# 变分推断

## 估计算法：变分推断

## 先验后验概率

## 变分贝叶斯

<https://www.zhihu.com/question/20993648>

## 估计算法：变分推断：

使用已知简单分布来逼近需推断的复杂分布，并通过限制近似分布的额类型，从而得到一种局部最优，但具有确定解的近似后验分布。

1. 理解变分的思想：

一个篮球队的5位队员的技术水平相似，那么一位队员的技术水平可以用其他队员的均值来代替（平均场方法），那么如果想要评估这支篮球队的技术水平，并且在不考虑任何战术配合的情况下，可以用一位队员的技术水平来近似这支球队的技术水平。这就是变分的思想。

（具体描述机器学习-周志华）

2、什么是推断

推断就是利用已知变量推测未知变量的分布。假定所关心的变量集合为Y，可观测变量集合为O，其他变量集合为R，推断就是要由P(Y,R,O)或P(Y,R|O)得到条件概率分布P(Y|O)。

## 先验概率与后验概率

1. 先验概率是指根据以往经验和分析得到的概率。是“由因求果”问题中的“因”。如全概率公式。
2. 后验概率是指在得到“结果”的信息后重新修正的概率以获得的更接近实际情况的概率估计，是“执果寻因”问题中的“因” 。如贝叶斯公式。

下面链接参考对全概率公式和贝叶斯公式的理解：

<http://blog.csdn.net/luc9910/article/details/54377626>

## 变分贝叶斯

<http://blog.huajh7.com/2013/03/06/variational-bayes/>

变分贝叶斯是一类用于贝叶斯估计和机器学习领域中近似计算复杂（intractable）积分的技术。它主要应用于复杂的统计模型中，这种模型一般包括三类变量：观测变量(observed variables, data)，未知参数（parameters）和潜变量（latent variables）。在贝叶斯推断中，参数和潜变量统称为不可观测变量(unobserved variables)。变分贝叶斯方法主要是两个目的:

（1）近似不可观测变量的后验概率，以便通过这些变量作出统计推断（简单理解成X->Y）。

（2）对一个特定的模型，给出观测变量的边缘似然函数（或称为证据，evidence）的下界。这是为了选择模型，认为模型的边缘似然值越高，则模型对数据拟合程度越好，该模型产生Data的概率也越高。

从某种角度看，变分贝叶斯可以看做是EM算法的扩展，因为它也是采用极大后验估计(MAP)，即用单个最有可能的参数值来代替完全贝叶斯估计。另外，变分贝叶斯也通过一组相互依然（mutually dependent）的等式进行不断的迭代来获得最优解。

注：MAP--

https://www.cnblogs.com/liliu/archive/2010/11/24/1886110.html

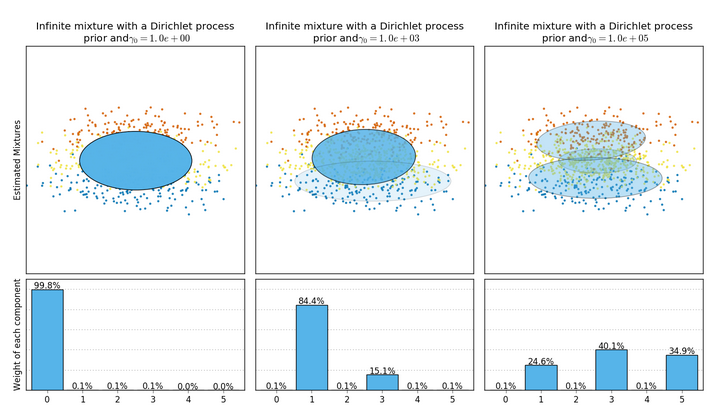
Scikit-Learn官方文档：

从先验分布中整合信息，避免了奇异点，但引入了模型的bias。

变分算法需要比期望最大化更多的超参数，其中最重要的是浓度参数weight\_concentration\_prior。给这个参数指定小值，就会使得模型给其他少数分量更多的权重。权重越大，作用越大。

[BayesianGaussianMixture](http://scikit-learn.org/stable/modules/generated/sklearn.mixture.BayesianGaussianMixture.html" \l "sklearn.mixture.BayesianGaussianMixture" \o "sklearn.mixture.BayesianGaussianMixture)该类的参数实现提出了权重分布的两种类型：具有Dirichlet分布的有限混合模型和具有Dirichlet过程的无限混合模型。Dirichlet过程推理算法近似使用固定数量的分量的截断分布。

weight\_concentration\_prior对active分量有很强的影响。当服从“dirichlet\_distribution”时，浓度权重越大导致更均匀的权重。



这里的就是weight\_concentration\_prior。