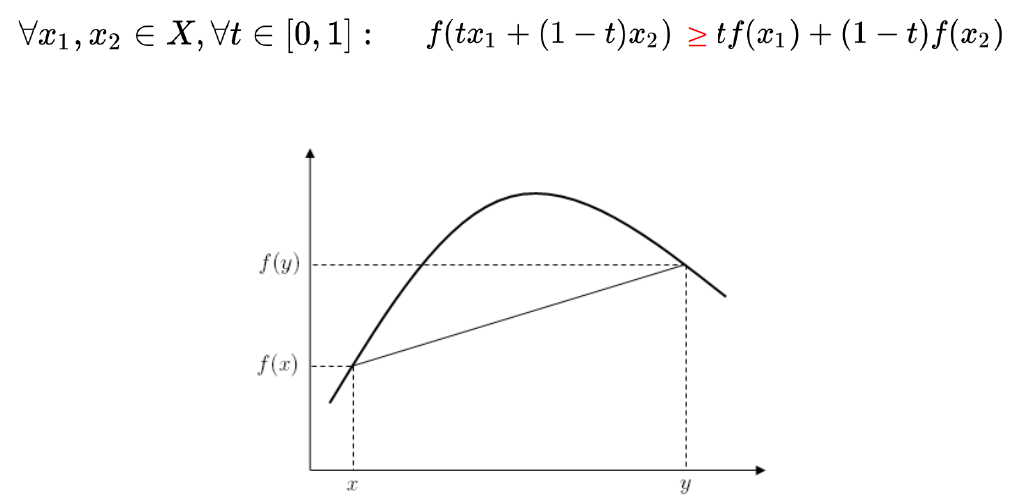
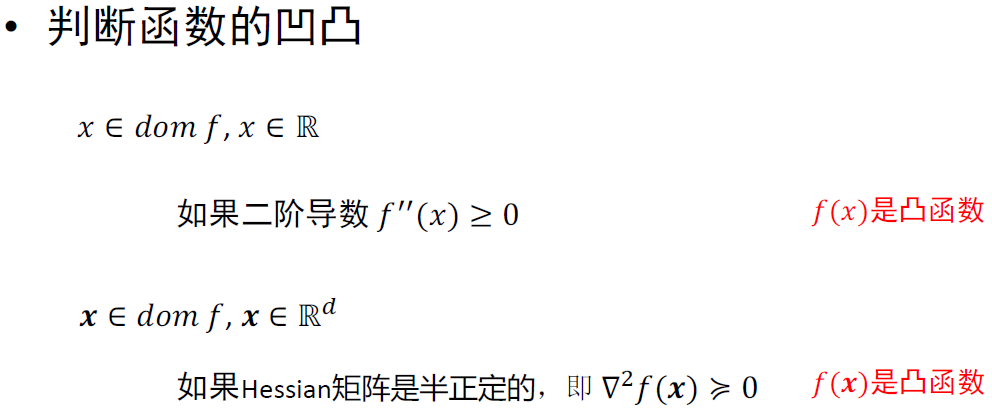
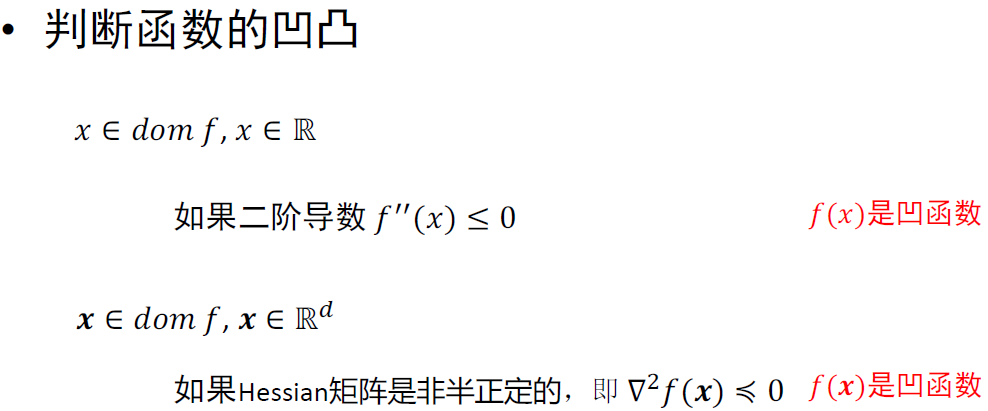
凹函数（**Concave** function）



凸函数（**Convex** function）

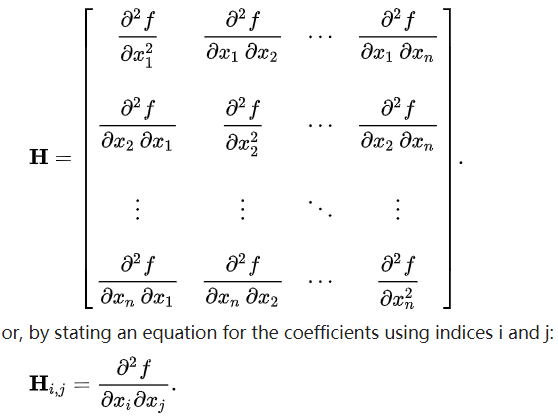
相反





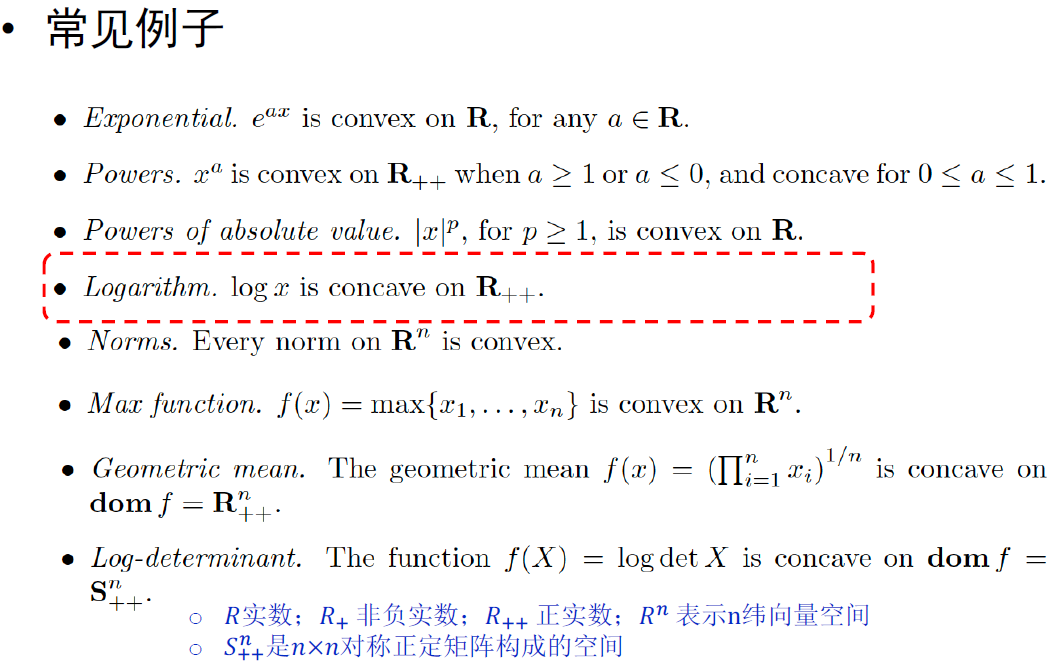
补充：

Hessian矩阵



凸函数可以找到最小值；凹函数可以找到最大值

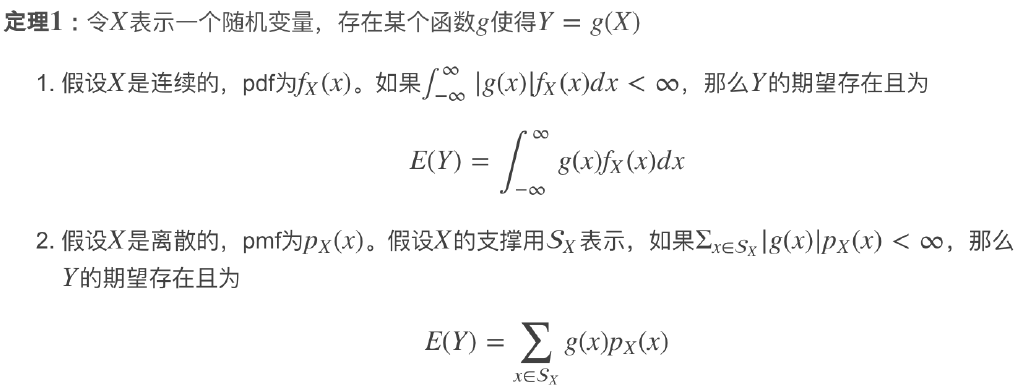
所以定义**损失函数**时，**最好是凸/凹函数**



Norm是范数

需要记住：**对数函数是凹函数**

随机变量的期望



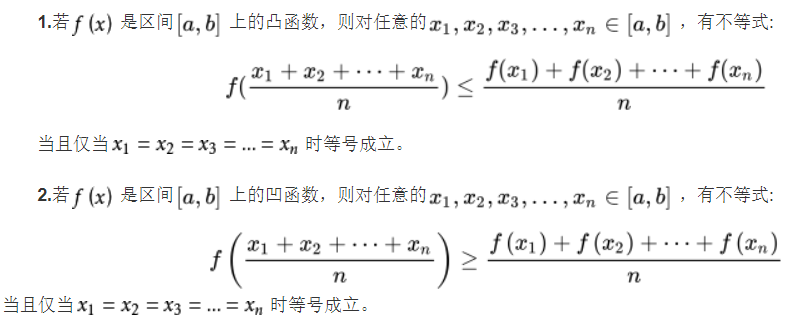
解释：

pdf是**概率密度函数**：随机变量的取值落在某个区域之内的概率为概率密度函数在这个区域上的积分

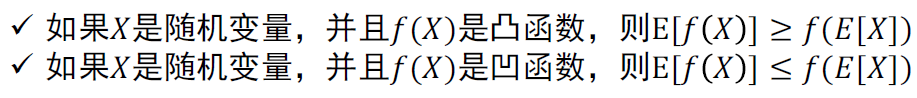
pmf**概率质量函数**：离散随机变量在各特定取值上的概率

**琴生不等式（Jensen's inequality）**

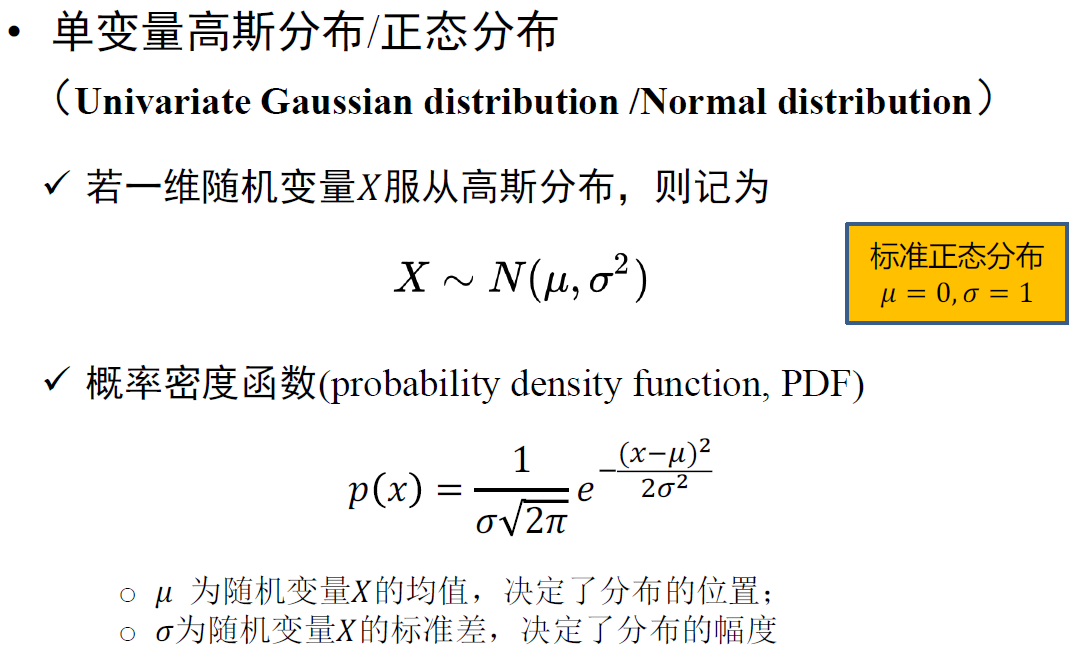
普通形式：

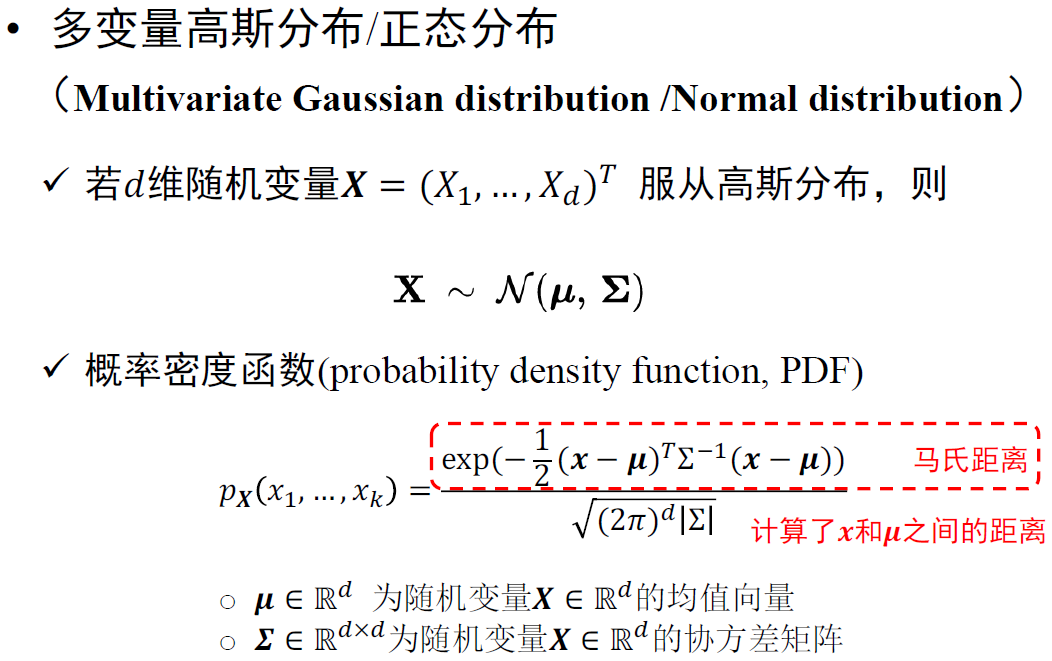


概率学上的形式：



高斯分布/正态分布





补充：

**协方差矩阵**都是**半正定**的

协方差矩阵正定需要条件，要求**满秩**

**多元高斯分布的协方差矩阵未必是正定的**

典型的混合模型（能概括所有混合模型的描述），来自Wikipedia

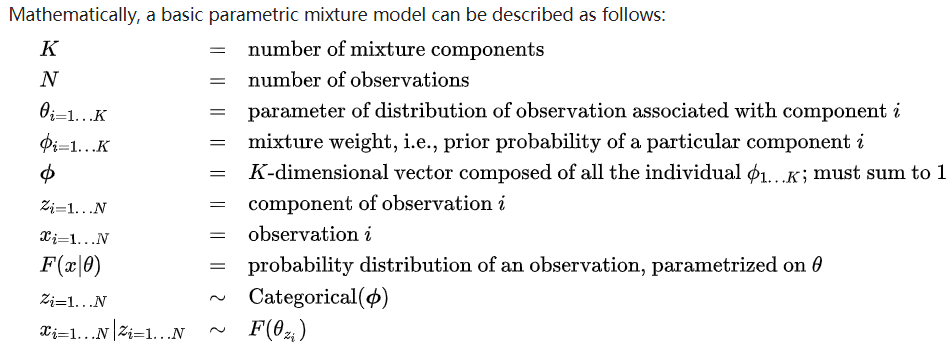
A typical finite-dimensional mixture model is a hierarchical model consisting of the following components:

**N random variables** that are **observed**, each distributed according to a mixture of **K components**, with the components belonging to **the same parametric family of distributions** (e.g., all normal, all Zipfian, etc.) but with different parameters

**N** random **latent variables** specifying the identity of the mixture component of each observation, each distributed according to a **K-dimensional categorical distribution**

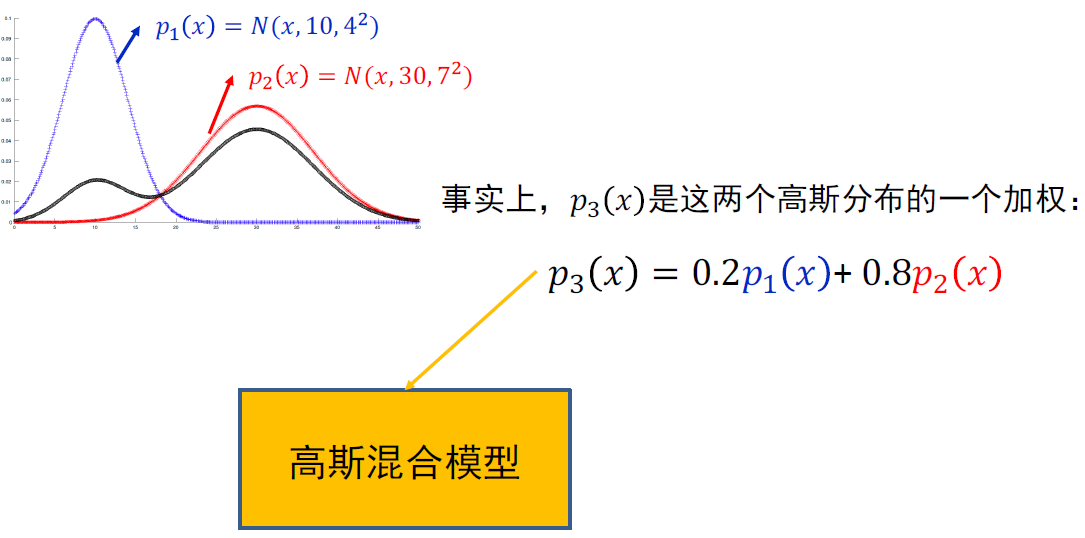
A set of **K mixture weights**, which are probabilities that **sum to 1**

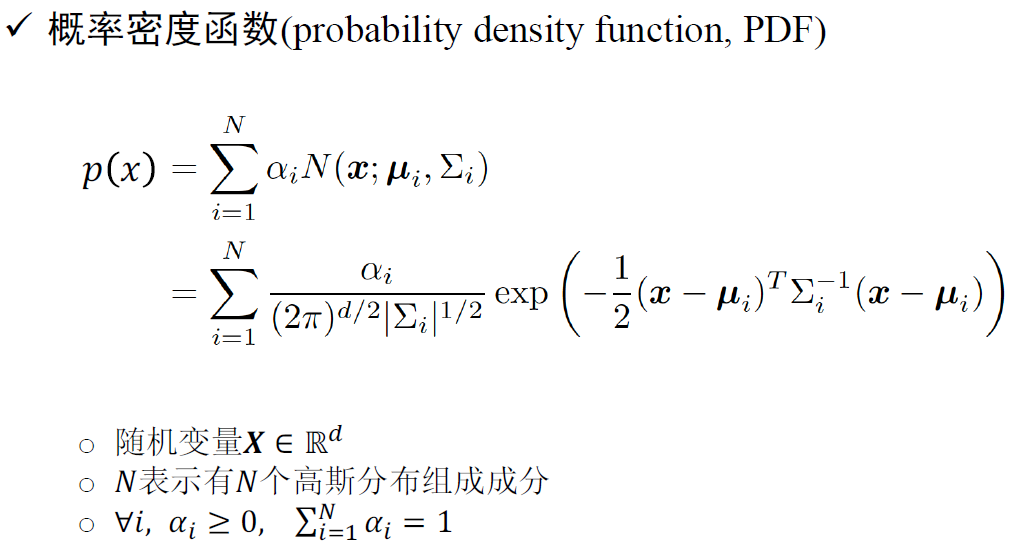
A set of **K parameters**, each specifying the parameter of the corresponding mixture component. In many cases, each "parameter" is actually a set of parameters. For example, if the mixture components are **Gaussian distributions**, there will be a **mean** and **variance** for each component. If the mixture components are categorical distributions (e.g., when each observation is a token from a finite alphabet of size V), there will be a vector of V probabilities summing to 1

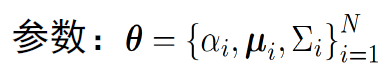


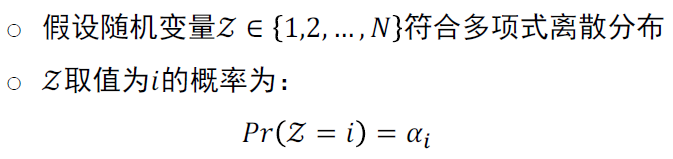
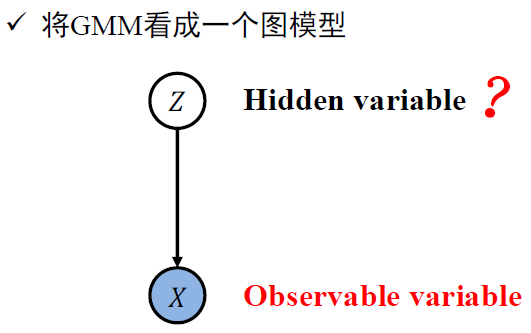
（下面的高斯混合模型中的N和这里的K含义相同，表**成分**数量；和这里的含义相同，表每个成分的权；M和这里的N含义相同，表样本数量）

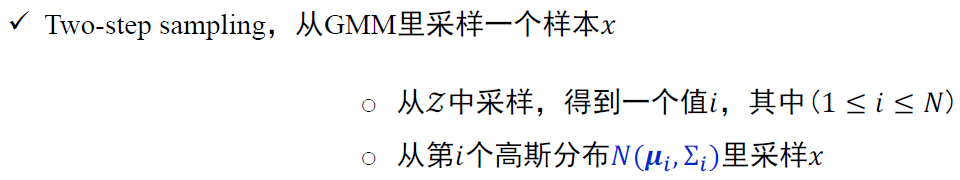
高斯混合模型（是混合模型中的一个特殊例子）





模型中需要需要求解



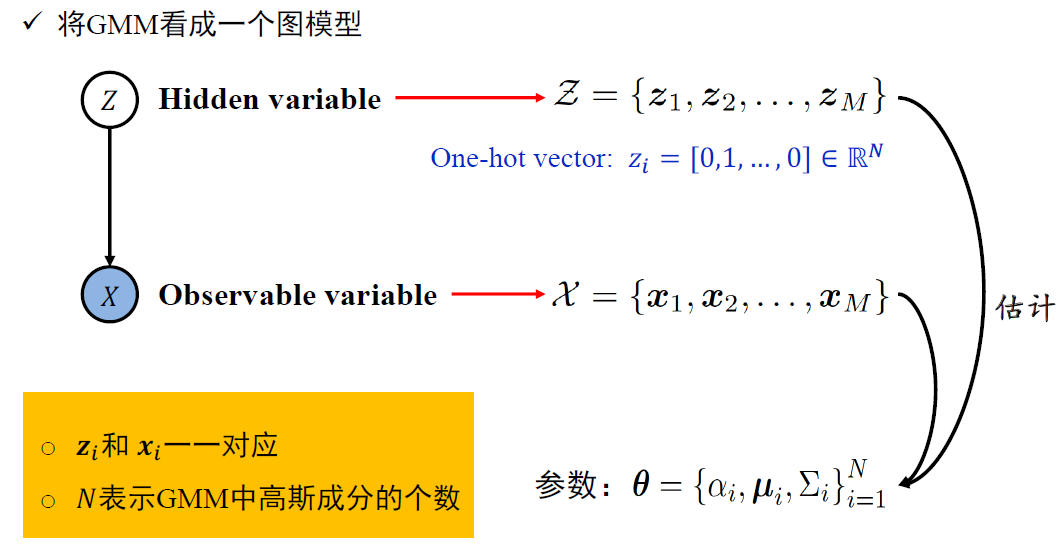


理解：

随机变量Z决定我们的样本使用哪个高斯分布。Z是隐藏变量。

X是可观测变量。

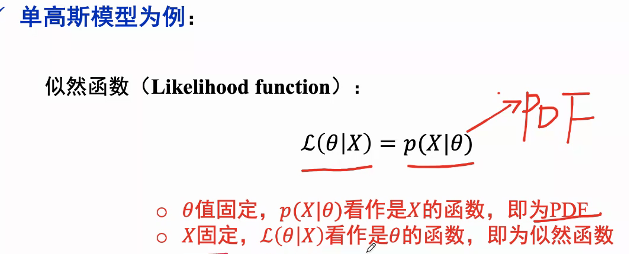
我们观察到X时并不知道是它从哪个高斯分布得到的。

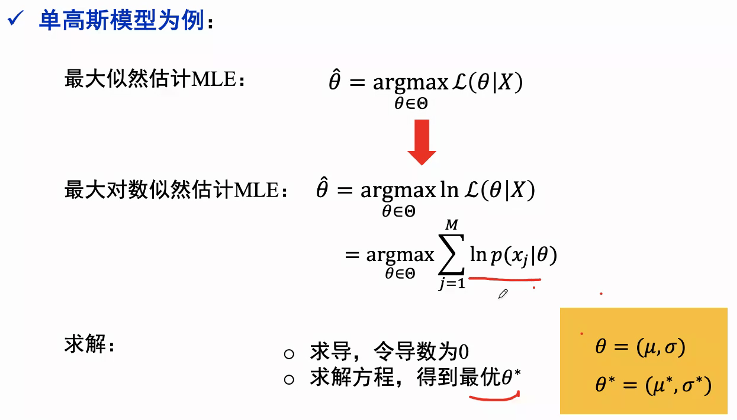


理解：

观察到M个样本（Observation），然后通过这M个样本求解出GMM模型的参数。

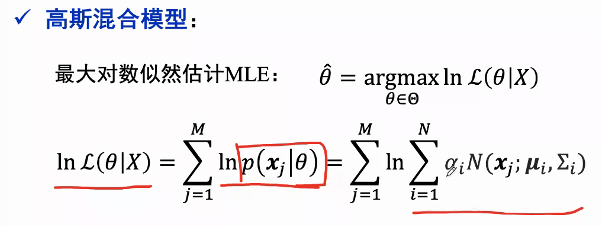
one-hot vector的意思是向量只有一个位置取1，其它位置取0。在这里表示意思是：取1的位置代表取某个成分。



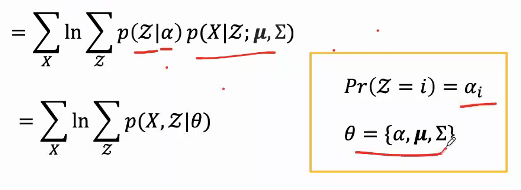


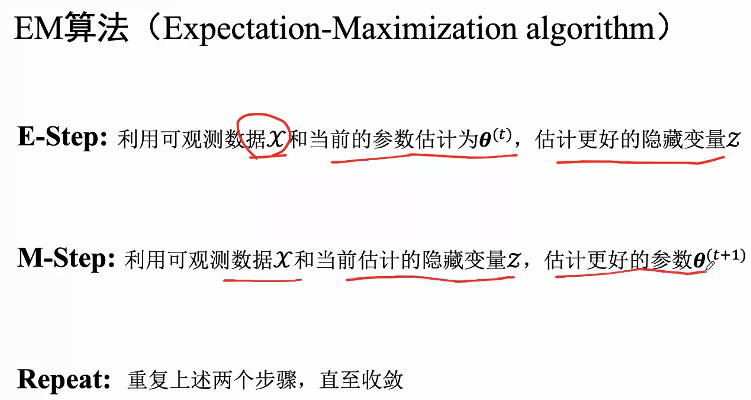
凹函数对数后还是凹函数？

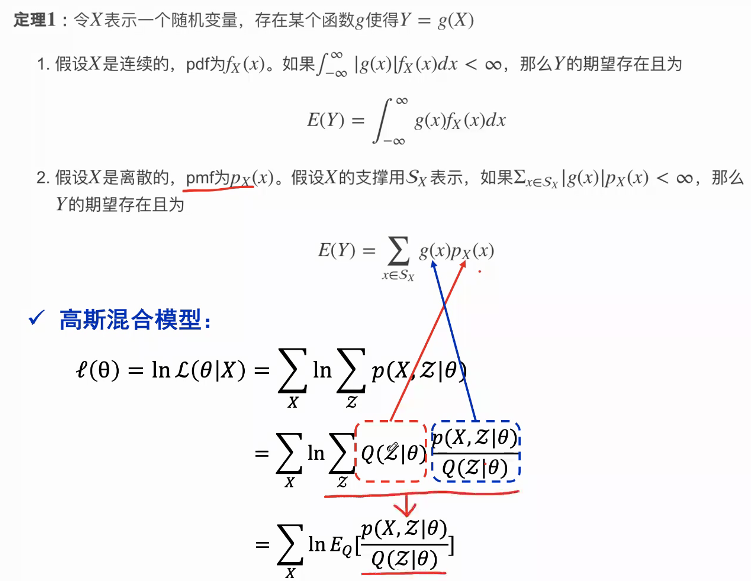
因为由多个高斯分布组成，所以

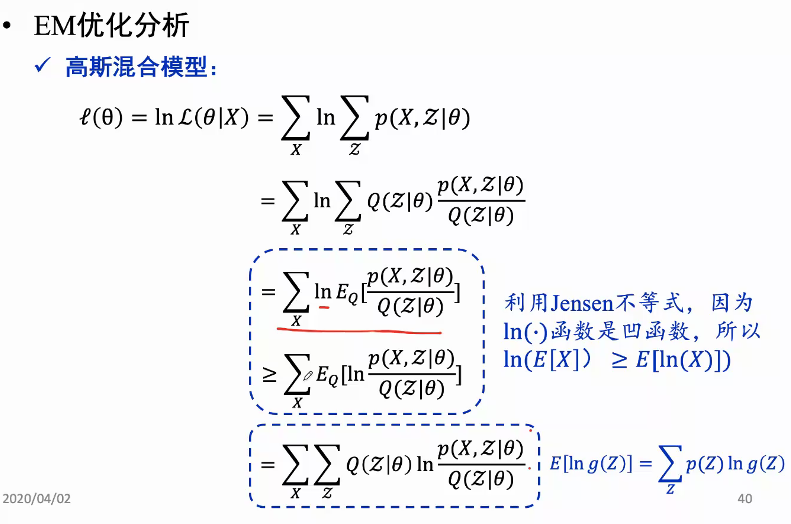


再把alpha[i]变成Z随机变量

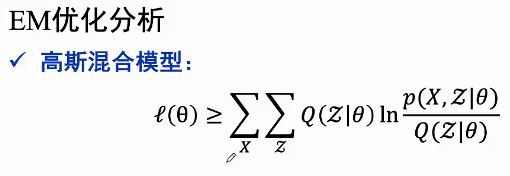


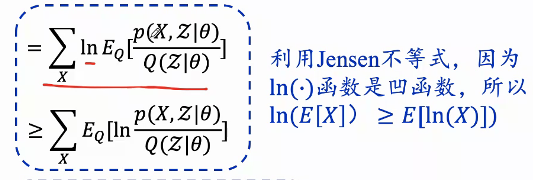




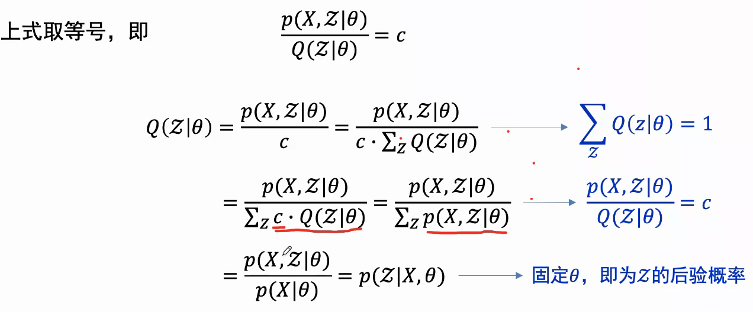


下式什么时候能取等号？



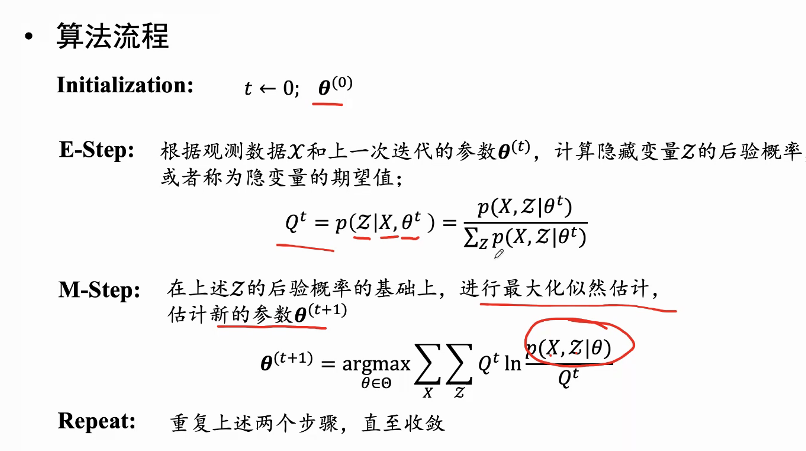
（>=出现的地方）

如何取到等号？要是常量



就可以得到高斯混合模型的下界：固定theta，计算后验概率，可以得到下界

就变成完全对数似然，是可求解的凹函数

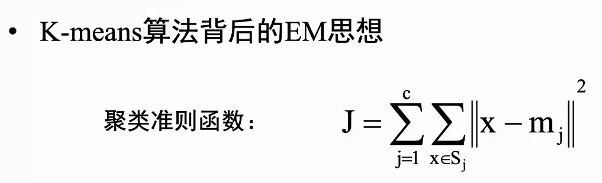


gamma I,j 表示

mi是指第i个高斯分布的统计

高斯混合模型是软的K-means

可以把类当成某个分布



mj是某个类的均值

