Gaussian Mixture Models:GMM

Mixture Distribution is defined as the linear combinations of more basic probability distributions such as Gaussians, which is called superposition. It overcomes the disadvantages that single basic distribution can't model the real-world datasets (where data density is high, the probability should be high too. But simple distribution can't capture this data structure well, e.g., 'Old Faithful' data set). By using a sufficient number of Gaussians, and by adjusting their means and covariances as well as the coefficients in the linear combination, almost any continuous density can be approximated to arbitrary accuracy. Mixture models provides us a framework for building more complex probability distributions, mixture models can also be used to cluster data.

consider a superposition of K Gaussian densities:

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

Each Gaussian density $N(x|\mu_k, \sigma_k)$ is called a component of the mixture model, and the parameter π_k is called mixing coefficients. Note that both $p(x)$ and individual Gaussian component are normalized, thus $\sum_{k=1}^{K} \pi_k = 1$. And, $p(x) \geq 0$, $N(x|\mu_k, \sigma_k) \geq 0$, implies that $\pi_k \geq 0$ for all k. Thus, $\pi_k \geq 0$ and $\pi_k \leq 1$

Follow the sum and product rules, the marginal density is given by:

> $p(x) = \sum_{k=1}^{K} p(k)p(x|k)$

which is equivalent to above equation, in which we can view $p(k) = \pi_k$ as the prior paobability of picking the k^{th} component, and $N(x|\mu_k, \sigma_k) = p(x|k)$.