统计学习读书报告

09017244 郑健雄

1. 自己提出的问题的理解：
2. P305页16.22如何理解？其中互不相关为什么可以让公式变形？

讨论后的理解：

此处的不相关说明每个y的分量都可以对x产生独立的影响，所以可以将一个整体的形式变为一个求和形式。而从整体上看，xi本身和y是线性组合关系，所以根据定义，相关性为1，所以最终结果相关性为1。

1. 主成分分析和SVD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：

可以将SVD理解为处理PCA的一种手段，也可以将其看作是PCA本质的思想，因为PCA的分解形式和SVD不谋而合，但PCA也可以通过特征分解来实现，总之，分解是手段，而提取主成分是目的。实际上，最终需要完成的任务就是对X协方差矩阵进行一次奇异值分解，然后按照顺序从大到小排列奇异值和对应的特征向量，也就得出了我们需要的主成分。

1. 别人提出的问题的理解
2. 为什么说16.1(a)中的x1和x2是线性相关的?

自己的理解:

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

1. 满足16.29式子说明什么? 是指使方差最大吗?

自己的理解：

这个问题可以看一下16.26公式，此处实际上是希望迹最大，而根据特征值的大小关系以及16.28，16.27的限制，只有让大的特征值系数尽可能大才能实现这个目标，也就是16.29的表达的意思，因为系数和为q，那么最理想的就是前q个特征值的系数都为1而其余系数都为0.

1. 在下图中，为什么说当知道其中一个变量y1的取值时，对另一个变量y2的预测是完全随机的，y2的取值范围难道不是跟y1有关吗？

自己的理解:

在知道y1的情况下，除了y2的取值范围，我们并没有得到任何其他关于y2的信息，y2的值是完全随机的，而如果是原始图中的x1，x2的话，知道x1，就能知道x2在哪个象限的可能性大，所以x1对x2有着一种预测的作用，但是y1就无法得出y2在哪个部分的概率大这种信息

1. 为什么需要将总体主成分和样本主成分分开讨论？

自己的理解：

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

1. 在主成分分析中，如何保证结果是线性无关的？

自己的理解：

这应该在使用拉格朗日乘数法的时候的条件就已经保证了这一点，每一次的计算都要保证与之前的分解方向线性无关，所以结果也就是线性无关的，这是一个固有属性。

1. 读书计划

1、本周完成的内容章节：复习16章并学习了22章内容

2、下周计划：准备复习一下17章并开始学习下一本书

四、读书摘要及理解

1． 主成分分析：主成分分析的目的是使用少量线性无关的变量来代替线性相关的变量，从而实现数据的压缩以及维度的下降，以此来挖掘数据的基本结构。主成分分析一般分为总体主成分分析和样本主成分分析。

2. 主成分的分析方法：主成分分析最直观的求解方法就是拉格朗日乘数法，其通过逐步分解的过程求解。其是一系列最优化问题，在保证系数向量线性无关的前提下来实现方差最大化，并且新的系数向量要与之前所有的维度都不想关，而经过计算，主成分的方差对应的是从大到小的原始矩阵的特征值。

3. 主成分分析k的确定：根据定理，保存前k个主成分造成的信息损失最小，而k的确定需要根据实际的应用。一般通过累计方差贡献率和针对某个特定变量xi的方差贡献率综合考虑k的值。

4. 矩阵规范化：一般在进行处理前，我们会将随机变量归一化，这样做的好处是引申出很多有用的而结论，比如特征值的和的计算以及特征值是相关矩阵R的特征值等结论。这些结论在是原始条件在特定情况下的计算结果，实际上从原始情况来直接推导也完全成立，但是因为其较为常用，所以额外给出其形式便于使用。

思考：PCA是一种非常经典且常用的数据降维方法，其目标是按照影响力的大小给出主成分，进而给出数据的基本结构。而其可以通过特征分解和奇异值分解来实现。PCA本身的目标就是尽可能多的保留信息，只是其偏重于研究最终保存的主成分向量，但是实际上，特征分解的步骤是一个必须的步骤，因为其会给出计算主成分必须的特征向量，所以可以将PCA理解为SVD的一个应用。此外，PCA在求解特征向量，计算结构等问题上有着很重要的应用价值，是NLP中很常用的方法。