读书报告

09118119 黄一凡

# 自己提出的问题

#### 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？

这个问题应该对于其他的分解方法也是适用的，根据非负矩阵分解的能力， 其应该是使用了加权和的思想，那么如果数据变多，其分解的零部件部分以及权重也就越准确，不过数据过多也需要考虑过拟合的问题。

# 别人提出的问题

#### 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？

这个意义应该仅仅存在于计算本身，也就是分解的结果，但是其应该不具有现实意义， 或者说类似于上面对于SVD的比喻，是对多张人脸的加减法，这种解释就不如非负矩阵直观合理。

#### 2.两种矩阵的分解方法各有何优劣？哪一种算法更好？

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，并且比较唯一(除去倍数关系)， 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。

#### 3.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"。

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。伪概率分布即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。

#### 4.P332定义的散度好像D(A||B)不等于D(B||A)？

这个应该是不相等的关系，或者说类似书上说的不对称的关系，不过如果两个矩阵完全相同的话，结果应该就会变成0，也就是会相等。

# 读书计划

#### 本周所读：

《统计学习方法》17章部分

#### 下周计划：

《统计学习方法》18章