**统计学习方法读书笔记**

61518122 丁自民

**读书内容**

第20章 潜在狄利克雷分布

**读书收获**

* LDA是基于贝叶斯学习的话题模型，是潜在语义分析，概率潜在语义分析的扩展，是文本集合的概率生成模型
* LDA的文本集合生成过程如下
  + 随机生成一个文本的话题分布
  + 在该文本的每个位置，根据文本的话题分布随机生成一个话题
  + 在该位置，根据话题随机生成一个单词
  + 重复以上过程，直至文本最后一个位置，生成整个文本。
  + 按以上方法，生成所有文本
* 几种分布
  + 二项分布
    - 一个事件只有两种状态，A和B。当然，也可以理解成A的发生和不发生
    - 重复此事件n次
    - 所以参数有两个，一个是A发生的概率p（自然B发生的概率为1-p），另一个是试验次数n
  + 多项分布
    - 进行n次重复独立实验，每次实验有k种结果
    - 概率参数有k-1个，最后一个事件的概率可由前面的概率推出
    - 用排列组合方式求概率
    - 当n=1时，多项分布变成类别分布，即单次实验的分布律
  + 贝塔分布
  + 狄利克雷分布
* LDA的吉布斯抽样算法
  + 在LDA文本模型中，单词分布z是可观测的，但是话题分布z却是不可观测的。
  + 吉布斯抽样算法先通过讲联合概率密度对隐变量\theta和\phi做积分，得到边缘概率分布p
* LDA的变分EM算法
  + 变分推理
    - MCMC通过抽样统计的方法计算模型的后验概率，变分推理则通过解析的方法计算模型和后验概率的近似值。
    - 但是模型的概率分布有时很复杂，没办法直接用参数估计，所以考虑用概率分布q(z)近似条件概率分布，用KL散度计算两者（观测变量和隐变量）的相似度
    - q(z)称为变分分布。如果能找到与p在KL散度下最近的分布q，

**读书疑问**

1. 使用变分法近似概率模型的原因是，待求概率模型可能很复杂。既然如此，怎么求KL散度？求KL散度相当于p已知，既然p已知，为何要用q去近似？
2. PCA，LDA和PLSA有什么区别和联系？
3. 为什么要使用狄利克雷分布作为概率模型？

**疑问解答**

回答王然问题1:

当随机事件有n中可能，第i种可能发生的概率为p\_i时，\Sigma p\_i = 1。这恰恰是一种n-1维单纯性。