0、概述

一、EM算法

EM算法用来解决含有隐变量的概率模型的参数估计问题。

我们知道，对于一般的概率模型（本质上就是一个随机变量X的分布），可以用极大似然等方法对其未知参数进行估计。一般的概率模型的形式为：，对于具体的问题，P有具体的形式，所以当固定时，就能基于P的具体形式，对任意给定的样本观测值x，给出其出现的概率。但是对于含有隐变量的概率模型，即便固定，也无法基于一个具体的形式直接给出任意给定样本观测值x的出现概率——还需要知道模型中隐变量Z的取值z才行。

隐变量Z，与直接观测到的随机变量X相对，是模型中不能直接观测到的随机变量。虽然不能直接观测到，但Z也是模型中必需因素。因为模型需要给出X和Z的联合观测值，才能基于具体形式给出其出现的概率值。

注意，隐变量Z并不是一个任意的随机变量，一般来说，它与X的取值空间是一一对应的。通常，Z和X对应于随机试验中先后执行的两个试验。

在《聚类算法》笔记中，介绍了高斯混合模型，它可以看作是一种含有隐变量的模型。它的隐变量是“某个样本由某个高斯分布生成”。

这里再介绍一个简单的隐变量模型：两硬币模型。话说有A和B两个硬币，它们不均匀，掷它们一次，正面朝上的概率分别是和，我们随机选取其中一枚硬币，掷之，可以观测到其正面朝上还是反面朝上，假设我们不知道选取的是哪个硬币。这个试验对应的概率模型就是典型的含有隐变量的模型，在试验中，我们先后执行了两个试验，一个是随机选一个硬币，它对应于隐变量；一个是随机投掷选中的硬币，它对应于观测变量。我们多次重复这个试验，能够多次观察到“硬币朝上还是朝下”这一结果，然后可以用这些观察结果去估计和。

是的，虽然相比一般的概率模型，隐变量模型中多出了不可观测到的隐变量，但是我们仍然能基于观测值，求解出最优的未知参数（而不会因为多出了未知因素，就使结果变成多解）。这就是统计学的魅力，我们总是依据观测到的值，推断总体的情况，我们给出基于已知认识的最好的推测。即使我们了解到的信息比较少，也要充分利用这些信息，给出这些信息所蕴含的最可能的结论。这是多么理性的方法论啊。

对于隐变量模型，无法使用极大似然法估计未知参数，一般使用EM算法。

EM算法是基于极大似然的思想实现的。它迭代地逐步逼近使似然函数最大的参数，每次迭代分为两步（E步和M步）。

EM的迭代公式如下：



该公式首先对隐变量Z求积分，得到关于的函数，这是E步；然后求的极大，这是M步。可以证明，每次迭代都将使似然函数增加，因此该迭代过程是有效的。

注意，虽然隐变量模型需要X和Z两个变量的观测值，才能基于具体形式直接给出观测值的概率值，但是似然函数也是有意义的，只不过它没有简单的形式而已。

在具体的隐变量概率模型中应用EM算法，需要先给出和的具体形式，然后按照EM算法的步骤迭代求解。

二、随机过程与马尔科夫链

三、统计推断与随机采样

对于形式复杂的分布，即使能够给出其解析的形式，但是从该分布中采样也是很难的。你可能觉得只要给定一个分布的形式，那么从该分布中采样就是一个很简单的事，实际上这是不对的。比如你可以想一下对标准正态分布进行采样，应该使用什么方法？该方法是否很简单？

实际上是不简单的，一般来说，只有均匀分布是比较容易采样的，其他形式的分布的采样都得动点脑筋。

四、概率图模型

**1、隐马尔可夫模型（HMM）**

**2、条件随机场（CRF）**