

第9章 EM 算法

在统计学中，似然函数（likelihood function，通常简称为likelihood，似然）是一个非常重要的内容，在非正式场合似然和概率（Probability）几乎是一对同义词，但是在统计学中似然和概率却是两个不同的概念。概率是在特定环境下某件事情发生的可能性，也就是结果没有产生之前依据环境所对应的参数来预测某件事情发生的可能性，比如抛硬币，抛之前我们不知道最后是哪一面朝上，但是根据硬币的性质我们可以推测任何一面朝上的可能性均为50%，这个概率只有在抛硬币之前才是有意义的，抛完硬币后的结果便是确定的；而似然刚好相反，是在确定的结果下去推测产生这个结果的可能环境（参数），还是抛硬币的例子，假设我们随机抛掷一枚硬币1,000次，结果500次人头朝上，500次数字朝上（实际情况一般不会这么理想，这里只是举个例子），我们很容易判断这是一枚标准的硬币，两面朝上的概率均为50%，这个过程就是我们运用出现的结果来判断这个事情本身的性质（参数），也就是似然。

EM 算法

EM 算法是一种迭代算法，用于就含有隐形变量的概率模型参数的极大似然估计，或极大后验概率估计。EM 算法的每次迭代由两步构成：E步，求期望（expectation）；M步，求极大（maximization）。所以这一算法称为期望极大算法，简称 EM 算法。

一般的，**观测随机变量的数据 Y (不完全数据) + 隐随机变量的数据 Z = 完全数据**

Q函数：完全数据的对数似然函数 $\log P(Y, Z | \theta)$ 关于给定观测数据 Y 和当前参数 $\theta^{(i)}$ 下读未观测数据 Z 的条件分布概率 $P(Z | Y, \theta^{(i)})$ 的期望称为 Q 函数。

$$Q(\theta, \theta^{(i)}) = E_Z[\log P(Y, Z | \theta) | Y, \theta^{(i)}]$$

EM算法步骤：

- 1. E 步求 $Q(\theta, \theta^{(i)})$
- 2. M 步求 $Q(\theta, \theta^{(i)})$ 的极大化，由此得到参数 $\theta^{(i+1)}$ ，完成迭代 $\theta^{(i)} \rightarrow \theta^{(i+1)}$ 。
- 3. 设置迭代终止条件，若满足条件，则停止迭代

EM 算法不能保证找到的是全局最优解。其结果与初值的选择有关，选择不同初值，可能得到不同的参数估计值。所以一般 **EM 算法的常用方法是选择几个不同初值进行迭代，从中选择最好的。**

EM 算法的应用

- 监督学习
- 生成模型的非监督学习(隐马尔可夫模型)
- 混合高斯模型（GMM）的参数估计