**第十七章读书报告**

09118223 吴亦珂

读书进度：读完17章

1. **问题列表**

（我提出）初始的W以及H该如何选择？是不是只能随机尝试？

讨论结果：初值应该是随机选取，再通过后面的步骤逐步进行优化。就和梯度下降一样，初值选取并不是非常重要，通过迭代可以找到局部极小值。

（我提出）两种矩阵的分解方法各有何优劣？哪一种算法更好？

讨论结果：SVD是直接进行矩阵的奇异值分解，优点是奇异值分解使得不同列内积为0，相当于不同话题直接的相似度几乎为0，但是其并没有考虑到非负的问题。而非负矩阵分解使得结果一定不为负值，但是不同话题之间的关联却没有考虑到。因此，两种模型各有有缺，应该在不同的场合有自己的用处。

（别人提出）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"。

讨论结果：本来矩阵统计的数应该都是正的，而非负矩阵分解则满足了这一点，符合实际情况。同时将X费解为 W和H的乘积，原矩阵每一列对应于一个线性组合，也满足实际的情况。伪概率分布我的理解就是除了话题数目时提前规定之外，而进行分解并没有说将话题进行很明确的说明具体是什么话题，对应什么文章，因此所得的概率分布只能是假设的，这也就是伪的原因。

（别人提出）乘法更新原则是否采用了类似EM算法的思路？

讨论结果：我觉得可能想法是都是通过迭代来求解一个优化问题。但是ER的思路是求解Q函数然后再求解求极值，不断循环知道收敛。这个相当于是对梯度下降进行了优化操作，交替对变量进行更新。

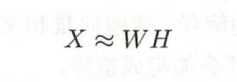
（别人提出）为什么WH与X完全相等很难实现，当m×k+k×n>m×n的时候，仅用解未知数的方法就能得出多组解了

讨论结果：我觉得虽然理论上当做求解方程组确实可以求出很多解，但是保证所有解均非负就很困难了。一般这种矩阵规模都不小。

1. **下周读书计划**

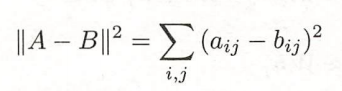
下周计划阅读第18章内容。

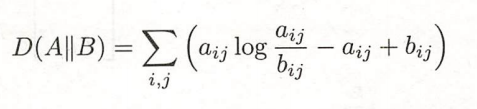
1. **读书收获**
2. 非负矩阵分解：将一个非负矩阵分解为两个非负矩阵乘积的形式。但一般很难做到完全相等，所以一般近似相等即可。

即

考虑到如果将X写成列向量的形式，则可以表示成W的一列数的线性组合形式。因此W的列向量为一组基，H为线性组合的系数。

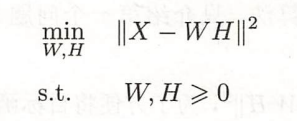
1. 潜在语义分析模型：如果将X视为单词文本矩阵，可以将W看作话题空间，H看作话题文本矩阵。所以，将一个单词文本矩阵进行分解，可以直接求得其话题空间以及话题文本矩阵。
2. 非负矩阵分解的形式化：可以把矩阵分解问题转化为一个数学问题进行求解。可以通过损失函数对这个问题建模。

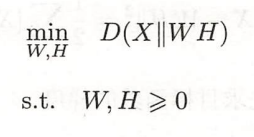
平方损失：

散度：

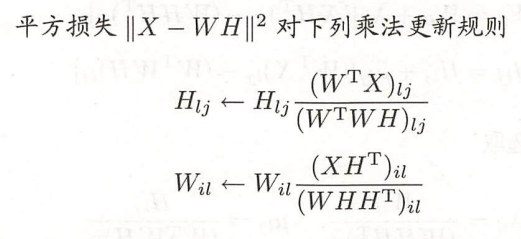
两者都是当A=B时，到达下界，因此适合作为损失函数。

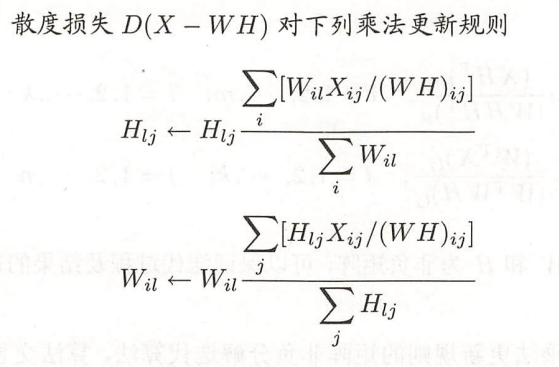
通过损失函数可以表示为一个最优化问题：





1. 优化算法：考虑到这次优化要痛死面对两个变量，因此传统的梯度下降并不适用，收敛的速度过慢，因此采用共轭梯度下降的算法。简单来说，就是交替对参数进行更新。





具体更新规则：

