EM算法

1. EM算法基本思想

最大期望算法（Expectation-Maximization algorithm, EM），是一类通过迭代进行极大似然估计的优化算法，通常作为牛顿迭代法的替代，用于对包含隐变量或缺失数据的概率模型进行参数估计。

最大期望算法基本思想是经过两个步骤交替进行计算：

　　第一步是计算期望（E），利用对隐藏变量的现有估计值，计算其最大似然估计值；

　　第二步是最大化（M），最大化在E步上求得的最大似然值来计算参数的值。

M步上找到的参数估计值被用于下一个E步计算中，这个过程不断交替进行。

2. EM算法推导

　对于m个样本观察数据$x=(x^{1},x^{2},...,x^{m})$，现在想找出样本的模型参数$\theta$，其极大化模型分布的对数似然函数为：

$$\theta = \arg \max *\theta \sum *i=1^m logP(x^{(i)};\theta)m$$

如果得到的观察数据有未观察到的隐含数据$z=(z^{(1)},z^{(2)},...z^{(m)})$，极大化模型分布的对数似然函数则为：

$$\theta = \arg \max *\theta \sum *i=1^m logP(x^{(i)};\theta) = \arg \max *\theta \sum *i=1^m log\sum_{z^{(i)}}P(x^{(i)}, z^{(i)};\theta) \quad \text{(a)}$$

由于上式不能直接求出θ，采用缩放技巧：

$$\begin{aligned} \sum_{i=1}^m log\sum_{z^{(i)}}P(x^{(i)}, z^{(i)};\theta) & = \sum_{i=1}^m log\sum_{z^{(i)}}Q_i(z^{(i)})\frac{P(x^{(i)}, z^{(i)};\theta)}{Q_i(z^{(i)})} \\ & \geqslant \sum_{i=1}^m \sum_{z^{(i)}}Q_i(z^{(i)})log\frac{P(x^{(i)}, z^{(i)};\theta)}{Q_i(z^{(i)})} \end{aligned} \tag{1}$$

上式用到了Jensen不等式：

$$log\sum_j\lambda_jy_j \geqslant \sum_j\lambda_jlogy_j;;,\lambda_j \geqslant 0, \sum_j\lambda_j =1$$

并且引入了一个未知的新分布$Q_i(z^{(i)})$。

此时，如果需要满足Jensen不等式中的等号，所以有：

$$\frac{P(x^{(i)},z^{(i)};\theta)}{Q_i(z^{(i)})}=c, c\text{为常数}$$

由于$Q_i(z^{(i)})$是一个分布，所以满足

$$\sum_z Q_i(z^{(i)})=1$$

综上，可得：

$$Q_i(z^{(i)})=\frac{P(x^{(i)}，z^{(i)};\theta)}{\sum\limits_z P(x^{(i)},z^{(i)};\theta)}=\frac{P(x^{(i)},z^{(i)};\theta)}{P(x^{(i)};\theta)}=P(z^{(i)}|x^{(i)};\theta)$$

如果$Q_i(z^{(i)})=P(z^{(i)}|x^{(i)};\theta)$，则第(1)式是我们的包含隐藏数据的对数似然的一个下界。如果我们能极大化这个下界，则也在尝试极大化我们的对数似然。即我们需要最大化下式：

$$\arg\max *\theta \sum *i=1^m \sum_{z^{(i)}} Q_i(z^{(i)}) log\frac{P(x^{(i)}，z^{(i)};\theta)}{Q_i(z^{(i)})}$$

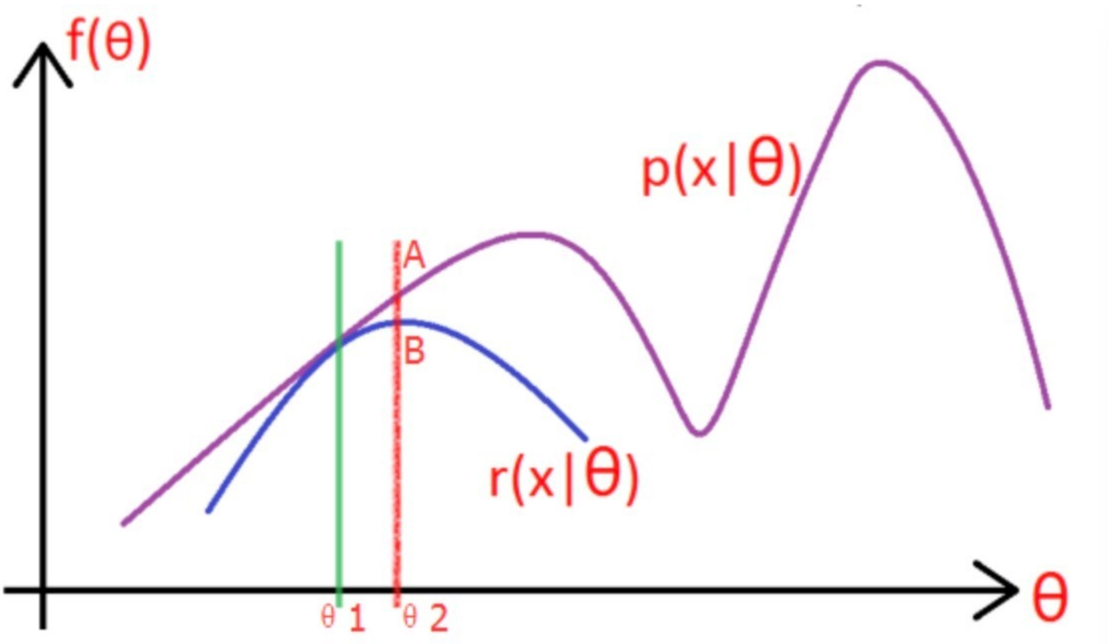
简化得：

$$\arg\max *\theta \sum *i=1^m \sum_{z^{(i)}} Q_i(z^{(i)}) logP(x^{(i)},z^{(i)};\theta)$$

以上即为EM算法的M步，$\sum\limits_{z^{(i)}} Q_i(z^{(i)}) logP(x^{(i)},z^{(i)};\theta)$可理解为$logP(x^{(i)},z^{(i)};\theta)$基于条件概率分布$Q_i(z^{(i)})$的期望。以上即为EM算法中E步和M步的具体数学含义。

3. EM算法原理

考虑式（a），表达式中存在隐变量，通过EM算法迭代求解下界的最大值到收敛为止。



图片中的紫色部分是我们的目标模型$p(x|\theta)$，该模型复杂，难以求解析解，为了消除隐变量$z^{(i)}$的影响，我们可以选择一个不包含$z^{(i)}$的模型$r(x|\theta)$，使其满足条件$r(x|\theta)\leqslant p(x|\theta)$。

求解步骤如下：

（1）选取θ_1，使得$r(x|\theta_1)=p(x|\theta_1)$，然后对此时的$r$求取最大值，得到极值点$\theta_2$，实现参数的更新。

（2）重复以上过程到收敛为止，在更新过程中始终满足$r\leqslant p$.

4.EM算法流程

输入：观察数据$x=(x^{(1)},x^{(2)},...x^{(m)})$，联合分布$p(x,z;\theta)$，条件分布$p(z|x;\theta)$，最大迭代次数$J$

1）随机初始化模型参数θ的初值θ^0。

2）$for\ j\ from\ 1\ to\ j$:

a） E步。计算联合分布的条件概率期望： $Q_i(z^{(i)})=P(z^{(i)}|x^{(i)},\theta^j)$

$$L(\theta,\theta^j)=\sum_{i=1}^{m}\sum_{z^{(i)}}P(z^{(i)}|x^{(i)},\theta^j)logP(x^{(i)},z^{(i)};\theta)$$

b） M步。极大化$L(\theta,\theta^j)$，得到θ^{j+1}: $\theta^{j+1}=\arg\max_\theta L(\theta,\theta^j)$ c） 如果θ^{j+1}收敛，则算法 结束。否则继续回到步骤a）进行E步迭代。

输出：模型参数θ。