EM 算法

W.J.Z

EM 算法是含有因变量的概率模型参数的极大似然估计方法，每代迭代由两步组成：E 步，求期望；M 步, 求极大。本文只讲述 EM 算法和在高斯函数上的应用, 算法的推导和收敛性证明请查阅相关资料。

1 EM 算法

一般用 Y 表示观测随机变量的数据，Z 表示隐随机变量的数据，Y 和 Z 连在一起成为完全数据，感测数据 Y 又称为不完全数据。假设给定观测数据 Y, 其概率分布式 $P(Y|\theta)$, 其 θ 是需要估计的模型参数，那不完全数据 Y 的似然函数 $P(Y|\theta)$. 对数似然函数 $L(\theta) = \log P(Y|\theta)$. 假设 Y 和 Z 的联合概率分布为 $P(Y, Z|\theta)$，那么完全数据的对数似然函数是 $\log P(Y, Z|\theta)$。

Algorithm 1: EM 算法

Input: 观测变量数据 Y，隐变量数据 Z，联合分布 $P(Y, Z|\theta)$, 条件分布 $P(Z|Y, \theta)$

Output: 模型参数 θ

选择参数的初值 θ^0, 开始迭代。

E 步：记 θ^i 为第 i 次迭代参数 θ 的估计值，在第 $i+1$ 次迭代的 E 步，计算

$$Q\left(\theta\ \theta^i\right) = \sum_Z \log P(Y, Z|\theta)\, P\left(Z|Y, \theta^i\right)$$

$P(Z|Y, \theta^i)$ 是在给定观测数据 Y 和当前的参数估计 θ^i 下隐变量数据 Z 的条件概率分布。

M 步: 求使 $Q(\theta, \theta^i) = \arg max_\theta Q(\theta, \theta^i)$

重复第二步和第三步，直到收敛。

---

2 EM 算法在高斯混合模型学习中的应用

高斯混合模型指具有以下形式的概率分布模型：

$$P\left(y|\theta\right)=\sum_{k=1}^{K}\alpha_k\phi\left(y|\theta_k\right)$$

其中,α_k 是系数,$\alpha_k\geq 0$,$\sum_{k=1}^{K}\alpha_k=1$;$\phi\left(\theta|\theta_k\right)$ 是高斯分布密度,$\theta_k=\left(\mu_k,\sigma_k^2\right)$

$$\phi\left(y|\theta_k\right)=\frac{1}{\sqrt{2pi}\sigma_k}exp\left(-\frac{\left(y-\mu_k\right)^2}{2\sigma_k^2}\right)$$

观测数据 $y_j,j=1,2,\ldots,N$ 产生过程为：首先依概率 α_k 选择第 k 个高斯分布模型 $\phi\left(y|\theta_k\right)$；然后依第 K 个分模型的概率分布 $\phi\left(y|\theta_k\right)$ 生成观测数据 y_j。

$$\gamma_{jk}=\begin{cases}1, & \text{第 j 个观测数据来自第 K 个分模型}\\ 0 & \text{否则}\end{cases}\quad\gamma_{jk}\ 0-1$$

完全数据的似然函数为

$$\begin{aligned}Py,\gamma|\theta&=\prod_{k=1}^{K}\prod_{j=1}^{N}\left[\alpha_k\phi\left(y_j|\theta_k\right)\right]^{\gamma_{jk}}\\&=\prod_{k=1}^{K}\alpha_k^{n_k}\prod_{j=1}^{N}\left[\phi\left(y_j|\theta_k\right)\right]^{\gamma_{jk}}\\&=\prod_{k=1}^{K}\alpha_k^{n_k}\prod_{j=1}^{N}\left[\phi\left(y|\theta_k\right)=\frac{1}{\sqrt{2pi}\sigma_k}exp\left(-\frac{\left(y-\mu_k\right)^2}{2\sigma_k^2}\right)]^{\gamma_{jk}}\right]^{\gamma_{jk}}\end{aligned}$$

其中 $n_k=\sum_{j=1}^{N}\gamma_{jk},\sum_{k=1}^{K}n_k=N$。完全数据的对数似然函数为

$$\log P\left(y,\gamma|\theta\right)=\sum_{k=1}^{K}\left\{n_k\log\alpha_k+\sum_{j=1}^{N}\gamma_{jk}\left[\log\left(\frac{1}{\sqrt{2pi}}\right)-\log\sigma_k-\frac{1}{2\sigma_k^2}\left(y_j-\mu_k\right)^2\right]\right\}$$

确定 Q 函数：

$$\begin{aligned}P\left(\gamma_{jk}|y,\theta\right)&=P\left(\gamma_{jk}=1|y,\theta\right)\\&=\frac{P\left(\gamma_{jk}=1\ y_j|\theta\right)}{\sum_{k=1}^{K}P\left(\gamma_{jk}=1\ y_j|\theta\right)}\\&=\frac{\alpha_k\phi\left(y_j|\theta_k\right)}{\sum_{k=1}^{K}\alpha_k\phi\left(y_j|\theta_k\right)}\end{aligned}$$

Q 函数为：

$$Q\left(\theta, \theta^{i}\right)=\sum_{Z} \log P(y, \gamma | \theta) P\left(\gamma_{jk} | y, \theta\right)$$

求函数 $Q(\theta, \theta^{i})$ 对 θ 的极大值，分别对 μ_k,σ_k^2 求偏导并令其为 0 即可得到解；在 $\sum_{k=1}^{K} \alpha_k=1$ 条件下求偏导并令其为 0 得到其解。

$$\mu_k=\frac{\sum_{j=1}^{N} P\left(\gamma_{jk} | y, \theta\right) \gamma_{jk} y_j}{\sum_{j=1}^{N} P\left(\gamma_{jk} | y, \theta\right) y_{jk}}$$

$$\sigma_k^2=\frac{\sum_{j=1}^{N} P\left(\gamma_{jk} | y, \theta\right) \gamma_{jk}\left(y_j-\mu_k\right)^2}{\sum_{j=1}^{N} P\left(\gamma_{jk} | y, \theta\right) \gamma_{jk}}$$

$$\alpha_k=\frac{\sum_{j=1}^{N} P\left(\gamma_{jk} | y, \theta\right) \gamma_{jk}}{N}$$

Algorithm 2: 高斯混合模型参数估计的 EM 算法

Input: 观测数据，高斯混合模型

Output: 高斯混合模型参数

1 取参数的初始值开始迭代

2 E 步: 计算分模型 k 对观测数据 y_j 响应度

$$\frac{\alpha_k \phi\left(y_j | \theta_k\right)}{\sum_{k=1}^{K} \alpha_k \phi\left(y_j | \theta_k\right)}$$

3 M 步：

$$\mu_k=\frac{\sum_{j=1}^{N} P\left(\gamma_{jk} | y, \theta\right) \gamma_{jk} y_j}{\sum_{j=1}^{N} P\left(\gamma_{jk} | y, \theta\right) y_{jk}}$$

$$\sigma_k^2=\frac{\sum_{j=1}^{N} P\left(\gamma_{jk} | y, \theta\right) \gamma_{jk}\left(y_j-\mu_k\right)^2}{\sum_{j=1}^{N} P\left(\gamma_{jk} | y, \theta\right) \gamma_{jk}}$$

$$\alpha_k=\frac{\sum_{j=1}^{N} P\left(\gamma_{jk} | y, \theta\right) \gamma_{jk}}{N}$$

4 重复 2 和 3 步，直至收敛。