读书报告

09118119 黄一凡

# 自己提出的问题

#### 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？

因为如果不进行规范化，不同变量的量纲可能相差较大。如果对人的体育进行预估。 跳远以米为单位，数值一般为各位数，而跳绳个数可能有几百个。 而进行主成分分析时，要选取方差最大。那么易知，跳绳个数数值很大，因而方差较大，对结果影响较大。 所以，这就违背了算法的思想。因而要先进行规范化才好操作。

# 别人提出的问题

#### 1.为什么需要将总体主成分和样本主成分分开讨论？

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。

#### 2.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？

主成分分析是找出方差最大的方向，将数据映射到新的空间当中，使得原有数据的维度降低。SVM是找出一个最优平面将两者分开。 SVM的核函数实际上是将原始数据映射到了一个高维空间中，使其变得线性可分。PCA其实是可以由特征值分解实现的。

#### 3.为什么说16.1(a)中的x1和x2是线性相关的?

它们之间的关系可能不是很好给出，但是直观上看，x1的值会影响x2的值， 粗略一点的想可以把这个椭圆看作一条直线，那么x2很明显是等于k\*x1的，也就是有关联的。

# 读书计划

#### 本周所读：

《统计学习方法》16.1

#### 下周计划：

《统计学习方法》16.2