**Roles（slots） overlap**

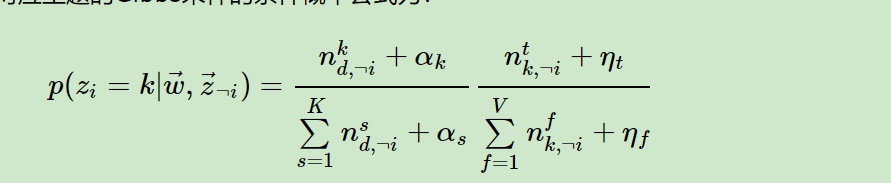
图模型分享。Vae。注意力机制。

图模型的局部图。。

迪利克雷的后验即是偏观测向量。

大概可以理解为先观测了阿尔法-1次。所以当阿尔法取1时是均匀分布，阿尔法必须大于0.。。在阿尔法小于1时会更偏0，1否则偏中间。

**Ooc，ool**



现在我们总结下LDA Gibbs采样算法流程。首先是训练流程：

　　　　1） 选择合适的主题数K, 选择合适的超参数向量α⃗ ,η⃗ α→,η→

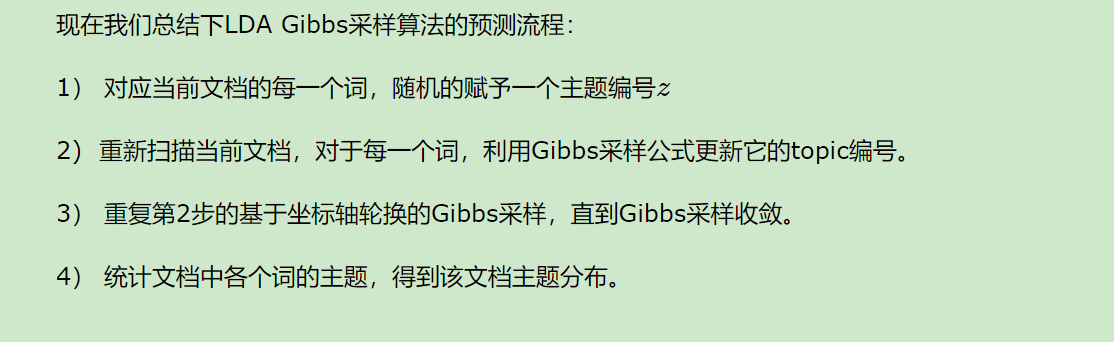
　　　　2） 对应语料库中每一篇文档的每一个词，随机的赋予一个主题编号zz

　　　　3)  重新扫描语料库，对于每一个词，利用Gibbs采样公式更新它的topic编号，并更新语料库中该词的编号。

　　　　4） 重复第3步的基于坐标轴轮换的Gibbs采样，直到Gibbs采样收敛。

　　　　5） 统计语料库中的各个文档各个词的主题，得到文档主题分布θdθd，统计语料库中各个主题词的分布，得到LDA的主题与词的分布βkβk。

下面我们再来看看当新文档出现时，如何统计该文档的主题。此时我们的模型已定，也就是LDA的各个主题的词分布βkβk已经确定，我们需要得到的是该文档的主题分布。因此在Gibbs采样时，我们的EDirichlet(βk)(βkt)EDirichlet(βk)(βkt)已经固定，只需要对前半部分EDirichlet(θd)(θdk)EDirichlet(θd)(θdk)进行采样计算即可。

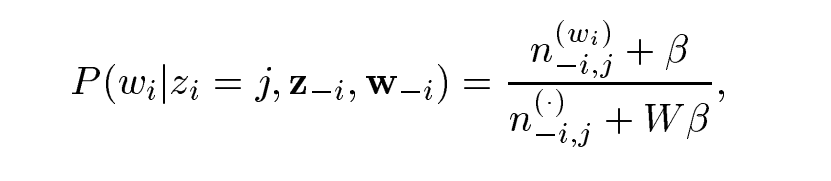


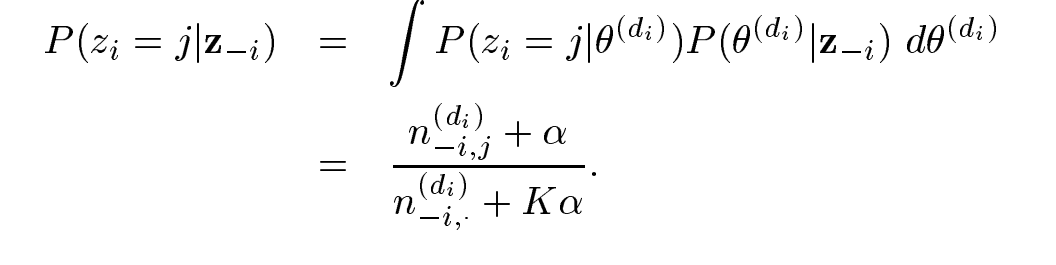
**LDA并不能聚出来监督信息，它的输出并不是具有明确可解释性的selta。但是你可以对应着去看相应的效果。**

LSA

LDA：gibbs采样，这样可以得到一个符合现有条件的样本。

（注意gibbs得到的是当前αβ的解。。）





非i只是除i以外的计数，j是第j个话题，wi是这个词语，di是这个文档。

对于许多应用，变分贝叶斯可以更快地提供与吉布斯采样精度相当的解决方案。但是，与推导类似的吉布斯采样方程相比，推导用于迭代更新参数的方程组通常需要大量工作。即使在概念上非常简单的许多模型中也是如此，如下面在仅具有两个参数且没有潜在变量的基本非分层模型的情况下所证明的。

而变分推断中我们得到的是

变分推断算法我们得到的更快的gibbs采样的精确度，

为什么能用gibbs采样解决这个问题，我们确实获得了一个样本，事实上我们可以获得更多。。

Gibbs采样通常用作[统计推断的](https://en.wikipedia.org/wiki/Statistical_inference" \o "统计推断)一种方法，尤其是[贝叶斯推断](https://en.wikipedia.org/wiki/Bayesian_inference" \o "贝叶斯推理)。它是一种[随机算法](https://en.wikipedia.org/wiki/Randomized_algorithm" \o "随机算法)（即利用[随机数](https://en.wikipedia.org/wiki/Random_number_generation" \o "随机数生成)的算法），并且是用于统计推断的[确定性算法](https://en.wikipedia.org/wiki/Deterministic_algorithm" \o "确定性算法)（例如[期望最大化算法](https://en.wikipedia.org/wiki/Expectation-maximization_algorithm" \o "期望最大化算法)（EM））的替代方法。

Gibbs采样通常用于[统计推断](https://en.wikipedia.org/wiki/Statistical_inference" \o "统计推断)（例如，确定参数的最佳值，例如确定给定日期在特定商店购物的人数，投票者最有可能投票的候选人等）。 。这个想法是通过为每个观察数据创建单独的变量并将相关变量固定为其观察值，而不是从这些变量中进行采样，将观察数据合并到采样过程中。剩余变量的分布则实际上是基于观察数据的[后验分布](https://en.wikipedia.org/wiki/Posterior_distribution" \o "后验分布)。

其实就是对参数，隐变量实现gibbs采样最后选择出现最多的。。。

我们是取 Gibbs Sampling 收敛之后的 *n*n 个迭代的结果进行平均来做参数估计，这样模型质量更高。

隐变量往往和样本数量有关，而参数不会，且通常被各个样本共享，可以通过极大似然，最大后验解决。