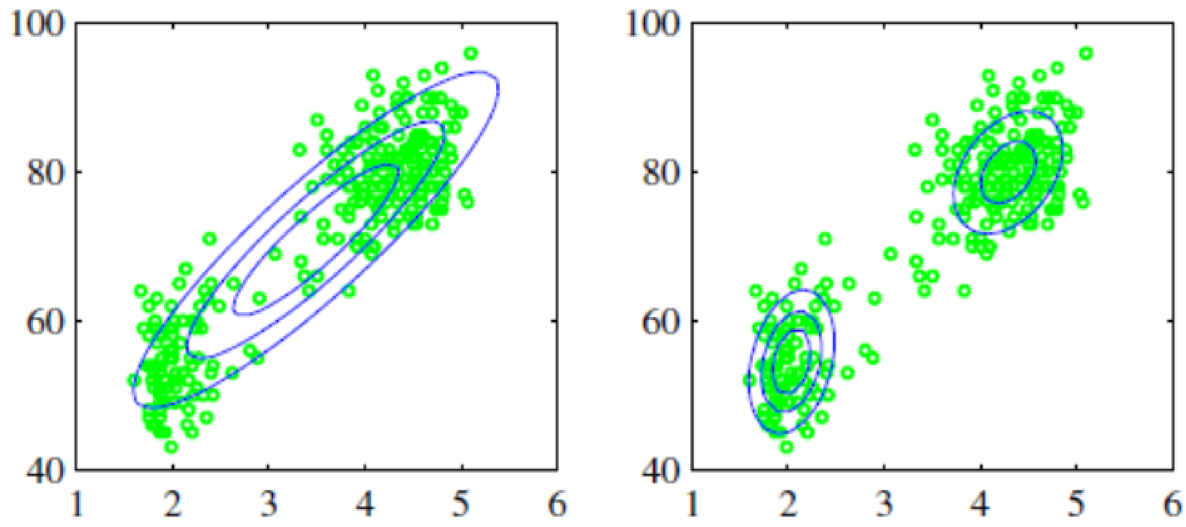
$$\mathcal{N}(X|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(X-\mu)^T\Sigma^{-1}(X-\mu)\right\}$$

8.2 Mixture of Gaussians

Gaussian distribution has significant limitations when modeling real data sets.

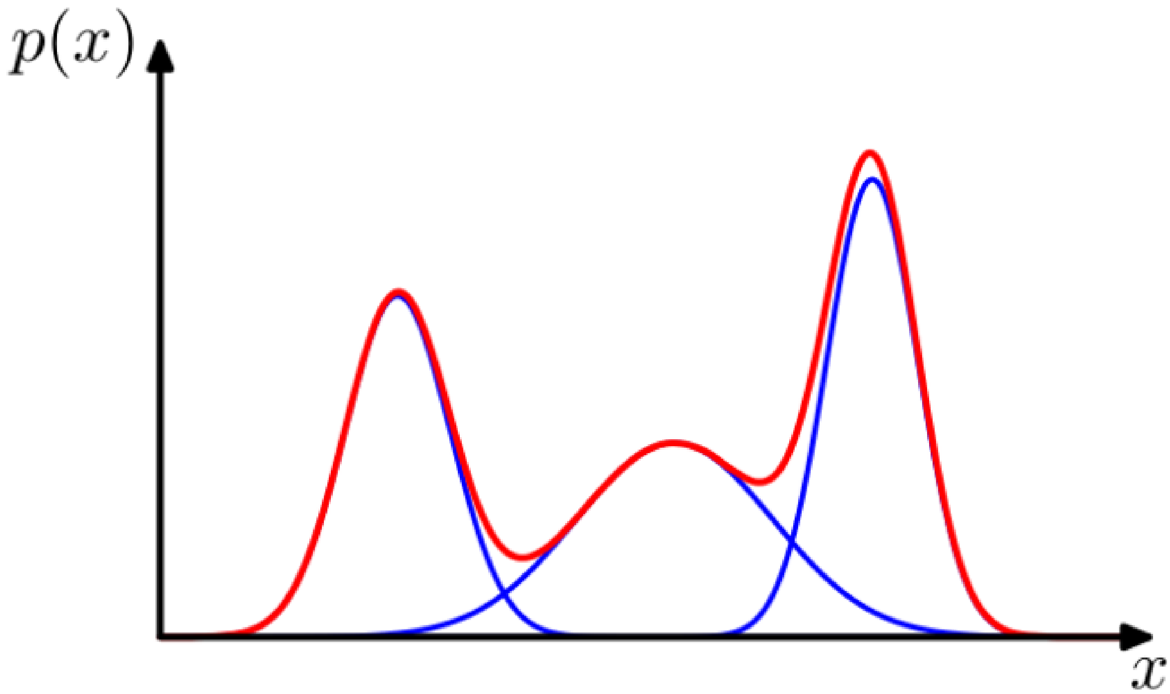
Let’s take this example :

- 272 measurements of eruptions of a geyser
- horizontal axis : duration of eruptions in minutes
- vertical axis : time to the next eruption in minutes



- The data set forms two dominant clusters ⇒ a simple gaussian distribution is unable to capture this structure whereas a linear distribution of gaussians gives a better characterization of the data set
- Superpositions formed by taking linear combinations of basic distributions like Gaussians can be formulated as probabilistics models known as mixture distribution

By a linear combination of gaussians, we obtain some complex densities like the red curve :



- With a sufficient number of Gaussians and with means and covariances adjusted, we can approximate almost any continuous density with an arbitrary accuracy

Let's consider a superposition of K Gaussians densities with the following formula :

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)$$

which is called "mixture of Gaussians" with the parameters :

- $\mathcal{N}(x|\mu_k, \Sigma_k)$ is called a "component" of the mixture and has its own mean μk and covariance Σ_k.
- PI-k are called mixing coefficients and verify the conditions $\sum_{k=1}^{K} \pi_k = 1 \quad \text{and} \quad 0 \leq \pi_k \leq 1$

Now we have to find a formula for our mixture which is introduced by a latent variable.

$$p(z_k = 1) = \pi_k$$

Note: A latent variable is a variable that is not directly measurable, but its value can be inferred by taking other measurements.
After, the conditional distribution of X given a particular value for z is a Gaussian:

$$p(x|z_k = 1) = \mathcal{N}(x|\mu_k, \Sigma_k)$$

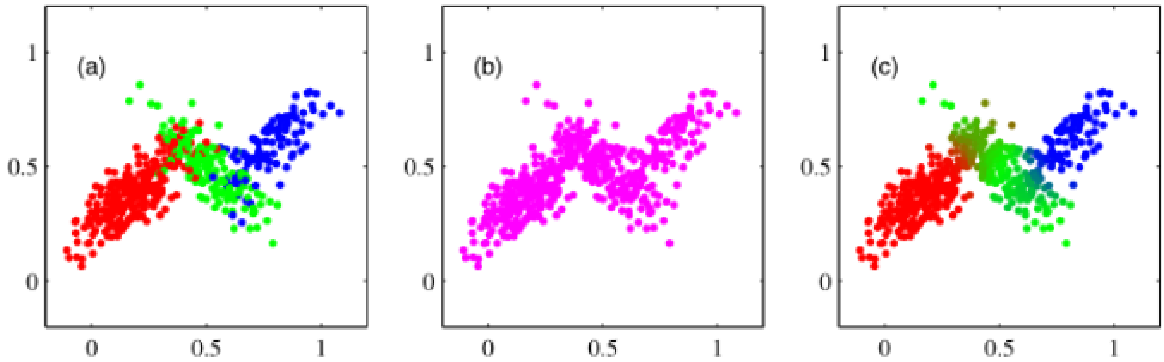
Further we obtain p(x) by summing the joint distribution over all possible states of z to give.

$$p(x) = \sum_{z} p(z)p(x|z) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)$$

Now, we can simplify even more. by using the join distribution formula

$$r(z_k) = p(z_k = 1|x) = \frac{p(z_k = 1)p(x|(z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(x|(z_j = 1)}$$
$$= \frac{\pi_k \mathcal{N}(x|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x|\mu_j, \Sigma_j)}$$

In the following image we can see 3 fazes of different Gaussian, the first one we can see that there are 3 clusters represented by 3 different collors



In conclusion, π, ɥ and ∑ are the most important variables for our Gaussian formulas. And we can set the values of these parameters by using the maximum likelihood.

$$\ln p(X|\pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k) \right\}$$

8.3 EM for Gaussian Mixtures

EM steps:

1. Initialize the means μk, the covariances Σk and mixing coefficients πk randomly. Evaluate the initial value of the log-likelihood.
2. E step :
   Evaluate the responsibilities using the current parameter values:

$$r(z_{nk}) = \frac{\pi_k \mathcal{N}(x|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x|\mu_j, \Sigma_j)}$$

3. M step :
   Re-estimate the parameters using the current responsibilities:

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} r(z_{nk}) x_n$$

$$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} r(z_{nk})(x_n - \mu_k)(x_n - \mu_k)^T$$

$$\pi_k = \frac{N_k}{N}$$

$$\text{where} \quad N_k = \sum_{n=1}^{N} r(z_{nk})$$

   Evaluate the log-likelihood:

$$\ln p(X|\pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k) \right\}$$

4. Repeat steps 2 & 3 until convergence of either the parameters or the log-likelihood. If the convergence criterion is not satisfied return to the E step.