Estimation of the Parameters of GMM with EM Algorithm by Minimizing Free Energy

赖泽强

2021 年 1 月 12 日

1 Introduction

回顾，在生成式模型中，我们希望求出数据的一个 parameterized distribution 最大化下面的似然概率：

$$P(X) = \prod_{i}^{m} P(x_i|\theta)$$

其中 θ 为分布的参数，总共有 m 个样本点。对上式取 log，改写成我们的目标函数：

$$L(\theta) = \sum_{i}^{m} logP(x_i|\theta)$$

2 Gaussian Mixture Model

高斯混合模型假设我们的训练数据服从一系列高斯分布，每个样本点 x 出现的概率与其属于哪个高斯分布 z 有关。具体来说，我们假设数据服从下面的联合概率分布：

$$P(x^{(i)}, z^{(i)}) = P(x^{(i)}|z^{(i)})Q(z^{(i)}) \tag{1}$$

其中 z 服从多项分布 $z^{(i)} \sim \text{Multinomial}(\phi)$，且 $\phi_j \geq 0$，$\sum_{j=1}^{k} \phi_j = 1$，$\phi_j = p(z^{(i)} = j)$。且 z 可以取 k 个值。x|z 服从正态分布，$x^{(i)} \mid (z^{(i)} = j) \sim \mathcal{N}(\mu_j, \Sigma_j)$。

换句话说，我们的模型假设 x 的生成过程是，先从 z 中取一个值出来表示这个 x 属于哪个高斯分布，然后再从对应的高斯分布生成 x。

3 Free Energies

基于高斯混合模型的假设，我们有：

$$\begin{aligned}
L(\theta) &= \sum_{i}^{m} logP(x_i|\theta) \\
&= \sum_{i}^{m} log \int P(x_i, z|\theta)dz && \text{全概率公式} \\
&= \sum_{i}^{m} log \int Q(z)\frac{P(x_i, z|\theta)}{Q(z)}dz \\
&\geq \sum_{i}^{m} \int Q(z)log\frac{P(x_i, z|\theta)}{Q(z)}dz && \text{Jensen 不等式}
\end{aligned} \tag{2}$$

由上式可知 $\sum_i^m \int Q(z)log\frac{P(x_i,z|\theta)}{Q(z)}dz$ 为 $L(\theta)$ 的一个下界，我们可以通过最大化这个下界 ELBO 来得到 $L(\theta)$ 的一个局部最优解。

令

$$\begin{aligned}F(Q,P) &= -\sum_i^m \int Q(z)log\frac{P(x_i,z|\theta)}{Q(z)}dz \\ &= \sum_i^m \int Q(z)log\frac{Q(z)}{P(x_i,z|\theta)}dz\end{aligned} \tag{3}$$

我们将 $F(Q,P)$ 称为 **free energies**，且最大化 ELBO 等价于最小化 free energies。

4 EM Algorithm

4.1 E Step

由 $P(z|x)=\frac{P(x,z)}{P(x)}$，可得

$$\begin{aligned}F(Q,P) &= \sum_i^m \int Q(z)log\frac{Q(z)}{P(x_i,z)}dz \\ &= \sum_i^m \int Q(z)log\frac{Q(z)}{P(z|x_i)P(x)}dz \\ &= \sum_i^m [D[Q(z)||P(z|x_i)] - logP(x_i)]\end{aligned}$$

在 E Step 中，我们固定 θ，优化 Q(z)。因为 $logP(x_i)=log\int P(x|z)P(z)dz$ 只与 θ 有关, 且 KL 散度 $D[Q(z)||P(z|x)]\geq 0$，仅当 $Q(z)\equiv P(z|x)$ 时取等，所以在 E Step 中，我们有：

$$E \text{ Step: } Q_{k+1} = \underset{p}{\text{argmin}}\ F[Q,P(\theta)] = P(z|x,\theta_k)$$

即当我们取 $Q(z)\equiv P(z|x)$ 时，$F(Q,P)$ 取到最小值。

4.2 M Step

在 M Step 中，我们将 E Step 估计出的 $Q(z)\equiv P(z|x)$ 回代到 $F(Q,P)$ 中：

$$\begin{aligned}F[Q_{k+1},P(\theta)] &= \sum_i^m \int Q(z)log\frac{Q(z)}{P(x_i,z)}dz \\ &= \sum_i^m \int P(z|x_i,\theta_k)log\frac{P(z|x_i,\theta_k)}{P(x_i,z|\theta)}dz \\ &= \sum_i^m \left[\int P(z|x_i,\theta_k)logP(z|x_i,\theta_k)dz - \int P(z|x_i,\theta_k)logP(x_i,z|\theta)dz\right]\end{aligned}$$

前一项使用的是上一步的 θ，因此是个定值，所以我们有：

$$M \text{ Step: } \theta_{k+1} = \underset{\theta}{\text{argmin}}\ F[Q_{k+1},P(\theta)] = \underset{\theta}{\text{argmin}}\ -\sum_i^m \int P(z|x_i,\theta_k)logP(x_i,z|\theta)dz \tag{4}$$

5 EM For GMM

现在我们将上一节推导的 EM 更新公式代入 GMM 的具体表达式，推导 GMM 的参数更新方法。

$$
\begin{aligned}
&M \text{ Step: } \theta_{k+1} = \underset{\theta}{\text{argmin}} \quad -\sum_i^m \int P(z|x_i, \theta_k) logP(x_i, z|\theta) dz \\
&E \text{ Step: } Q_{k+1} = P(z|x, \theta_k)
\end{aligned}
$$

在 GMM 中，我们假设

$$
\begin{aligned}
P(z) &\equiv \text{Multinomial}(\phi) \ \ , \sum_{j=1}^c \phi_j = 1 \\
P(x|z=j) &\equiv \mathcal{N}\left(\mu_j, \Sigma_j\right)
\end{aligned}
$$

5.1 E Step

E Step 直接使用贝叶斯公式即可

$$
Q(z=i) = P(z=i|x, \theta_k) = \frac{P(x|z=i)P(z=i)}{\sum_j^c P(x|z=j)P(z=j)}
$$

5.2 M Step

M Step 需要**最大化**:

$$
\begin{aligned}
\sum_i^m \int P(z|x_i, \theta_k) logP(x_i, z|\theta) dz &= \sum_i^m \int P(z|x_i, \theta_k) log[P(x_i|z, \mu, \Sigma)P(z|\phi)] dz \\
&= \sum_i^m (E[logP(x_i|z, \mu, \Sigma)] + E[logP(z|\phi)])
\end{aligned} \tag{5}
$$

优化 ϕ ：

ϕ 只与后一项 $E[logP(z|\phi)]$ 有关，

$$
\begin{aligned}
E[logP(z|\phi)] &= \sum_j^c P(x|z=j) logP(z=j) \\
&= \sum_j^c P(x|z=j) log\phi_j
\end{aligned}
$$

使用拉格朗日法引入约束条件，

$$
L = F(Q, P) + \lambda(\sum_{j=1}^k \phi_j - 1)
$$

求导，令导数等于零:

$$
\begin{aligned}
\frac{\partial L}{\partial \phi_j} &= -\sum_i^m \frac{P(x_i|z=j)}{\phi_j} + \lambda = 0 \\
\frac{\partial L}{\partial \lambda} &= \sum_{j=1}^c \phi_j - 1 = 0
\end{aligned}
$$

令

$$w_j^{(i)} = P(x_i|z=j)$$
$$s_j = \sum_i^m P(x_i|z=j) = \sum_i^m w_j^{(i)}$$

解方程组5.2，可得

$$\lambda = \sum_{j=1}^c s_j \phi_j = \frac{s_j}{\sum_{j=1}^c s_j}$$

因为

$$\sum_{j=1}^c s_j = \sum_{j=1}^c \sum_{i=1}^m P(x_i|z=j) = \sum_{i=1}^m \sum_{j=1}^c P(x_i|z=j) = \sum_{i=1}^m 1 = m$$

所以

$$\phi_j = \frac{1}{m} \sum_{i=1}^m w_j^{(i)} \tag{6}$$

优化 μ ：μ 只与公式5中的 $E[logP(x_i|z,\mu,\Sigma)]$ 有关：

$$\begin{aligned} E[logP(x_i|z,\mu,\Sigma)] &= \sum_j^c P(x|z=j) logP(x_i|z=j,\mu,\Sigma) \\ &= \sum_j^c w_j^{(i)} logP(x_i|z=j,\mu,\Sigma) \\ &= \sum_j^c w_j^{(i)} [log\sigma_j + \frac{1}{2}log2\pi + \frac{(x_i-\mu_j)^2}{2\sigma_j^2}] \end{aligned}$$

求导，令导数等于零：

$$\frac{\partial L}{\partial \mu_j} = \sum_{i=1}^m w_j^{(i)} \frac{(x_i-\mu_j)}{\sigma_j^2} = 0$$

解得

$$\mu_j = \frac{\sum_{i=1}^m w_j^{(i)} x_i}{\sum_{i=1}^m w_j^{(i)}} \tag{7}$$

优化 $\Sigma = \sigma_1, \sigma_2, ..., \sigma_c$ ：

同样只用考虑公式5的 $E[logP(x_i|z,\mu,\Sigma)]$：

$$\begin{aligned} E[logP(x_i|z,\mu,\Sigma)] &= \sum_j^c w_j^{(i)} [log\sigma_j + \frac{1}{2}log2\pi + \frac{(x_i-\mu_j)^2}{2\sigma_j^2}] \\ &= \sum_j^c w_j^{(i)} [log\sqrt{\sigma_j^2} + \frac{1}{2}log2\pi + \frac{(x_i-\mu_j)^2}{2\sigma_j^2}] \end{aligned}$$

求导，令导数等于零：

$$\frac{\partial F}{\partial \sigma_j^2} = -\frac{1}{2} \sum_{i=1}^m w_j^{(i)} [\frac{1}{\sigma_j^2} - \frac{(x_i-\mu_j)^2}{\sigma_j^4}] = 0$$

解得

$$\sigma_j^2 = \frac{\sum_{i=1}^m w_j^{(i)} (x_i - \mu_j)^2}{\sum_{i=1}^m w_j^{(i)}} \tag{8}$$

5.3 Summary

最终的 GMM 的 EM 更新公式如下:

$$\begin{aligned}
\phi_j &= \frac{1}{m} \sum_{i=1}^m w_j^{(i)} \\
\mu_j &= \frac{\sum_{i=1}^m w_j^{(i)} x_i}{\sum_{i=1}^m w_j^{(i)}} \\
\sigma_j^2 &= \frac{\sum_{i=1}^m w_j^{(i)} (x_i - \mu_j)^2}{\sum_{i=1}^m w_j^{(i)}}
\end{aligned} \tag{9}$$