EM算法及其应用技术报告

目录

[混合模型与EM算法 3](#_Toc526781448)

[高斯混合模型 3](#_Toc526781449)

[多元高斯分布的性质 4](#_Toc526781450)

[GMM模型的隐变量形式 5](#_Toc526781451)

[Jenson不等式与EM 框架 7](#_Toc526781452)

[基于EM算法的GMM模型推断 8](#_Toc526781453)

[隐马尔科夫HMM隐变量模型 10](#_Toc526781454)

[基于EM算法的HMM模型框架 12](#_Toc526781455)

[基于EM算法的时序高斯聚类GM-HMM 14](#_Toc526781456)

[利用GMM与GM-HMM进行聚类标注 15](#_Toc526781457)

[基于EM算法的混合广义线性回归 17](#_Toc526781458)

[基于EM算法的混合时序预测器 21](#_Toc526781459)

# **基础篇—混合模型与EM算法**

## **高斯混合模型**

观测数据记为,独立同分布于某个参数化密度函数,并记一组确定一组密度函数。对于其中任意一个维的随机变量，我们定义服从于一个混合密度分布：



其中，满足，为密度分布的权重，对应相应的。如果我们令每个参数确定下的多元高斯分布：

上面的混合密度分布为：

N个观测数据的联合LOG似然为：

## **多元高斯分布的性质**

基于上面的梯度，进行批量随机梯度下降。由于我们知道混合高斯密度函数是一个多峰函数，混合的密度函数个数越多（K越大），似然函数不具有全局最优值点。另一方面，参数之间交替优化的过程较难以有收敛性保证。我们引入混合密度模型的一种隐变量拓展形式，并介绍EM算法。

## **GMM模型的隐变量形式**



我们引入密度选择变量，它是只依赖参数的多项式分布随机变量，即有：

满足一个多项式分布，即。是一个指示函数，当且仅当表达式为真值时为1，非真则为0。那么我们有。假设集合包含了的所有可能取值，即满足每个。在给定情况下，则有：

这样我们得到，隐变量与观察变量的联合概率为：

可见，积分隐变量并最大化模型的LOG联合似然，等价于最大化混合密度模型似然函数。因而，N个观察变量的似然优化目标函数为：

* 联合似然函数

* LOG联合似然函数

## **Jenson不等式与EM 框架**

EM算法给出一个隐变量模型最大化似然的迭代算法，对于一个LOG似然

利用Jenson不等式，我们有：

当且仅当 有下界与上界相等。期望计算，通常用手动推导得到 函数，其它变种则依赖近似或者蒙特卡罗方法，EM算法给出第j步的更新为如下：

* **E-Step** 确定隐变量的后验概率参数
* **M-Step**

事实上，EM迭代满足如下单调性：

这个单调性提升分别对应E-Step和M-Step。

## **基于EM算法的GMM模型推断**

根据EM算法，在高斯混合模型GMM中已知，则有的后验分布中的每一个条件独立，因而满足：

依赖与以及联合分布的形态，GMM中：

那么可以得到E-Step，很明显，的后验分布为多项式分布：

最大化目标函数M-Step：

即求解下式：

：

增加拉格朗日约束的凸优化问题

对求导为0得到：

进而得到：

公式，只与包含的部分有关：

关于多元高斯分布的一些结论，考虑下面的参数化分布

因而我们对目标函数中的参数求导并令导数为0，可以得到：

上面的式子导出满足最大化条件的更新：

重复上述迭代直到收敛。

## **隐马尔科夫HMM隐变量模型**

马尔可夫模型Hidden Markov Model是一种时序数据生成的先验假设框。时序观测数据记为,是一个T长度的时间序列观测数据，它们仍然满足条件独立的假设，引入在时刻t处的密度选择隐变量，这与GMM的模型假设相同。

它依赖参数和前一个时刻的隐变量，即有：

参数是一个概率转移矩阵，隐变量表示离散状态选择，由离散状态转移到的概率为。值得注意的是，我们使用连续乘积的方式表达概率，仅为了随机变量的概率表示更为简洁。这样，我们有隐变量与观察变量的联合概率表达式：

因而联合似然函数是时序观测的概率乘积，可以设置为随机的，即初始时刻之前的离散状态等概率。

## **基于EM算法的HMM模型框架**

下面进行EM迭代算法的具体推导。每个只依赖前一个时刻的变量，马尔可夫性决定了LOG似然可以分解为成对的隐变量函数。具体地，联合分布表达式如下：

上面的表达式就是E步需要进行数学演算的部分，我们发现对隐变量基于后验参数求期望，只需要关于时刻t隐变量的边缘分布及的联合分布，也即：

在没有后验分布推导前，我们约定有如下的概率分布：

也即 ,那么就有：

是一个离散概率矩阵，。这样我们就有：

M步更新为：

E步还需确定的表达式，

由决定了具有同样的链式结构：

其中

也就说满足：

可以得到E步更新为：

其中，t=-1时刻令。

## **基于EM算法的时序高斯聚类GM-HMM**

此时有如下概率分布：

E步更新为：

M步需要求解下式：

其中 ,有满足。这个条件得到：

则有M步更新：

## **利用GMM与GM-HMM进行聚类标注**

EM标准框架中的E步基于上一步参数确定的似然后验分布来消除联合似然的隐变量。在GMM和HMM中，我们总结后验分布的形式:

* GMM后验概率形式：

因而为进行标注的k满足概率最大化：

* HMM后验概率形式：

确定了参数，我们可以得到t时刻的标注：

它依赖，进而需要依赖，所以这个计算过程仍然遵循“维特比算法”从t=0时刻开始。

## **基于EM算法的混合广义线性回归**



在前面的介绍中，我们引入隐变量模型和EM算法来解决混合高斯密度估计问题，单一的高斯分布参数估计可以认为是一种无监督学习方法。隐变量混合策略可以增强模型的表达能力，这里我们介绍它如何应用在监督学习问题中。

* 线性回归问题

维的随机变量，实值输出。线性回归需要学习一个线性函数，通常基于最小二乘损失,也即假设预测和真实偏差基于高斯噪音：

* logistic回归问题

维的随机变量，二分类输出。逻辑回归需要学习一个线性函数 ，通常基于逻辑二分类损失，也即

与GMM隐变量混合类似，这里只需要将密度函数换成上面的两种概率分布形式即可：

与GMM不同之处在于，GMM做无监督密度聚类所学习得到的是一个多峰密度分布，这里得到的混合分布需要转换为对的预测形式，也即：

这个表达式揭示了混合模型(隐变量层次贝叶斯)具有Model Averaging特性，很多应用中，我们利用这种特性构建“混合专家模型”Mixture Of Experts。下面完成具体算法推导。

* 学习混合线性回归

E步：

M步：

M步还需要优化

似然函数为：

对其中的参数求导并置为0得到：

上式中有如下表达式：

对其中的参数求导并置为0得到:

预测公式为：

其中,上面的式子有封闭解：

* 学习混合逻辑回归

由于算法范式相同，所以只需要推导M步联合概率参数，也即：

其中，对其中的求导以及参数求导：

是一个负定的矩阵，我们无法得到满足一阶梯度为0的封闭解形式，可以通过梯度下降来进行优化。

E步：

M步：

的更新则通过梯度下降完成，可以使用牛顿法、拟牛顿法来加快算法收敛。