读书报告

09118119 黄一凡

# 自己提出的问题

#### 样本的协方差矩阵和相关矩阵分别代表着什么？

#### 这两个都是表示样本各个变量之间的线性相关关系。当均值为0，方差为1时两者一致。这两者是对变量间线性相关关系的一个度量。

# 别人提出的问题

#### 1.主成分分析得出的结果除了可以用于聚类分析还有什么常见应用吗？

主成分分析可以用于数据降维，可作为SVM，线性回归等的数据预处理。数据经过降维后可以提高效率，减少需要的存储空间。

#### 2.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?

对数据进行规范化处理，可以简化后续的计算，提高效率。

#### 3.P315实例中给出了第一、第二主成分代表的含义，但似乎是事后从观察得到的，从实际意义上怎么理解主成分对问题的意义呢？

主成分分析应该是按照重要性给出对于数据最重要的维度或者说因素，所以我们只知道哪些因素是重要的。 而它的涵义应该是我们观察其影响之后赋予的，但是如果首先不知道重要因素，也就说不上赋予意义了。

#### 4.书上介绍了两种求解PCA的方法，为什么现在常用第二种方法，是因为第二种方法效率更高吗？因为一个求解特征值和特征向量，另一个需要进行SVD。

可能主要是因为特征分解本身的局限性，特征分解只能用于方阵，并且必须满足对角化才能使用，而奇异值分解可以用在任何矩阵上。

# 读书计划

#### 本周所读：

《统计学习方法》16.2

#### 下周计划：

《统计学习方法》17