**开学第八次读书报告——郭昊南**

1. **别人提出的问题的理解：**

**问题1：**为什么主成分分析时需要对数据进行规范化，使其方差为1？

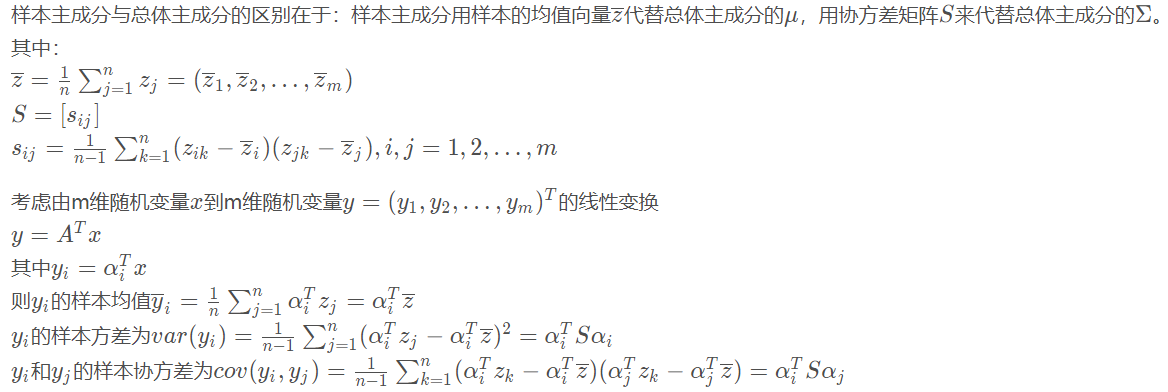
**讨论后的理解：**

主成分分析，是考察多个变量间相关性一种多元统计方法，研究如何通过少数几个主成分来揭示多个变量间的内部结构，即从原始变量中导出少数几个主成分，使它们尽可能多地保留原始变量的信息，且彼此间互不相关.通常数学上的处理就是将原来P个指标作[线性组合](https://baike.baidu.com/item/%E7%BA%BF%E6%80%A7%E7%BB%84%E5%90%88)，作为新的综合指标。这个在神经网络中是为了让数据有着相同的权重，使得某一个属性不要有着太大的影响。此外在梯度下降过程中，数据之间的范围相似，可以加快梯度下降的过程

**问题2：**为什么需要将总体主成分和样本主成分分开讨论？

**讨论后的理解：**

总体主成分是假设我们已经有了这种情况的所有数据以及数据分布的模式和情况，知道了所有的情况，感觉是一种理论情况。样本分析感觉倾向于数据统计，也就是我们并不知道数据分布的所有信息，只能通过随机采样来进行一种近似的分析，偏向于实际的应用，比如显示生活中的数据肯定不太可能对应到数学中的一个具体的分布模式，所以样本分析就是用局部估计总体的情况然后在此基础上进行PCA

****

**问题3：**主成分分析和SVM有什么关联吗？是否可以从奇异值分解这个角度考虑主成分分析是如何推导的？

**讨论后的理解：**

**二、读书计划**

1、本周完成的内容章节：《统计学习方法》第十六章