一、为什么需要EM算法

概率模型有时既含有观测变量，又含有隐变量或者潜在变量。如果概率模型的变量都是观测变量，那么给定数据，可以直接用极大似然估计，或者贝叶斯估计法估计模型参数。但是，当模型含有隐变量时，就不能简单地使用这些估计方法。EM算法就是含有隐变量的概率模型参数的极大似然估计法。

EM算法的每次迭代由两步组成：E步，求期望；M步，求极大。所以算法称为期望极大算法，简称EM算法。

二、EM算法的例子

《统计学习方法》例子9.1（三硬币模型）：

假设又三枚硬币，分别记作A，B，C。这些硬币正面出现的概率分别是π，p和q。进行如下掷硬币实验：先掷硬币A，根据其结果选出硬币B或硬币C，正面选硬币B，反面选硬币C；然后掷选出的硬币，掷硬币的结果，出现正面记作1，出现反面记作0；独立地重复n次实验（这里，n=10），观察结果如下：

1，1，0，1，0，0，1，0，1，1

假设只能观测到掷硬币的结果，不能观测掷硬币的过程。问如何估计三枚硬币正面出现的概率，即三枚硬币模型的参数

解：对每一次实验可以进行如下建模

前提须知公式：

求解公式：

其中，随机变量y是观测变量，表示一次实验观测的结果是1或0；随机变量Z是隐变量，表示未观测到的掷硬币A的结果；是模型参数。

将观测数据表示为,未观测数据表示为则观测数据的似然函数为：

考虑求模型参数的极大似然估计，即使用对数似然函数来进行参数估计可得：

三、EM算法的导出

3.1 Jensen（琴生不等式）

若f是凸函数，则：

其中，。同理，若f是凹函数，则只需将上式中的换成即可。

将上式中的t推广到n个同样成立，也即：

其中，，。在概率论中常以以下形式出现：

其中X是随机变量，是凸函数，表示X的期望。

3.2 EM算法的推导

我们面对一个含有隐变量的概率模型，目标是极大化观测数据Y关于参数的对数似然函数，即极大化：

注意到这一极大化的主要困难是上式中有未观测数据Z并有包含和（Z为离散型时）或者积分（Z为连续型时）的对数。EM算法采用的是通过迭代逐步近似极大化：假设在第次迭代后的估计值是，我们希望新的估计值能使增加，即

并逐步达到极大值。为此，我们考虑两者的差：

套用琴生不等式可得：

令

即函数是的一个下界，此时若设使得达到极大，也即

因此，任何可以使增大的也可以使增大，由于问题转化为了求解能使得达到极大的，即

四、EM算法求解例子

使用EM算法求解三硬币模型