统计学习读书报告

09017244 郑健雄

1. 自己提出的问题的理解：
2. 非负矩阵分解的k值是否有类似于奇异值分解方法的选择依据？其信息的压缩程度似乎并不能直观的表现出来。

讨论后的理解：

这个情况下k值可以根据经验来选择，当然k值越大，保存的信息越多，但是由于本身非负矩阵分解就是一种近似的分解方式，所以选择k值可以通过尝试的方法和以往的研究经验以及研究的问题来进行决定。

1. P332中为什么说非负矩阵分解有着很直观的解释，如何理解其"伪概率分布"。

讨论后的理解：

其直观解释是相对于其他分解方法而言的。非负矩阵分解的结果都是非负的，而负数在现实中实际上是很难进行解释的，所以非负矩阵意义的解释相比于SVD或者特征分解而言肯定是更加直观的。其伪概率分布的理解是不是真正的概率分布，是一种假设。经过查阅资料和别人的讲解，我对这个问题的理解是非负矩阵分解出的话题矩阵中的列向量并非严格的不相关的，也就是说话题可能会产生重叠性甚至内部有所关联，但是理想情况应该是话题独立。

1. 乘法更新原则是否采用了类似EM算法的思路？

讨论后的理解：

严格来说，两者的思路不同，EM算法分为E步和M步，也是一种交替的过程，属于交替更新的策略。乘法更新的思路也是交替更新两边的变量，但是其更类似梯度下降的思路，EM算法中间有着极大化的步骤，所以两者并不等同。但是证明两者收敛到局部最优点的思路都是构造辅助函数来证明并且其都是局部最优，所以两者也有类似之处。

1. 别人提出的问题的理解
2. 初始的W以及H该如何选择？是不是只能随机尝试？

自己的理解:

非负矩阵分解求解其最优解是NP问题，现在只有近似解法，所以初始值的选择其实不用考虑太多，因为往往是收敛到局部最优解，所以可以保证初始矩阵都是非负的并且多随机几次，这样可以有更好的效果。

1. 两种矩阵的分解方法各有何优劣？哪一种算法更好？

自己的理解：

非负矩阵分解的优势在于其分解的矩阵都是非负的，在实际生活中，负数是没什么意义的，所以得到的结果具有可解释性，并且其处理大规模数据更快捷，效率更高，近似效果也很好。SVM的问题是不一定有可解释性，但是SVM的结果是可以还原原始矩阵的，并且比较独特(除去倍数关系)，可以到达最优，而非负矩阵分解的求解比较近似，而且不唯一。它们都是比较好的方法，可能在不同的应用上有着不同的效果，要根据实际应用来判断。

1. P332定义的散度好像D(A||B)不等于D(B||A)？

自己的理解:

两者确实是不相等的关系，或者说是不对称的关系，两者相等代表两个矩阵相等，也就是最小化的情况了。

1. 非负矩阵分解是不是需要较多的文本数量才可达到较好的效果，如果文本数量较少是否会导致得到的矩阵不具有很好的代表性？

自己的理解：

这个问题应该对于其他的分解方法也是适用的，非负矩阵分解使用了加权和的思想，将数据拆分成不同的块进行重组，那么如果数据变多，其分解的零部件部分以及权重也就更加准确，所以多一点效果是更好地。不过也需要考虑过拟合问题。

1. 奇异值分解后矩阵中出现的负值有什么特殊含义吗？或者说，正常情况下允许分解后的矩阵中出现负值吗？

自己的理解：

与非负矩阵分解不同，SVD的思路应该是整体的加权和，也就是多个完整部分的叠加，比如图象拼接人脸，是多张脸的叠加，那么叠加的话，就会有加减法的情况，也就是负数，但是负数仅仅在数学中可以解释，现实中是比较难解释的，这也是为什么SVD等方法的解释性差的原因。

1. 读书计划

1、本周完成的内容章节：复习17章，阅读完Speech and Language Processing第三章。

2、下周计划：复习18章，阅读完第四章内容。

四、读书摘要及理解

1. 求解的方法：求解LDA的方法有SVD和非负矩阵分解。SVD的思想是整体的加权和，奇异值代表了不同矩阵的权重，是经典的方法并且效果很好，但是解释性比较差。而非负矩阵分解是部分加权和的思想，有较好的解释性。

2. 非负矩阵分解：非负矩阵分解是直接将非负的矩阵分解为两个非负矩阵的乘积，其采用的是加权和的思想，与SVD等方法有所不同。其有很多优点，比如可解释性和计算的方便性，但是其缺点在于不能得到最优解，只有近似解法和局部最优，而SVD可以得到最优解，两者都是较好的方法。求解非负矩阵分解的方法是乘法更新算法，其设定散度或平方损失作为目标函数，进行交替更新的优化，其使用了梯度下降的方法，可以到局部最优。

思考：

非负矩阵分解和SVD象征了分解思路的差异，非负矩阵分解是分解部件的思想，而SVD则是整体分解，层层叠加的思想。不同的分解方法有着不同的效果和应用，所以不能直接评判其优劣，因为其有着各自的优势和缺点，适合的方法才是最优的。比如SVD用于降维和挖掘主成分就是很好的工具，其考虑整体结构的时候更好，而对于图象分解或者文本的分解，可能非负矩阵分解的效果更好，因为它可以分解不同的部分，比如眼睛，嘴等部位，然后是通过组合的方法来构成人的脸，那么这种方法就更直观，也更好解释，因而有着更好的应用效果。