统计学习方法读书报告

61518122 丁自民

**读书内容**

1. 第1节

**读书收获与感悟**

1. 主成分分析是利用正交变换的方法把有线性相关表示的观测数据转换为少数几个线性无关的变量。
2. 总体主成分分析

找到线性无关的变量，就是找到这个问题的本质。在后面数据处理时，调整其中一个变量的值，也不会引起其他变量的连带变化（因为他们是线性无关的）

但在将数据进行正交化之前，还需进行规范化，使数据（此时还是线性相关的数据）均值为0，方差为1.

因为有些数据天生体量就比较大，但起伏并不大。例如身高，大多数人的身高都在160cm到190cm之间。而体重则跨度从不足50kg到超过100kg。

所以需要先对数据进行规范化。

1. 线性变换的基应该满足如下性质

自己和自己的内积为1，自己和别人的内积为0.由于之前数据已经标准化处理过了，所以只需要满足后者。

1. 求特征值，特征向量的过程老生常谈。
2. 总体主成分

把随机变量x转化为随机变量y，x是一个向量所以y也是一个向量。

y的每一个元素，都是x的一个线性组合。

这样的话，y的每一个元素，对应一个\alpha向量。

所以，一个y对应一个\alpha矩阵。

1. 最后经过一系列推倒，y的各个分量应为x的协方差矩阵的特征向量。特征值则是y的某个分量的方差

**读书疑问**

为什么要规定因子负能量这个量，有什么用？

**疑问解答**

王然问题1:

因为从图中可以看出，当x1越大时，x2也更加容易取大。

当x2取大时，x1也更容易取大。

王贵涛问题1:

Y1和y2无关。因为不论y取什么值，y2的分布都是对称的，反之亦然。

郑健雄

我觉得，主成分分析和SVM关系不大，但是和奇异值分解倒是很像。从思想上来说，奇异值分解也是找出基，主成分分析也是找基。