

EM

EM算法 (Expectation-maximization) , 又称最大期望算法, 是一种迭代算法, 用于含有隐变量的概率模型**参数**的极大似然估计 (或极大后验概率估计)

从定义可知, 该算法是用来估计参数的, 这里约定参数为 θ 。既然是迭代算法, 那么肯定有一个初始值, 记为 $\theta^{(0)}$, 然后再通过算法计算 $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(t)}$

通常, 当模型的变量都是观测变量时, 可以直接通过极大似然估计法, 或者贝叶斯估计法估计模型参数。但是当模型包含隐变量时, 就不能简单的使用这些估计方法

举个具体的栗子:

K-means算法中, 除了给定的样本 (也就是观测变量) X 以及参数 θ (也就是那些个聚类的中心) 之外, 还包含一个隐变量 (记为 Z) , 它是每个样本的所属类别。

可以理解为, 我们之所以对一批样本进行聚类, 也是因为认为这些样本是有它们潜在的类别的, 也就是说还有一个隐变量是我们没有 (或者无法) 观测到的

下面先给出EM算法的步骤公式, 然后再对公式进行推导。假设在第 i 次迭代后参数的估计值为 $\theta^{(i)}$, 对于第 $i+1$ 次迭代, 分为两步

- E步, 求期望:

$$\begin{aligned} Q(\theta, \theta^{(i)}) &= \sum_Z P(Z|X, \theta^{(i)}) \log P(X, Z|\theta) \\ &= \mathbb{E}_{Z|X, \theta^{(i)}} [\log P(X, Z|\theta)] \end{aligned}$$

关于的随机变量的函数的期望, 公式在后面会给出

- M步, 最大化:

$$\theta^{(i+1)} = \arg \max_{\theta} Q(\theta, \theta^{(i)})$$

其中, $Q(\theta, \theta^{(i)})$ 称为 Q 函数, 是EM算法的核心。下面就来对公式进行推导

给定一组观测数据记为 $X = (x_1, x_2, \dots, x_n)$, 以及参数 θ 。因为 $x_1, x_2, \dots, x_n$ 是独立同分布, 所以有以下对数似然函数:

$$\begin{aligned}
 \ell(\theta|X) &= \log P(X|\theta) \\
 &= \log \left(\prod_{i=1}^n P(x_i|\theta) \right) \\
 &= \sum_{i=1}^n \log P(x_i|\theta)
 \end{aligned}$$

可以通过极大似然估计来求解最优参数，即：

$$\begin{aligned}
 \hat{\theta} &= \arg \max_{\theta} \log \ell(\theta|X) \\
 &= \arg \max_{\theta} \sum_{i=1}^n \log P(x_i|\theta)
 \end{aligned}$$

但是由于隐变量的存在， $\log P(X|\theta)$ 变为

$$\begin{aligned}
 \log P(X|\theta) &= \log \sum_Z P(X, Z|\theta) \\
 &= \log \left(\sum_Z P(X|\theta, Z)P(Z|\theta) \right)
 \end{aligned}$$

这样直接求解就变得困难，一个办法是构造一个容易优化的——关于对数似然函数的——下界函数，通过不断的优化这个下界，迭代逼近最优参数。为了方便下面推导流畅，提前先贴几个公式

随机变量的数学期望

$$\mathbb{E}[X] = \sum_{x \in X} xP(x)$$

随机变量函数的数学期望。设 $Y = g(X)$ ，则 Y 的期望为：

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \sum_{x \in X} g(x)P(x)$$

相对熵

$$KL(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$$

下面是具体的推导。首先引入隐变量 Z 的概率分布 $q(Z)$ ，满足

$$\sum_Z q(Z) = 1$$

并且以下等式成立

$$P(X|\theta) = \frac{P(X, Z|\theta)}{P(Z|X, \theta)} = \frac{P(X, Z|\theta)/q(Z)}{P(Z|X, \theta)/q(Z)}$$

两边同时取对数

$$\log P(X|\theta) = \log \frac{P(X, Z|\theta)/q(Z)}{P(Z|X, \theta)/q(Z)}$$

同时求两边在 Z 上的期望

$$\mathbb{E}_Z[\log P(X|\theta)] = \mathbb{E}_Z\left[\log \frac{P(X, Z|\theta)/q(Z)}{P(Z|X, \theta)/q(Z)}\right]$$

因为 $\log P(X|\theta)$ 与 Z 无关，所以求期望仍然不变：

$$\begin{aligned}\mathbb{E}_Z[\log P(X|\theta)] &= \sum_Z q(Z) \log P(X|\theta) \\ &= \log P(X|\theta) \sum_Z q(Z) \\ &= \log P(X|\theta)\end{aligned}$$

然后将右边展开

$$\begin{aligned}\mathbb{E}_Z\left[\log \frac{P(X, Z|\theta)/q(Z)}{P(Z|X, \theta)/q(Z)}\right] &= \sum_Z q(Z) \log \frac{P(X, Z|\theta)/q(Z)}{P(Z|X, \theta)/q(Z)} \\ &= \sum_Z q(Z) \log \frac{P(X, Z|\theta)}{q(Z)} - \sum_Z q(Z) \log \frac{P(Z|X, \theta)}{q(Z)} \\ &= \sum_Z q(Z) \log \frac{P(X, Z|\theta)}{q(Z)} + \sum_Z q(Z) \log \frac{q(Z)}{P(Z|X, \theta)} \\ &= \sum_Z q(Z) \log \frac{P(X, Z|\theta)}{q(Z)} + KL(q(Z)||P(Z|X, \theta)) \\ &\geq \sum_Z q(Z) \log \frac{P(X, Z|\theta)}{q(Z)}\end{aligned}$$

由此得到对数似然函数的下界。并且当 $KL(q(Z)||P(Z|X, \theta)) = 0$ ，上式可以取到等号，由相对熵的性质可知，相对熵为0，也就是 $q(Z) = P(Z|X, \theta)$

其中 $q(Z)$ 是 Z 的概率分布，但是因为无法观测 Z ，所以 $q(Z)$ 未知，可以假设其等于 $P(Z|X, \theta)$ ，也就是 Z 关于给定 X 与 θ 的后验，且 θ 是由初始值 $\theta^{(0)}$ 一次次迭代计算而来，所以此处的 θ 是迭代 i 次后的值

$$P(Z|X, \theta) = P(Z|X, \theta^{(i)})$$

然后通过极大似然估计得到：

$$\begin{aligned}
\hat{\theta} &= \arg \max_{\theta} \sum_Z q(Z) \log \frac{P(X, Z|\theta)}{q(Z)} \\
&= \arg \max_{\theta} \sum_Z P(Z|X, \theta^{(i)}) \log \frac{P(X, Z|\theta)}{P(Z|X, \theta^{(i)})} \\
&= \arg \max_{\theta} \sum_Z P(Z|X, \theta^{(i)}) \left(\log P(X, Z|\theta) - \log P(Z|X, \theta^{(i)}) \right) \\
&= \arg \max_{\theta} \sum_Z P(Z|X, \theta^{(i)}) \log P(X, Z|\theta) \\
&= \arg \max_{\theta} \mathbb{E}_{Z|X, \theta^{(i)}} [\log P(X, Z|\theta)]
\end{aligned}$$

以上，就是EM算法中E步的由来，然后令 $\hat{\theta} = \theta^{(i+1)}$ ，就得到了M步的公式

$$\theta^{(i+1)} = \arg \max_{\theta} \mathbb{E}_{Z|X, \theta^{(i)}} [\log P(X, Z|\theta)]$$

以上就是EM算法的推导过程，为了加深理解，我们可以换一个角度来总结一下。前面我们定义了似然函数

$$\ell(\theta|X) = \log P(X|\theta) = \log \left(\sum_Z P(X|\theta, Z)P(Z|\theta) \right)$$

由于累加号嵌套在 $\log$ 函数中，难以直接进行求解，如果换一个似然函数，就容易的多

$$\ell(\theta|X, Z) = \log P(X, Z|\theta)$$

但是，又由于的 Z 是隐变量，无法得到它的概率分布，只能通过给定的 X 和 θ 来计算它的后验分布，然后求似然函数在此分布上的期望

$$\mathbb{E}_{Z|X, \theta^{(i)}} [\log P(X, Z|\theta)] = \mathbb{E}_{Z|X, \theta^{(i)}} [\ell(\theta|X, Z)]$$

最后，再寻找能使似然函数的期望最大化的参数。