**读书报告**

61518424 王贵涛

1. **问题与解答**

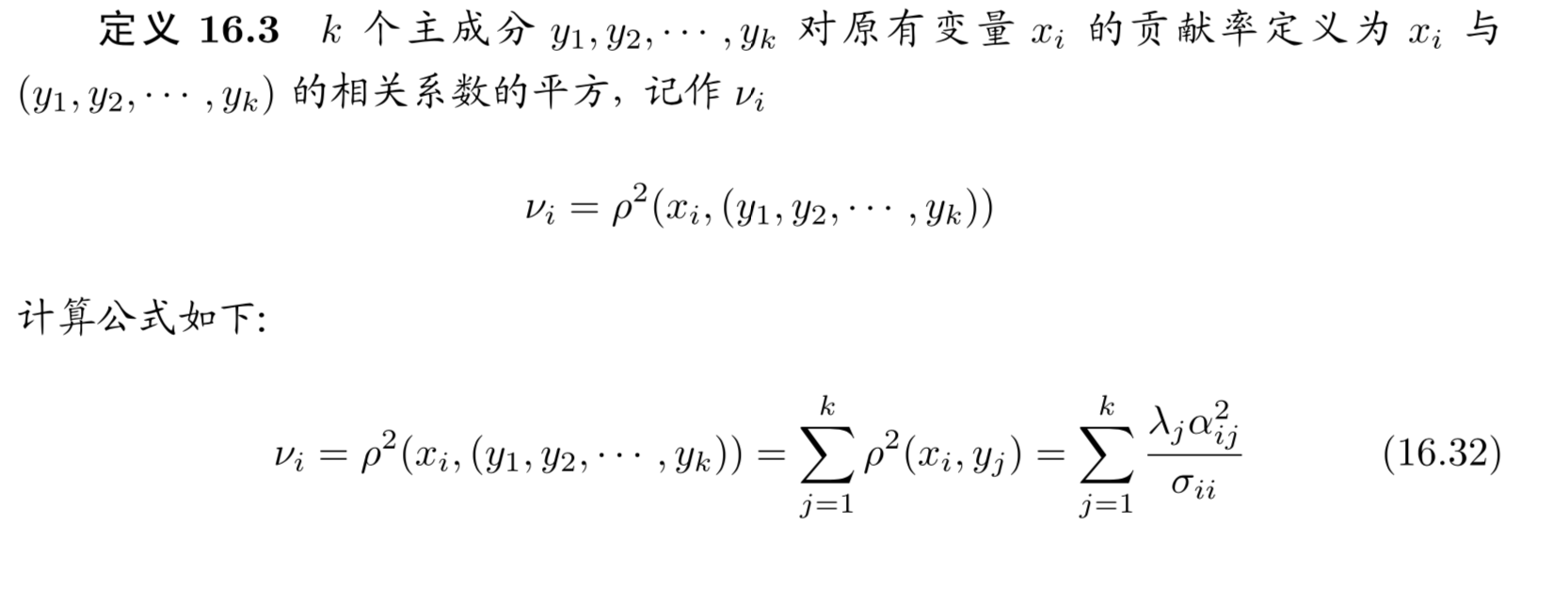
我提出问题：

1. 例题中，文科成绩提高都可使 y2 提高，理科成绩提高都可使 y2 降低，为什么不是恰好相反，这是由什么决定的？

讨论结果：这个应该就是分析出的结果，主成分分析不会赋予主成分意义，它只是认为这个文科成绩提高而理科成绩降低的主成分很重要吧，可能也有完全相反的主成分，只是没有给出，决定这种关系的应该还是数据内在的关系，应该比较抽象，这个只能通过主成分分析和观察把它表示出来，决定因素可能不是很好理解。

1. 第一、第二主成分对变量x的贡献率怎么求？

讨论结果：



别人提出的问题：

1. 样本协方差矩阵S为什么是总体协方差矩阵的近似？  
   我的解答：总体协方差矩阵除以的是n，样本除以的是n-1，和概率统计里的总体和样本一样，当n足够大时，两者接近。
2. 在实际使用中，如何选取合适的信息保留比例？

我的解答：根据实际需求吧，比如需要将文件压缩到多大，也有不那么主观的方法：把它看成一个 model selection 问题，然后用 model selection 的方法解决。

1. P316主成分分析算法中为什么要求X每一行元素的均值为0?

我的解答：均值为0时矩阵的协方差矩阵就是相关矩阵，方便计算。

1. **下周计划安排**

看完第十七章并参加讨论。

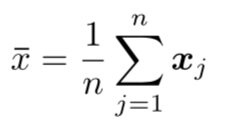
1. **读书收获**

总体主成分分析，是定义在样本总体上的。

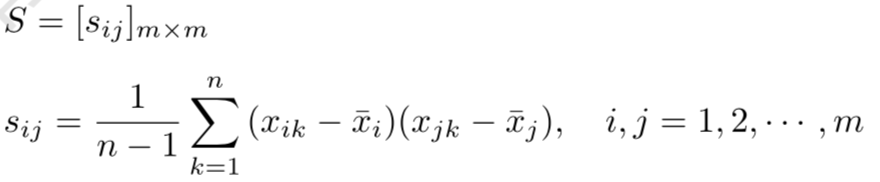
在实际问题中，需要在观测数据上进行主成分分析，这就是样本主成分分析。

样本主成分也和总体主成分具有相同的性质。

给定样本矩阵X，可以估计样本均值，以及样本协方差。样本均值向量为：

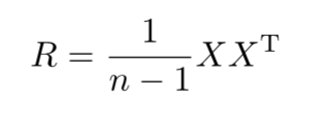


样本协方差矩阵S为：



与总体协方差的不同之处在于总体协方差矩阵除以的是n，样本除以的是n-1，和概率统计里的总体和样本一样，当n足够大时，两者接近。因此样本协方差矩阵S是总体协方差矩阵的近似。

在使用样本主成分时，先将样本数据规范化。此时协方差矩阵就是相关矩阵：



样本协方差矩阵S是总体协方差矩阵∑的无偏估计。

样本相关矩阵R是总体相关矩阵的无偏估计。

S的特征值和特征向量是∑的特征值和特征向量的极大似然估计。

奇异值分解和主成分分析可以相互转化，假设 X’ 的截断奇异值分解为 那么V的列向量就是 的单位特征向量。因此，V的列向量就是X的主成分，于是，求X主成分可以通过求X’的奇异值分解来实现。