统计学习方法读书笔记

61518122 丁自民

**读书内容**

第16章第2节

**读书笔记**

1. 向量x为处理前数据向量，基为默认基。经过主成分分析（乘一个m\*m的矩阵）后，成为向量y。变换后的基为标准正交基。但是可供选择的标准正交基有很多，于是再加上限制条件，就是第一个基要使得样本方差最大。
2. 协方差矩阵是用来描述两个随机向量之间的相关关系的
3. \vec x 的第k主成分是y\_k；也就是说，一个随机向量的第k主成分是一个随机向量。这里相当于在说，第k个标准基。一个多元向量的标准基自然是一个向量。
4. 向量乘法是没有消去律的。以特征向量的定义式为例：Ax = \lambda x。此时，如果两边直接消去x，就错了。这恰恰证明A是特征值\lambda的特征向量。因为不同的线性变换可能会对某个向量产生同样的效果。
5. 期望，分布和方差是用来形容一个随机变量的
6. \vec y的协方差矩阵为对角矩阵：因为任意两个不同的分量协方差为0，自己和自己的协方差为特征值。
7. 如果两个随机变量期望和方差是一样的，他们同分布吗？
8. 样本主成分分析：对一个随机变量，要有n次观测，但每种观测不再是随机变量，而是一个确定对值。

**读书疑问**

“因子负能量”这个量究竟是描述数据哪个特征的？为什么要定义这个量？

**疑问解答**

回答吴亦珂问题1:

后面应该是对主成分分解过程对评价。通过这几步判断主成分分析效果怎么样，能不能符合预期。

如果不能进行主成分分析，可能要用其他方法处理数据

回答吴亦珂问题2:

这里我不同意郑健雄同学对观点，书上特征值分解法也可以用于非方阵。

我想应该是，奇异值分解更加符合要求吧。毕竟做特征值分解对时候，X^TX就已经很像求奇异值对过程了。

回答吴亦珂问题3:

k应该是提前指定，是一个超参数；相当于工厂里“次品率”之类需要控制对指标吧。