Gaussian Mixture Models

Rebecca C. Steorts

Gaussian Mixture Models

Assume that K mixture components, where μ_k and Σ_k are the mean and covariance matrix of the k-component.

Let $\pi_k > 0$ be mixture weights, which represents how much each component contributes to the final distribution. Note that $\sum_{k=1}^{K} \pi_k = 1$.

Then

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

is called a Gaussian mixture model.

Marginal and Joint Distributions

Consider the joint distribution

$$p(x, z) = p(z)p(x \mid z)$$

where z is a discrete random variable between 1 and K.

Let $\pi_k = P(z = k)$.

Assume the conditional distributions are Gaussian:

$$p(x \mid z = k) = N(x \mid \mu_k, \Sigma_k).$$

Then the marginal of distribution of x is

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

Parameter Estimation

The parameters of the model are $\theta = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^{K}$.

Let's consider the unrealistic case where we know the labels, z.

Define $\mathcal{D}' = \{x_n, z_n\}_{n=1}^{N}$ and $\mathcal{D} = \{x_n\}_{n=1}^{N}$ which represents the complete data and incomplete data.

How can we learn our parameters? Given $\mathcal{D}'$, the maximum likelihood estimation of θ is given by

$$\theta = \arg\max \sum_{n} \log p(x_n, z_n).$$

Parameter Estimation for Complete Data

The complete likelihood is decomposable across labels:

$$\sum_n \log p(x_n, z_n) = \sum_n \log p(z_n)p(x_n \mid z_n) \tag{1}$$

$$= \sum_k \sum_{n:z_n=k} \log p(z_n)p(x_n \mid z_n), \tag{2}$$

where we have grouped the data by the cluster labels z.

Let r_{nk} be a binary variable that indiates whether $z_n = k$.

Then it follows that

$$\sum_n \log p(x_n, z_n) = \sum_n \sum_k r_{nk} \log p(z = k)p(x_n \mid z = k) \tag{3}$$

$$= \sum_n \sum_k r_{nk} \left[\log \pi_k + \log N(x_n \mid \mu_k, \Sigma_k)\right] \tag{4}$$

$$\tag{5}$$

The MLE can be shown to be the following:

$$\pi_k = \frac{\sum_n r_{nk}}{\sum_{k'} r_{nk'}} \tag{6}$$

$$\mu_k = \frac{1}{\sum_n r_{nk}} \sum_n r_{nk} x_n \tag{7}$$

$$\Sigma_k = \frac{1}{\sum_n r_{nk}} \sum_n r_{nk}(x_n - \mu_k)(x_n - \mu_k)^T \tag{8}$$

Parameter Estimation for Incomplete Data

In this situation, we have observed and unobserved data, which is called an incomplete setting.

The observed data is $\mathcal{D} = \{x_n\}_{n=1}^N$ and the unobserved or hidden data is $\{z_n\}$.

Our goal is to find the MLE of θ where

$$\theta = \arg\max \sum_n \log P(\mathcal{D}) \tag{9}$$

$$= \arg\max \sum_n \log p(x_n) \tag{10}$$

$$= \arg\max \sum_n \log \sum_{z_n} p(x_n, z_n). \tag{11}$$

This objective function is called the incomplete log-likelihood, where there is no simple way to optimize it.

The EM algorithm provides a way to iteratively optimize this type of function.

E-step

The E-step guesses values of z_n using existing values of $\theta = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^{K}$ How does this work?

When z_n is not given, we can guess these using Bayes' rule in the following way:

$$
\begin{aligned}
p(z_n = k \mid x_n) &= \frac{p(x_n \mid z_n = k)p(z_n = k)}{p(x_n)} && (12) \\
&= \frac{p(x_n \mid z_n = k)p(z_n = k)}{\sum_{k'=1}^{K} p(x_n \mid z_n = k)p(z_n = k)} && (13) \\
&= \frac{N(x_n \mid \mu_k, \Sigma_k)\pi_k}{\sum_{k'=1}^{K} N(x_n \mid \mu_k, \Sigma_k)\pi_k} && (14)
\end{aligned}
$$

Re-define $r_{nk} = p(z_n = k \mid x_n)$. Recall previously it was binary, however, now it is a soft assignment of x_n to the kth component. So, each x_n is assigned to a component fractionally according to $p(z_n = k \mid x_n)$.

M-step

If we solve for the MLE for θ give the soft r_{nk} assignment, we get the same expressions as before. (Remember, we are cheating by using θ to compute r_{nk}.)

EM Algorithm

0. Initialize θ
1. E-step: Set $r_{nk} = p(z_n = k \mid x_n)$ with the current values of θ
2. M-step: Update θ using r_{nk} using MLE
3. Go back to step 1 until convergence.