GMM

Yuanye Liu
2023.10

Introduction

It'a a very simple and trivial derivation for GMM.

Content

Likelihood

The likelihood function in GMM for each observation x_i is,

$$P(x_i \mid \theta) = \sum_{k=1}^{K} \pi_k f_k(x_i; \mu_k, \sigma_k^2) = \sum_{k=1}^{K} \pi_k \Phi_k(x_i; \mu_k, \sigma_k^2)$$

Under the independence of pixel assumption, the likelihhod is,

$$L(\theta) = P(x \mid \theta) = \prod_i P(x_i \mid \theta) = \prod_i \sum_{k=1}^{K} \pi_k \Phi_k(x_i; \mu_k, \sigma_k^2)$$

Then log-likelihood

$$l(\theta) = \log L(\theta) = \sum_i \log(\sum_{k=1}^{K} \pi_k \Phi_k(x_i; \mu_k, \sigma_k^2))$$

we can note here we have a log-summation which is difficult to solve in MLE.

Introduce a latent variable Z

To address the problem above, we introuduce a latent variable Z, where $Z_i = k$ implies i_{th} observation belonging to k_{th} class.

$$P(x_i \mid \theta) = \sum_k P(x_i, Z_i = k \mid \theta) = \sum_k P(Z_i = k \mid \theta) P(x_i \mid Z_i = k, \theta) = \sum_k \pi_k \Phi_k(x_i; \mu_k, \sigma_k^2)$$

Prior distribution

We can find that the prior distribution of Z is,

$$P(Z_i = k \mid \theta) = \pi_k$$

Posterior distribution

And the posterior of Z is,

$$P(Z_i = k \mid x_i, \theta) = \frac{P(Z_i = k \mid \theta) P(x_i \mid z_i = k, \theta)}{\int_l P(Z_i = l \mid \theta) P(x_i \mid z_i = l, \theta)}$$

which is the result we want for segmentation.

Joint distribuiton

We compute the joint distribution

$$P(x_i, Z_i \mid \theta) = \prod_{k=1}^{K} (P(x_i, Z_i = k \mid \theta))^{1(Z_i = k)} = \prod_{k=1}^{K} (\pi_k \Phi_k(x_i; \mu_k, \sigma_k^2))^{1(Z_i = k)}$$

Complete data likelihood

The complete likelihood (including latent variables) is,

$$\begin{aligned}
L_{com}(\theta \mid x, Z) &= P(x, Z \mid \theta) \\
&= \prod_{i=1}^{N} P(x_i, Z_i \mid \theta) \\
&= \prod_{i=1}^{N} \prod_{k=1}^{K} (P(x_i, Z_i = k \mid \theta))^{1(Z_i = k)} \\
&= \prod_{k=1}^{K} (\pi_k \Phi_k(x_i; \mu_k, \sigma_k^2))^{1(Z_i = k)}
\end{aligned}$$

log-likelihood:

$$
\begin{aligned}
l_{com}(\theta \mid x, Z) &= \log L(\theta \mid x, Z) \\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} 1(Z_i = k)\{\log \pi_k + \log \Phi_k(x_i; \mu_k, \sigma_k^2)\}
\end{aligned}
$$

EM derivation

Now we derive the EM solution,

The objective is to maximize log-likelihood, we do a decomposition first,

$$\log P(x \mid \theta) = \log P(x, Z \mid \theta) - \log P(Z \mid x, \theta)$$

Take conditional expectation on both sides w.r.t. $Z \mid x, \theta^{[m]}$

$$\mathbb{E}_{Z|x,\theta^{[m]}} \log P(x \mid \theta) = \mathbb{E}_{Z|x,\theta^{[m]}} \log P(x, Z \mid \theta) - \mathbb{E}_{Z|x,\theta^{[m]}} \log P(Z \mid x, \theta)$$

We denote as

$$\log P(x \mid \theta) := l(\theta) = Q(\theta \mid \theta^{[m]}) - C(\theta \mid \theta^{[m]})$$

The update in $l(\theta)$:

$$l(\theta^{[m+1]}) - l(\theta^{[m]}) = Q(\theta^{[m+1]} \mid \theta^{[m]}) - Q(\theta^{[m]} \mid \theta^{[m]}) - \{C(\theta^{[m+1]} \mid \theta^{[m]}) - C(\theta^{[m]} \mid \theta^{[m]})\}$$

For C:

$$
\begin{aligned}
C(\theta^{[m+1]} \mid \theta^{[m]}) - C(\theta^{[m]} \mid \theta^{[m]}) &= \mathbb{E}\{\log \frac{P(Z \mid x, \theta^{[m+1]})}{P(Z \mid x, \theta^{[m]})} \mid x, \theta^{[m]}\} \\
&\leq \log \mathbb{E}\{\frac{P(Z \mid x, \theta^{[m+1]})}{P(Z \mid x, \theta^{[m]})} \mid x, \theta^{[m]}\} \\
&= \log \int_Z \frac{P(Z \mid x, \theta^{[m+1]})}{P(Z \mid x, \theta^{[m]})} P(Z \mid x, \theta^{[m]}) \mathrm{d}Z \\
&= 0
\end{aligned}
$$

So we have,

$$l(\theta^{[m+1]}) - l(\theta^{[m]}) \geq Q(\theta^{[m+1]} \mid \theta^{[m]}) - Q(\theta^{[m]} \mid \theta^{[m]})$$

So Q fucntion is a lower bound of l maximizeing $l(\theta)$ is reduced to,

$$\max_\theta Q(\theta \mid \theta^{[m]}) = \max_\theta \mathbb{E}\{\log P(x, Z \mid \theta) \mid x, \theta^{[m]}\}$$

EM solution for GMM

E-step: compute Q function

$$
\begin{aligned}
Q(\theta \mid \theta^{[m]}) &= \mathbb{E}\{\log P(x, Z \mid \theta) \mid x, \theta^{[m]}\} \\
&= \mathbb{E}\{\sum_{i=1}^{N}\sum_{k=1}^{K} 1(Z_i = k)(\log \pi_k + \log \Phi_k(x_i; \mu_k, \sigma_k^2)) \mid x, \theta^{[m]}\} \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K} \mathbb{E}\{1(Z_i = k) \mid x, \theta^{[m]}\}(\log \pi_k + \log \Phi_k(x_i; \mu_k, \sigma_k^2)) \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K} P\{Z_i = k \mid x, \theta^{[m]}\}(\log \pi_k + \log \Phi_k(x_i; \mu_k, \sigma_k^2)) \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K} \gamma_{ik}^{[m+1]}(\log \pi_k + \log \Phi_k(x_i; \mu_k, \sigma_k^2))
\end{aligned}
$$

We denote the posterior as,

$$
\gamma_{ik}^{[m+1]} = P(Z_i = k \mid x_i, \theta^{[m]}) = \frac{P(Z_i = k \mid \theta^{[m]})P(x_i \mid z_i = k, \theta^{[m]})}{\int_l P(Z_i = l \mid \theta^{[m]})P(x_i \mid z_i = l, \theta^{[m]})} = \frac{\pi_k^{[m]}\Phi(x_i; \mu_k^{[m]}, \sigma_k^{2[m]})}{\int_l \pi_l^{[m]}\Phi(x_i; \mu_l^{[m]}, \sigma_l^{2[m]})}
$$

M-Step:Maximize Q function

Simplily letting the derivative equalling to zero, and we have

π_k:

$$
\pi_k^{[m]} = \frac{1}{N}\sum_i \gamma_{ik}^{[m]}
$$

μ_k :

$$
\mu_k^{[m]} = \frac{\sum_i \gamma_{ik}^{[m]} x_i}{\sum_i \gamma_{ik}^{[m]}}
$$

σ_k^2:

$$
\sigma_k^{2[m]} = \frac{\sum_i \gamma_{ik}^{[m]}(x_i - \mu_k^{[m]})^2}{\sum_i \gamma_{ik}^{[m]}}
$$

A variational perspective of EM

Here I want to introduce a variational perspective of EM.

Similarly, we do a decomposition,

$$\begin{aligned}
l(\theta) = \log P(x \mid \theta) &= \log P(x, Z \mid \theta) - \log P(Z \mid x, \theta) \\
&= \log \frac{P(x, Z \mid \theta)}{q(z)} - \log \frac{P(Z \mid x, \theta)}{q(z)} \\
&= \sum_z q(z) \log \frac{P(x, Z \mid \theta)}{q(z)} - \sum_z q(z) \log \frac{P(Z \mid x, \theta)}{q(z)} \\
&= L(q, \theta) + KL(q(z) \| P(Z \mid x, \theta))
\end{aligned}$$

We know KL divergence is non-negative, so the $L(q, \theta)$ is a lower bound for $l(\theta)$.

E-step

Given $\theta^{[m]}$ fixed, maximize $L(q, \theta^{[m]})$ w.r.t $q(z)$.

$l(\theta)$ is not related to q, so if KL term equals zero, we can have max $L(q, \theta)$. That is to say, $q(z) = P(Z \mid x, \theta^{[m]})$, and meanwhile,

$$l(\theta^{[m]}) = L(q, \theta^{[m]})$$

M-step

Given $q(z) = P(Z \mid x, \theta^{[m]})$ fixed, maximize $L(q(z), \theta)$ w.r.t θ.

$$\begin{aligned}
L(P(Z \mid x, \theta^{[m]}), \theta) &= \sum_z P(Z \mid x, \theta^{[m]}) \log P(x, Z \mid \theta) - \sum_z P(Z \mid x, \theta^{[m]}) \log P(Z \mid x, \theta^{[m]}) \\
&= Q(\theta \mid \theta^{[m]}) + const
\end{aligned}$$

Which is the update we have proved above,

$$\theta = \arg\max_\theta Q(\theta \mid \theta^{[m]}) = \arg\max_\theta \mathbb{E}\{\log P(x, Z \mid \theta) \mid x, \theta^{[m]}\}$$

Get segmentation from GMM

The label is the latent variable value whose posterior is biggest.

$$label = \arg\max_k P(Z_i = k \mid x, \theta) \arg\max_k = \frac{P(Z_i = k \mid \theta) P(x \mid z_i = k, \theta)}{\int_l P(Z_i = l \mid \theta) P(x \mid z_i = l, \theta)} = \arg\max_k P(x, Z_i = k \mid \theta)$$

Multi-component GMM

Multi-component GMM is several level of GMM, i.e., the component of GMM is a GMM.

$$\begin{aligned}
P(I_i(x) \mid \theta) &= \sum_{k \in K} \pi_k P(I_i(x) \mid s(x) = k, \theta) \\
&= \sum_{k \in K} \pi_k \sum_{c \in C_k} \tau_{ick} P(I_i(x) \mid z_i(x) = c, s(x) = k) \\
&= \sum_{k \in K} \pi_k \sum_{c \in C_k} \tau_{ick} \Phi(I_i(x); \mu_{ikc}, \sigma_{ikc}^2)
\end{aligned}$$