统计学习读书报告

09017244 郑健雄

1. 自己提出的问题的理解：
2. 奇异值分解算法如果直接求解ATA的特征值，会带来哪些效率上的提升？计算的步骤与15.2中的基本方法会有很大差异吗？

讨论后的理解：

实际上直接求解的话，可以不用去计算矩阵相乘的结果，这样对于一个比较 复杂的矩阵来说实际上我们可以让效率提升，因为矩阵的结果对于奇异值分 解并不重要。

1. 矩阵的外积展开形式是否是奇异值分解对矩阵近似的一种直观表现？能否从外积展开式角度直观证明或者说明最小平方损失是如何得出的？

讨论后的理解：

矩阵的外积形式实际上是按照奇异值作为权值的加权和，其很直观地展示了大的奇异值使得其对应的矩阵权重更大，也就表现出了更重要的地位，也直观的说明了奇异值分解的逐步近似性。除此之外，最小平方损失也可以理解为舍弃了外积展开中最小的那些奇异值加权和，进行变化后实际上可以观察到其损失也是最小平方的。

1. 别人提出的问题的理解
2. 矩阵奇异值分解计算问题中，我看网上有的说法计算U矩阵是计算AAT的特征值，与书上不同，这两者有什么关系？

自己的理解:

两种方法应该是等价的，说U的列向量是AAT的特征向量，所以U和V的地位是一样的，也可以用特征值的方法求出。而书中的方法应该是直接用了构造奇异值分解时使用的那些等式关系，也就避免了再计算一次AAT矩阵并求其特征向量的开销，但是结果应该一样。

1. 在定理15.3中，A'是阶段奇异值矩阵，奇异值分解时用的是U，V，但按照书上前面15.1的定义，A'用的U，V应是前k列，而不是A的U，V。这两个有区别吗？

自己的理解：

奇异值分解直接使用U,V的话是完整的奇异值分解，和使用秩r的奇异值分解是等价的，而k的话是截断奇异值分解的情况，是近似。它们都是奇异值分解，只是近似程度不同。比较直观的理解就是外积展开式15.47，在秩为r的情况下，sigma r+1到n实际上都是0，有效的只是前r个加权和，而使用k个奇异值则代表了进一步的压缩，也就是保留了一部分加权和的值。我的理解是直接使用U,V进行奇异值分解得到的是完整的形式，然后如果只保留r个奇异值就变为紧奇异值分解，但是根据外积展开式，两者等价，再向下压缩就会造成实际的压缩，也就是k的情况，从外积展开式中可以直观的观察到只保留了1到k的奇异值的权值和，其对应的是截断奇异值分解。在不压缩的情况下，两者等价，压缩的情况下，是最小平方损失的近似。

1. 书上之前提到奇异值分解的几何解释，那么利用截断奇异值分解对矩阵进行压缩在几何上是否有直观的解释呢？

自己的理解:

这个可以根据书中15.4的例子来说明一下，并且根据p279的内容对照一下。如果将P280的奇异值矩阵只保留一个奇异值，那么对于两个标准正交基的旋转就会不同于整体情况而是变为一个一维转换的情况，然后缩放也会变为1维的缩放，最后的旋转还原也是一个一维的还原，感觉直观一点的理解就是变换的维度压缩了，实际变换的信息有损失。原始的V矩阵式2\*2的结构，所以它对正交基进行旋转变换之后得出的向量也是一个二维的结果，感觉是在两个维度进行了旋转，而如果截断了，正交基的旋转变换会变为一个数字，也就是说只是一个维度进行了旋转，奇异值也是，只对一个维度进行了放大，所以最后的旋转结果肯定会由于维度的压缩而导致信息丢失，可以实际对p279的图进行一下实验，可以更直观的理解压缩对于几何产生的影响。

1. 奇异值分解时使用ATA和AAT得到的结果有什么不同？为什么选取ATA的方法？

自己的理解：

两种方法本质上没有区别，结果也是一致的。ATA得到的是V的特征向量和奇异值的值，我们可以用它直接推导U的信息。但是如果使用AAT来推导的话，实际上得到U的信息和奇异值的信息，根据关系也可以直接推导V。选择哪种方法应该是一种习惯问题。

1. 读书计划

1、本周完成的内容章节：复习15章并且复习和完成了21章剩下的一点内容

2、下周计划：22章

四、读书摘要及理解

1. 奇异值分解的计算：

奇异值分解的计算通过计算ATA的特征值和特征向量来实现，特征值提供了奇异值的值，特征向量构成了V矩阵，而通过矩阵的关系可以使用A和奇异值，特征向量推导出U，从而实现奇异值分解。但是实际情况里不直接计算ATA的结果，而是直接计算特征值。

1. 奇异值分解的意义：奇异值分解其实是对于矩阵平方损失最小情况下的近似，是最优矩阵近似，所以奇异值分解有着很广泛的应用前景。而从另一个角度看，奇异值分解实际上可以写成外积展开式，也就是按照奇异值的矩阵加权和，这样很直观的表现出了奇异值分解对于矩阵本身的近似性，并且可以很好的理解其结构。

思考：奇异值分解可以理解为外积展开式实际上是一种本质的体现，它非常明确的说明了奇异值作为一个加权的作用，也很直观的解释了为什么保留大的奇异值可以更好的近似矩阵本身，其在PCA里也有很广泛的应用，是经典降维方法的基础。