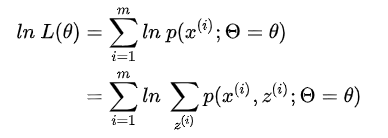
# EM算法报告

时间：2020年3月8日星期天

## 一 问题列表

许茜

1. 为什么不能直接对参数求导令其等于0进行求解？EM算法是如何求解的？



由于上式在对数中还套了求和，因此导数（梯度）的形式将会非常复杂，不利于求解。EM算法要利用Jensen不等式构造出一个下界，极大化下界求得局部最优解。

1. EM算法的流程

（1）选择一个初始值

（2）选择在与似然函数相切的函数

（3）求出（2）中函数的极大值，将更新为，重复上述过程直至收敛。

3. EM算法给出的局部最优解，这个“局部”是哪个局部？什么情况下局部最优解可以是全局最优解？

求出的是哪个局部最优解，与函数的形状和给定的初值有关。在样本规模不太大的情况下，可以对参数空间进行网格搜索，选取最优的那个解。当函数只有一个极大值点，是一个凸优化问题时，局部最优是全局最优。

王鹏

总体理解：

EM算法是一种迭代算法，它主要是引入了隐变量，来解决含隐变量的概率模型参数的极大似然估计，或极大后验估计。EM一共包含两个步骤，E布，求期望（expectation）；M步，求极大（maximization）。

1. EM算法的一般步骤：

写出似然函数；

对似然函数取对数，并整理；

求导数，令导数为0，得到似然方程；

解似然r方程，得到的参数即为所求；

1. 请说明Q函数是什么？

Q函数是完全数据的对数似然函数关于在给定的观测数据Y和当前的参数下，对未观测数据Z的条件概率分布的期望称为Q函数。

1. EM算法如何与无监督学习联系？

这个问题很好，首先EM算法中引入了隐变量的概念，我们可以把隐变量当作我们的训练数据中的“标签值”，观测数据当作样本值，然后通过EM的方法求其联合概率分布。

秦旭

1、 EM算法是一种局部下限构造么？

是的，因为EM算法在最大化观测数据Y关于参数的对数似然函数的过程中，能够求解通过之前所求的参数变量计算得到的下界，不断优化这个下界，就能不断提升最大似然函数值。故从该角度看，EM算法是一种局部下限构造的方法。

2、 K-means跟EM算法有关么？

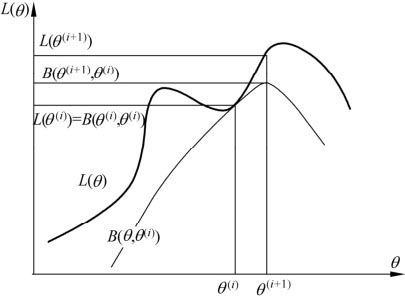
聚类问题是给定数据点X1-XN，给定分类数目K，在各类内点至其中心点的平方距离和最小的约束条件下，求K个类的中心。

问题中其实就包含着隐变量，对于任意一个点i，问题就是确定其归属于哪个类，等同于判断其关于K个类的隐变量，其中哪一个隐变量的取值为1，其余值为0。

如果切换到EM算法中，E步可以转换求距离平方和倒数。M步可以转化为求解当前类别内距离损失最小的点。

3、 EM算法一定能找到全局最优么？

不一定，EM算法在不停的优化下界，可能会陷入局部最优。李航书中的图就可以作为举例。



4、 EM算法的一般步骤？

１.　参数进行初始化；

２.　E步：利用已知参数，计算得到下一步要优化的最大似然函数；

３.　M步：找到能够使最大似然函数极大化的参数，可以通过求导置零取得；

４.　不断重复２，３步骤，当变化量小于最小控制值时停止。

5、 请说明Q函数是什么？

Q函数是关于当前步参数和观测结果对未知数据的条件概率分布。

6. EM算法怎么和无监督学习联合在一起？

通过隐变量可以得到非直接观测数据的标签，类似于数据中没有直接进行标注的部分。使其最大似然概率最大，从思想上与无监督学习类似。

李志强

1. EM算法适用于样本中含有隐变量

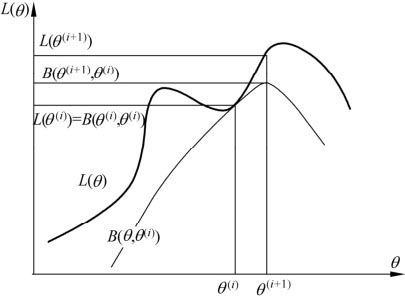
对于一个函数f(x;θ)，里面x是已知的，但θ未知，于是使用极大似然估计求θ，但如果x是由(x, z)两部分组成，其中x已知z未知，也就是说样本中含有隐变量z，对于这种情况极大似然估计就没办法解决了，而为了解决这样的情况而发明的算法之一就是EM算法。

1. EM算法的缺陷

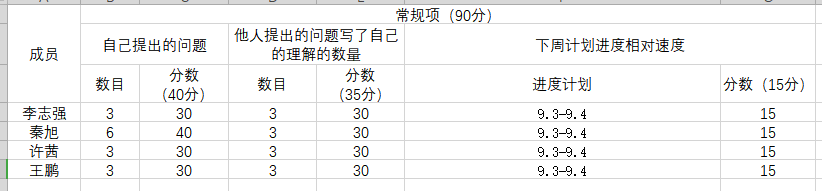
只能得到局部极值点，不能得到全局极值点。

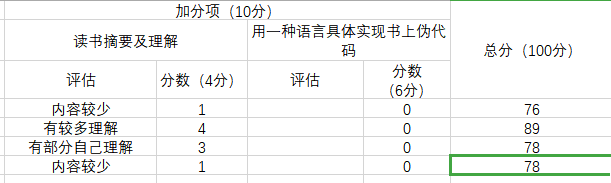
1. 为什么EM算法采用迭代的方式进行？

因为在对观测变量的似然函数进行极大似然估计时，由于隐变量的存在无法直接求得解析解，所以EM考虑通过迭代的方式，不断极大化该似然函数的下界，最终达到收敛。



## 二 各成员读书情况汇总





## 三 读书会情况总结

在讨论过程中，对于同一个问题，每个人都有自己的理解，可能每个人的想法都有偏差，但是最终的结果是得到了一个公认的解释，比如这次讨论中关于EM迭代时优化的过程图的理解。

大家的进度都保持一致，讨论效果理想。