方骏-2020年5月3日-读书报告

1. 自己提出的问题的理解：
2. 提出的问题1：后面的GEM算法2和算法3一定可以有和算法1一样的效果吗？

讨论后的理解：差不多。因为后面的GEM算法其实就是一个前面的EM算法的一个改版，本质上并没有什么特别大的区别，F函数和EM算法的目标都是一样的，而且收敛也是一致的，算法2和算法3就是方法收敛方法不一样，但总归是会收敛，只要模型一样，最后函数的收敛就是差别不大的。

1. 别人提出的问题的理解：
2. 问题2：高斯混合模型可以用在什么地方？可以举一个具体的例子吗？

自己的理解：高斯混合模型是多个高斯模型通过权重组合起来的，所以适用于图像处理，再识别前景和背景的时候，前景和背景的灰度图是不一样的，所以灰度分布图的模型建模就可以用高斯混合模型去做，利于分类和分别分析。

1. 问题3： E步和M步分别想要得到什么结果？

自己的理解：E步是想直接计算出可以用来下面i+1步求最大值的Q值，然后在M步求出对应的θ值使Q最大，这样E步和M步交替计算，就可以算出让Q最大的θ，即参数。

1. 问题4：怎么理解EM算法不能保证找到全局最优解？

自己的理解：EM算法的期望值是有下限的，只要不断提高这个下限，就可以让期望值一直增大，虽然每一次迭代都可以让对应的Q值达到最大，但是Q值最大，未必就一定是期望为对应的最大，所以未必是全局最优解。

1. 读书计划

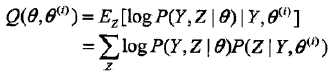
1、本周完成的内容章节：《统计机器学习》第九章

2、下周计划：《统计机器学习》第十章

四、读书摘要及理解

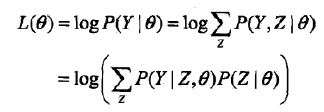
1、概率模型有时候既含有观测变量，又含有隐变量或潜在变量。隐变量就是影响观测结果的变量，但是我们在实验的时候不能直接观测到影响，所以就是隐变量。对于概率模型的变量都是观测变量，那么只要有观测数据，就可以用极大似然估计或者贝叶斯估，但是对于隐变量，就不能直接这样估计。EM算法就是含有隐变量的概率模型参数的极大似然估计法。

2、EM是：给定观测变量数据Y，隐变量数据Z，联合分布，条件分布，那么选择参数的初值，开始迭代，E步是计算：



用第i次迭代的参数的估计值，计算第i+1次迭代的Q值，计算出来，然后在M步，求使Q值极大化的，确定第i+1次迭代的参数的估计值，然后重复2和3步，直到收敛。停止迭代的条件可以是Q值的变化小于一个阈值，或者是值得变化小于一个阈值，均可。

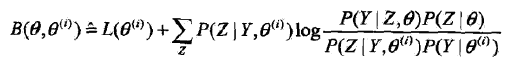
3、我们的目标就是极大化观测数据关于参数的对数似然函数，这是直接的想法，即：



这里经过计算，可以得到：



其中：



那么B函数就是L的一个下界，只要最大化B，就可以让L增大，这里的最大化B就是最大化Q，所以EM算法是通过不断求解下界的极大化逼近求解对数似然函数极大化的算法。

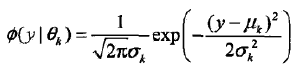
4、EM算法有这样几个不难证明的特性：是单调递增的，即：

如果有上界，那么收敛到某一值，在函数Q与L满足一定条件下，有EM算法得到的参数估计序列的收敛值是L的稳定点。这里最终得到的是随着初值不同而不同的，所以最好是多选几个初值进行迭代，找到最好的。

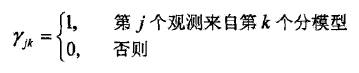
5、EM算法可以在高斯混合模型中应用来计算每一个模型对结果的贡献，混合模型为：



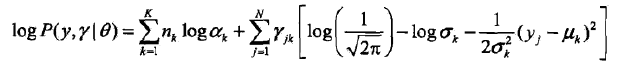
其中：



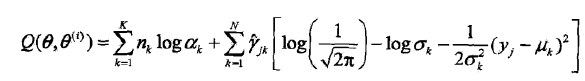
这里EM算法中首先要确定隐变量，就是第j个观测来自第k个分模型：



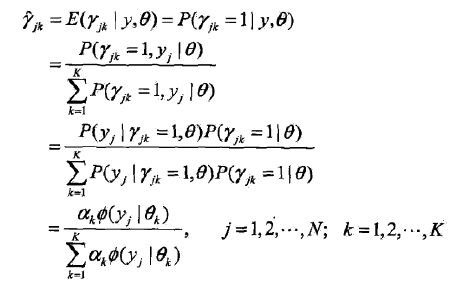
那么可以写出我们估计的最大似然函数就是：



有了这个，就可以计算Q函数：



其中：



然后M步就是极大化Q函数找到对应的值。

6、F函数的极大极大算法就是类似于Q值函数的，F函数如下：



对固定的，求P使得F极大化，对固定的P，再求使F极大化。这就是GEM算法1，GEM算法2就是找到一个使得Q值大于的Q值，就可以了。GEM算法3就是的每一个分量找到一个最优值，再继续下一次迭代，这是为了防止Q的最大值太难算了。