高斯混合模型的 EM 算法推导

赖泽强

2021 年 1 月 12 日

1　高斯混合模型

高斯混合模型尝试用多个高斯分布对输入数据建模。我们假设模型符合一个联合概率分布

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

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。

2　参数估计

我们的模数包含三个参数，首先是多项分布的参数 ϕ，然后是正态分布的均值 μ 和方差 σ^2。(只考虑 x 是一维，x|z 符合一维正态分布的情况)

估计这三个参数，我们采用最大化公式2所示的似然概率的方法。

$$\begin{aligned}
\ell(\phi, \mu, \sigma) &= \sum_{i=1}^{m} \log p\left(x^{(i)}; \phi, \mu, \sigma\right) \\
&= \sum_{i=1}^{m} \log \sum_{z^{(i)}=1}^{k} p\left(x^{(i)} \mid z^{(i)}; \mu, \sigma\right) p\left(z^{(i)}; \phi\right)
\end{aligned} \tag{2}$$

如果 z 的类别已知的话，我们可以去掉公式2里面的求和符号，因此我们只需要最大化3所示的似然概率即可。

$$\ell(\phi, \mu, \sigma) = \sum_{i=1}^{m} \log p\left(x^{(i)} \mid z^{(i)}; \mu, \sigma\right) + \log p\left(z^{(i)}; \phi\right) \tag{3}$$

我们要求似然函数的最大值点，分别对 ϕ, μ, σ^2 进行求导。

求解 $\frac{\partial \ell}{\partial \phi_j}$　：约束条件 $\sum_{j=1}^{k} \phi_j = 1$。

使用拉格朗日乘数法求解条件极值，令

$$F(\phi, \lambda) = \ell(\phi, \mu, \sigma) + \lambda(\sum_{j=1}^{k} \phi_j - 1)$$

$$s_j = \sum_{i=1}^{m} 1\left\{z^{(i)} = j\right\}$$

分别对 ϕ_j 以及 λ 进行求导，可得:

$$\begin{cases} \frac{\partial F}{\partial \phi_j} = \frac{s_j}{\phi_j} + \lambda \\ \frac{\partial F}{\partial \lambda} = \sum_{j=1}^{k} \phi_j - 1 \end{cases}$$

令导数等于零，联立方程组求解，可得

$$\lambda = -\sum_{j=1}^{k} s_j$$
$$\phi_j = \frac{1}{m}\sum_{i=1}^{m} 1\left\{z^{(i)} = j\right\} \tag{4}$$

求解 $\frac{\partial \ell}{\partial \mu_j}$ ：似然函数只有前一项与 μ 有关，只考虑前一项即可。

将正态分布的公式 $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ 带入似然函数可得：

$$\ell(\phi, \mu, \sigma) \approx \sum_{i=1}^{m} log(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}) = -\sum_{i=1}^{m}[log\sigma + \frac{1}{2}log2\pi + \frac{(x_i-\mu)^2}{2\sigma^2}]$$

求导令其导数为零：

$$\frac{\partial \ell}{\partial \mu_j} = \sum_{i=1}^{m} 1(z^{(i)} = j) \cdot 2(x_i - \mu_j) = 0$$

可得：

$$\mu_j = \frac{\sum_{i=1}^{m} 1\left\{z^{(i)} = j\right\} x^{(i)}}{\sum_{i=1}^{m} 1\left\{z^{(i)} = j\right\}} \tag{5}$$

求解 $\frac{\partial \ell}{\partial \sigma_j^2}$ ：同样只用考虑似然函数的前一项。

$$\ell(\phi, \mu, \sigma) \approx -\sum_{i=1}^{m}[log\sigma + \frac{1}{2}log2\pi + \frac{(x_i-\mu)^2}{2\sigma^2}]$$

求导令其导数为 0

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{1}{2}\sum_{i=1}^{m} 1(z^{(i)} = j)[\frac{1}{\sigma_j^2} - \frac{(x_i-\mu_j)^2}{\sigma_j^4}] = 0$$

可得

$$\sigma_j^2 = \frac{\sum_{i=1}^{m} 1(z^{(i)} = j)(x_i - \mu_j)^2}{\sum_{i=1}^{m} 1(z^{(i)} = j)} \tag{6}$$

3 EM 算法

在上一小节，我们固定 z，求解出了模型参数的最优解。在 EM 算法中，我们采用迭代的方法进行优化。

在 EM 算法的 **Expectation** 阶段，我们首先使用贝叶斯公式求解出当前参数下，x 属于某个高斯分布的概率。

$$p\left(z^{(i)} = j \mid x^{(i)}; \phi, \mu, \Sigma\right) = \frac{p\left(x^{(i)} \mid z^{(i)} = j; \mu, \Sigma\right) p\left(z^{(i)} = j; \phi\right)}{\sum_{l=1}^{k} p\left(x^{(i)} \mid z^{(i)} = l; \mu, \Sigma\right) p\left(z^{(i)} = l; \phi\right)} \tag{7}$$

令

$$w_j^{(i)} := p\left(z^{(i)} = j \mid x^{(i)}; \phi, \mu, \Sigma\right)$$

在 **Maximization** 阶段，我们期望代替之前的 indictor 函数, 并使用上一节推导出来的公式更新最优参数值。

$$\begin{aligned}
\phi_j &:= \frac{1}{m}\sum_{i=1}^{m} w_j^{(i)} \\
\mu_j &:= \frac{\sum_{i=1}^{m} w_j^{(i)} x^{(i)}}{\sum_{i=1}^{m} w_j^{(i)}} \\
\sigma_j^2 &:= \frac{\sum_{i=1}^{m} w_j^{(i)} (x_i - \mu_j)^2}{\sum_{i=1}^{m} w_j^{(i)}}
\end{aligned} \tag{8}$$

如此迭代进行，直至收敛。