读书报告

09118119 黄一凡

# 自己提出的问题

#### 1.共轭分布与共轭函数的共轭是否代表相同的意思？

如果后验分布与先验分布属于同类，则先验分布与后验分布称为共轭分布。 共轭函数指的是对偶函数，所以我认为它们两者中关于共轭的概念是不同 的，只是借用了这一概念来表示两个分布或两个函数间的关系。

# 别人提出的问题

#### 1.LDA中，狄利克雷分布的重要性是什么？能否使用其他分 布来充当先验分布？

我们希望先验分布和后验分布属于同类，也就是共轭分布。对数据可以采 用多项式分布来拟合，而多项式分布的共轭分布是狄利克雷分布，所以我 们通常采用狄利克雷分布。

#### 2.吉布斯抽样和变分EM推理分别适合在什么情况使用？

变分推理用的是EM算法，EM算法是肯定收敛的，因而收敛性一定可以得到 保证。吉布斯抽样迭代次数较多但是实现较为简单，但变分推理的准确性 以及效率更高。

#### 3.狄利克雷分布的参数a代表什么，先验分布的a怎么选取？

a可以理解为引入的一种不存在的假设计数，类似于在没有实验数据的情况 下的一种假设，其选择很多都是根据实际应用的经验来确定的。

#### 4.LDA与潜在语义分析和概率潜在语义分析之间有什么区别 和联系？

LDA就是在PLSA的基础上再加上一个潜在狄利克雷的先验分布。两者很大 的不同在于PLSA是直 捷确定分布，相当于分布是确定的，而LDA则需要用 狄利克雷分布生成。但是两者的基本思想度可以看成是由文本到话题再到 单词。

#### 5.为什么LDA的狄利克雷先验比起PLSA的可以达到更好的效 果？

PLSA主要是没有一个固定的先验分布，所以其产生的概率分布的形式是完 全依赖数据的，而LDA引入了狄利克雷分布作为基础的先验分布，而迪利 克雷分布又是一个非常合理的先验分布，其优势类似书上说的，可以更好 的防止过拟合，提升学习的效果。

# 读书计划

#### 本周所读：

《统计学习方法》20章

#### 下周计划：

《统计学习方法》21章