# GMM-HMM

1. GMM-HMM思想

1）Hmm是隐状态是离散的，观测状态也是离散的。hmm中发射概率B是离散的。

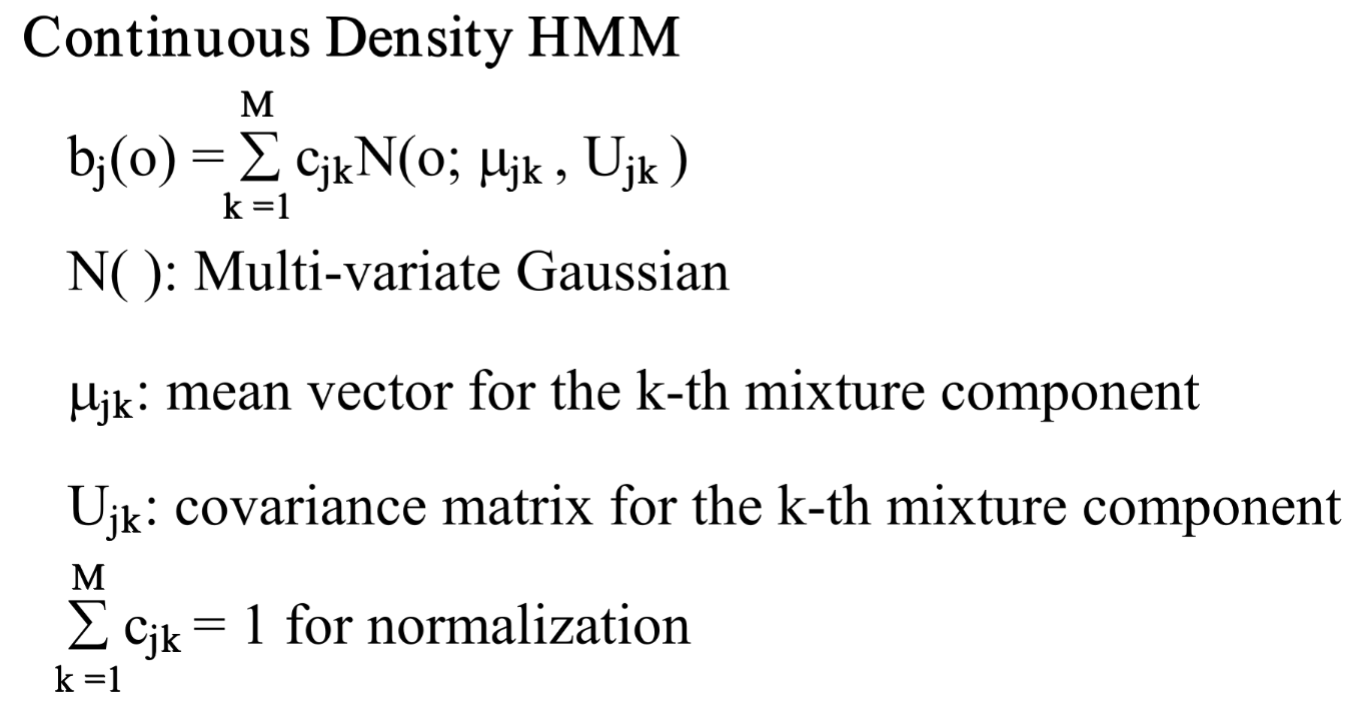
2）Gmm-hmm就是把发射概率替换成gmm（高斯混合模型）。

1. gmm参数求解是用em算法
2. hmm参数求解也是用em算法
3. Gmm-hmm参数求解也是这样的。。。
4. 拿语音识别为例，隐状态是三音素。语音识别的目的是将一串语音识别成一串音素，相当于hmm的解码的任务。首先会用mfcc提语音的特征，这样每一帧的语音都会对应成一个13维的向量表示。目标就是将13维的向量所对应的隐状态（三音素，常用三音）求出来。当然一串语音是由一连串的帧表示成的。所以一串语音最终解码成一串三音素。 这个任务很适合用hmm来做。但是hmm的发射概率是离散的，也就是说观测状态是有限的，而语音识别中的观测状态是无限的13维向量。所以需要连续型hmm来做。将发射概率，也就是一个三音素（隐状态）的发射概率变成gmm。gmm中高斯的个数可以自由设定，每个高斯都是13维的，就是（X1,X2,...,X13)服从联合正态分布。 那如果知道这个三音素服从的高斯混合模型的参数，将这个帧对应的mfcc特征向量代入这个gmm就可以得到这个三音素转移到这个状态的发射概率。

值得注意的是一个三音素就要对应一个gmm（里面有很多参数），如果有n个三音素，那么就要有n个gmm模型，每个gmm模型都对应很多个参数。

参数个数计算：13个均值，13个方差，k个高斯模型对应k个权重。就是26\*k+k个参数。 如果有n个三音素，则有n\*k\*27个参数

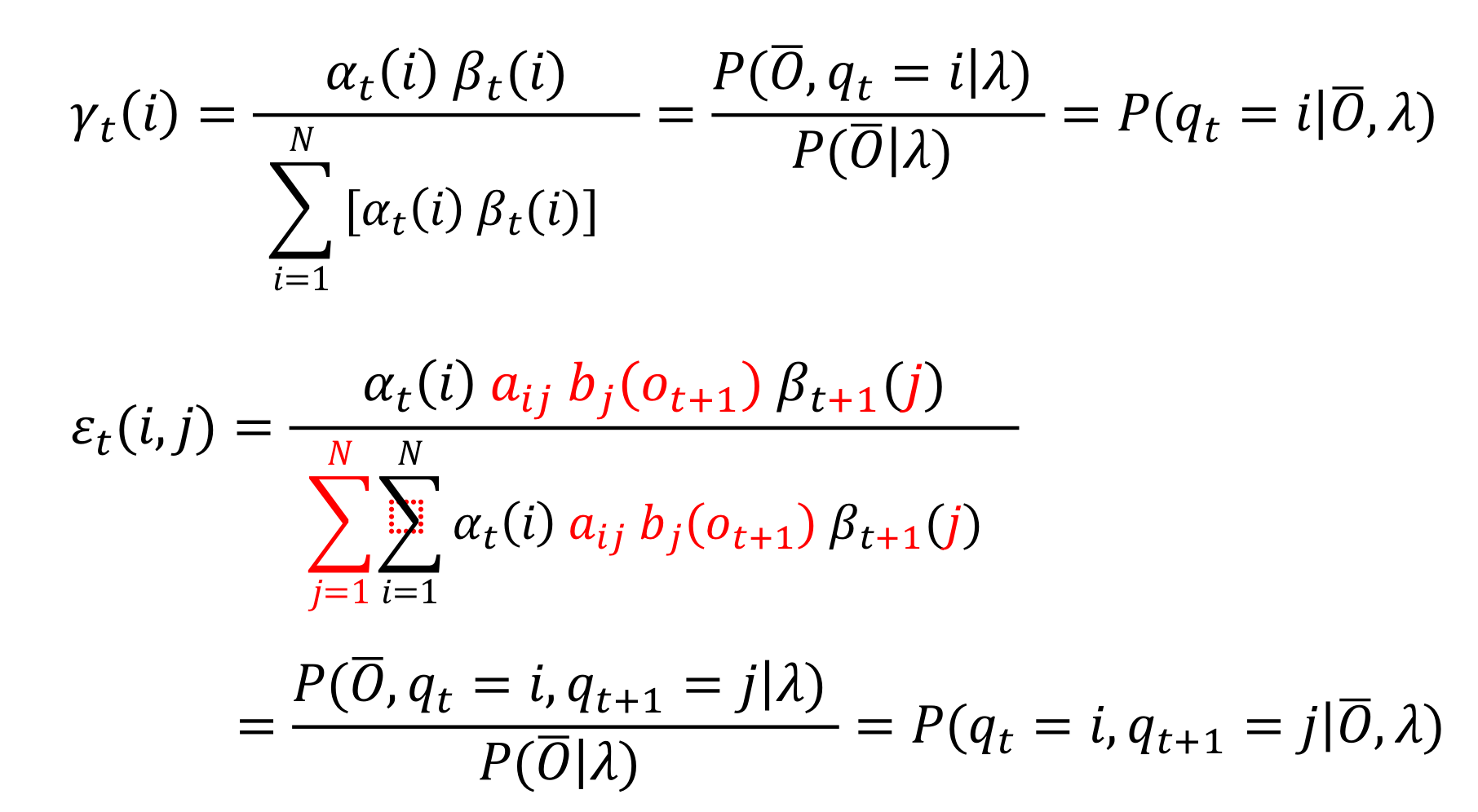
为了方便学习算法，不拿13维高斯模型举例子，太复杂，拿1维。同时1维和13维的计算方法是相似的。



1. GMM-HMM参数求解

1）gmm-hmm中的参数有hmm中的参数：A、B、Pi 和gmm中的参数：多个高斯模型的均值、方差还有每个高斯的权重。

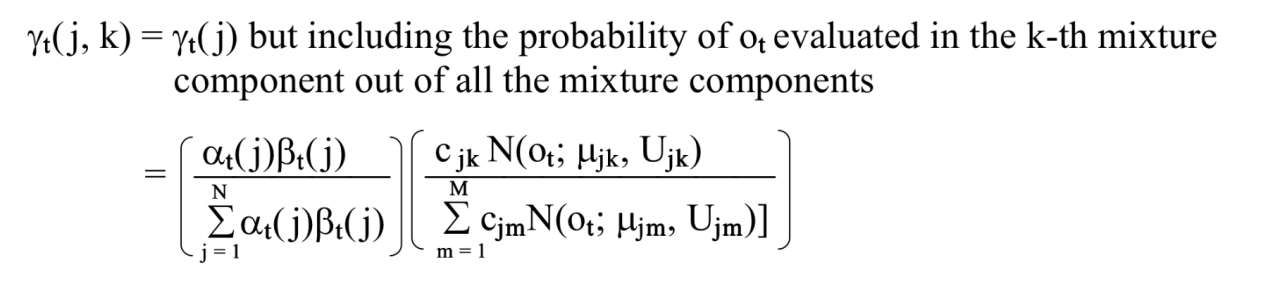
首先观察离散hmm中构造的两个变量：

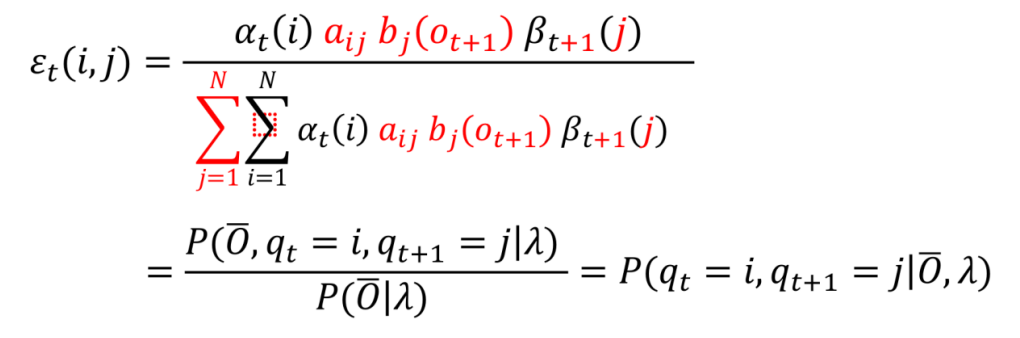


思考：

虽然A的转移概率是没有变的，但是B的发射概率变了。一个隐状态要对应一个gmm，而gmm中的参数并没有在上述变量中进行体现。

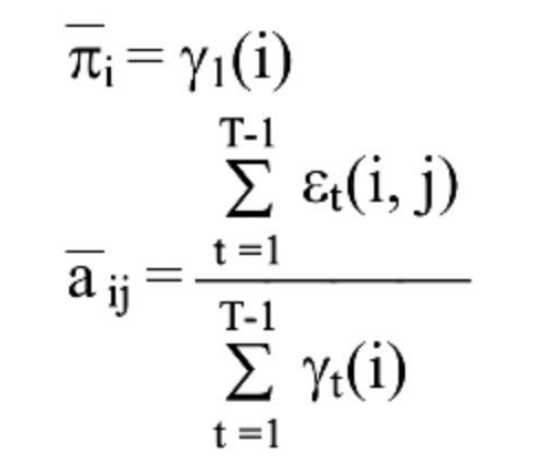
构造出来的两个变量：

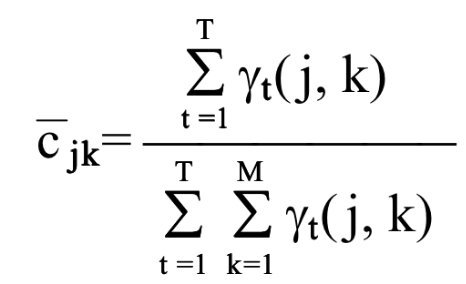


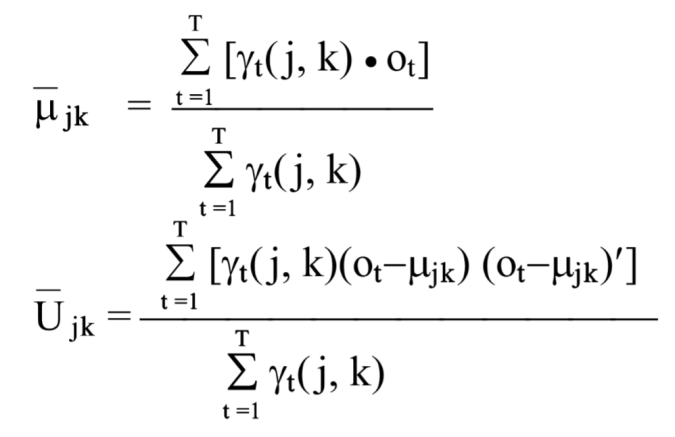


因为待求解的参数必须都要由这两个变量可以计算出来：

待求解变量：







Gmm-hmm算法的伪代码：

1. 初始化gmm-hmm中参数

2. 计算构造出来的那两个变量

3. 计算待求参数

4. 循环2，3直到收敛

从语音识别角度理解gmm-hmm训练过程：

1. 将语音使用mfcc抽取特征
2. 初始化gmm-hmm参数
3. 根据初始化参数计算构造出来的两个变量
4. 计算带球参数
5. 循环3，4直到收敛