浅谈EM算法

—— fangz_z

EM算法也是一个非常常见的算法了，它可以通过迭代为很多没有解析解的问题提供一个局部最优解，在很多模型中都有应用。仔细梳理一下算法的思路流程，就发现这个算法其实还是很容易掌握的。

算法描述

考虑一个带隐变量的模型，观测数据集为X，潜在变量集合为Z，参数为θ，则可以写出它的对数似然函数：

$$\ln p(X|\theta) = \ln \sum_{Z} p(X, Z|\theta)$$

一般而言，$\ln p(X|\theta)$的形式是无法直接求解的，而联合分布$p(X, Z|\theta)$可能形式上更好看，比如属于指数族分布，但由于求和符号在求对数内部，这样计算时仍然会无从下手，于是我们的EM算法就要派上用场了。
首先我们来定性地理解一下，{X,Z}可以称作完整数据集，但这个完整数据集我们是无法直接获取的，因为Z是潜在变量。那么我们可以试试从Z的后验分布下手，因为$p(Z|X,\theta) \propto p(X, Z|\theta)$，所以这个后验分布的形式往往也是简单的，然后我们就考虑联合分布的对数似然在这个后验分布下的期望，其实就是E step：

$$Q(\theta, \theta^{旧}) = \sum_{Z} p(Z|X, \theta^{旧}) \ln p(X, Z|\theta)$$

在M step中我们就去最大化这个期望。这里区分$\theta, \theta^{旧}$，是因为后验分布是根据旧的参数算出来的，在进行M step时，它已经被固定了。
所以我们就已经能写出EM算法的流程了：

1. 为参数设置一个初始值$\theta^{旧}$
2. E step: 计算$p(Z|X, \theta^{旧})$，得到$Q(\theta, \theta^{旧})$
3. M step: $\theta^{新} = \arg\max_{\theta} Q(\theta, \theta^{旧})$
4. 检查是否收敛，否则回到2

一言以蔽之，EM算法就是计算联合分布对数似然在潜在变量的条件分布下的期望，然后求极大，反复迭代达到收敛，最终求得观测变量似然函数的极大值

理论证明

上面只是我们定性的理解，下面我们来做一个定量的证明，为什么EM算法能够收敛
取关于Z的任一分布q(Z)，对于任意的q，我们都有以下式子成立：

$$\ln p(X|\theta) = L(q,\theta) + KL(q||p)$$

其中，

$$L(q,\theta) = \sum_Z q(Z) \ln \frac{p(X,Z|\theta)}{q(Z)}, KL(q||p) = -\sum_Z q(Z) \ln \frac{p(Z|X,\theta)}{q(Z)}$$

很容易验证上述等式的成立。
我们可以发现，$KL(q||p)$刚好是KL散度的形式，于是有$KL(q||p) \geq 0$，当且仅当$q(Z) = p(Z|X,\theta)$时等号成立。因此，

$$L(q,\theta) \leq \ln p(X|\theta)$$

也就是说，$L(q,\theta)$是$\ln p(X|\theta)$的下界。
于是，对于$\ln p(X|\theta)$的最大化我们可以分成两个步骤：

- **E step**
  可以看到，$\ln p(X|\theta)$的表达式是与q无关的，所以为了提升下界$L(q,\theta)$，我们可以减小$KL(q||p)$，注意，在这一步$\ln p(X|\theta)$是不变的。显然，当$q(Z) = p(Z|X,\theta)$时，KL散度最小，也就是$L(q,\theta)$取最大。可以发现，我们选择的$q(Z)$正是上一节中提到Z的后验分布
- **M step**
  这一步我们固定已经选好的$q(Z)$，再对$L(q,\theta)$做最大化，这一步会使$L(q,\theta)$继续增大，同时$KL(q||p)$只会增大而不会减小。因此，$\ln p(X|\theta)$一定会增大。
  我们来观察一下此时$L(q,\theta)$的形式，

  $$L(q,\theta) = \sum_Z q(Z) \ln \frac{p(X,Z|\theta)}{q(Z)} = \sum_Z p(Z|X,\theta^{旧}) \ln p(X,Z|\theta) + const$$

  这不正是我们前面提到的Q函数吗！

这样，我们就已经证明EM算法的收敛性。

其实还有一种从Jensen不等式出发来进行证明的角度，这里就不细说了，大致就是从

$$\ln \sum_Z p(X, Z|\theta) = \ln \sum_Z \frac{q(Z)}{q(Z)} p(X, Z|\theta) \geq \sum_Z q(Z) \ln \frac{p(X, Z|\theta)}{q(Z)}$$

开始，也能得出一样的结果

实战

EM算法的一个经典例子就是用来解高斯混合模型，很多参考书上都有，我就不写了（因为马上要和人去打球，懒得写了hhhh）

—— 2020.9.4