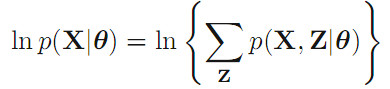
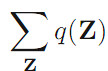
# EM算法

如果我们碰到一个这样的数学模型，它的参数θ（高维）控制着两层变量，一层变量为隐层变量，一层变量为观察变量，那么假设现在我们观察到了观察变量，想估计θ，那么我们可以用最大似然去估计它。既要求lnp(X|θ),但是我们只知道联合分布，而不知道边缘分布，那么就要转化成为（1）

但是右式难以计算，那么要想办法把求和符号放在对数外面。下面叙述如何变成想要的数学形式。利用贝叶斯公式可知p(X|θ)=p(x,z|θ)/p(z|x,θ)。同时取对数得lnp(X|θ)=lnp(x,z|θ)-lnp(z|x,θ)（2）。我们现在观察右式，由于x是观察变量而z是隐层变量，我们仍然无法计算，怎么办呢？我们的想法就是首先处理隐层变量让这个隐层变量消失，那么就要加上求和函数，如何加这个求和函数呢？利用等式两边可以同时乘以1的性质，而1如何换做求和函数呢?设一个z上的一个函数q(z),使其在z上求和为1，这样可以再（2）式左右同乘以1，然后换成可得下式：http://fmn.rrimg.com/fmn057/20111126/2205/b_large_n8G8_32540003b9921263.jpg（3）

这样我们处理掉了隐层变量，但是q(z)的形式未知，还是不能计算，如何处理q(z)？我们想想是不是可以利用数学形式上的方便来估计q(z), 现在（3）式做变形如下：

http://fmn.rrimg.com/fmn062/20111126/2210/b_large_cMYc_05840003ff91121b.jpg（4）

后面的一项就是所谓的KL距离，通过某种限制使其是大于等于0的，前面的一项称之为下界。如果P(Z|X,θ)后验概率可以计算的话，我们令q(z)=p(z|X,θ),（也就是令kl距离为0，这样下界可以最大），（这里的θ可以预先设置一个值，以后在最大似然的过程中，迭代更新θ，这里到底为什么可以通过迭代更新找到最佳，有时候只有利用这种迭代的方式去解决没有解析解的数学问题）。这样就剩第一项下界了。这样的做法好处有2，1是把二式划归为一式，方便计算，2是我们达到了最初想要的目的，把求和符合很自然的过渡到了对数的外面。（这就是伟大的数学，艺术啊）（有人会问为什么不让q(z)=p(x,z|θ),消去第一式，利用第二式去迭代更新，原因很简单，就是一个是二元函数，一个是一元函数）。

      以上就是EM算法的数学本质，一句话概括就是，目的是方便计算，方式是迭代更新。

      迭代更新的实质：没办法一次性解决问题的时候，只能迭代解决，一次解决一个小问题，然后往后滚，每次利用上一次的计算结果，依次下去，这在算法上本质就是动态规划算法，所以我猜测最早做动态规划算法的计算机科学家一定是个计算数学家。

      我们注意到如果z的后验概率难以计算,可能因为维度太高，或者数学形式很复杂，怎么办呢？2种方法去解决，一种是采样方法，一种是变分法去估计后验概率，其中变分法的理论基础就是泛函分析，变分法在数学领域里面本身就是一个分支，这里是利用其成果。关于这2种方法都有很深的研究工作。要想继续研究这些知识，需要更深入的学习。

                                                                                                                                            2011年11月26日晚

                                                                                                                                                   于燕园