主成分分析法（PCA算法）

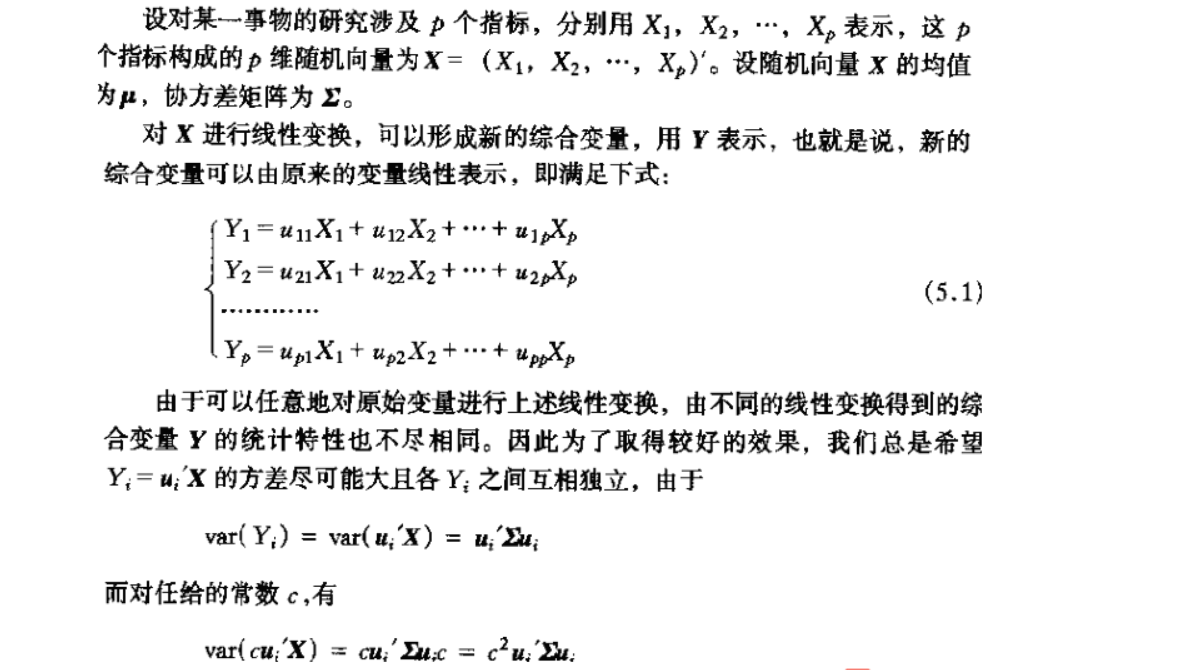
1. 算法概述

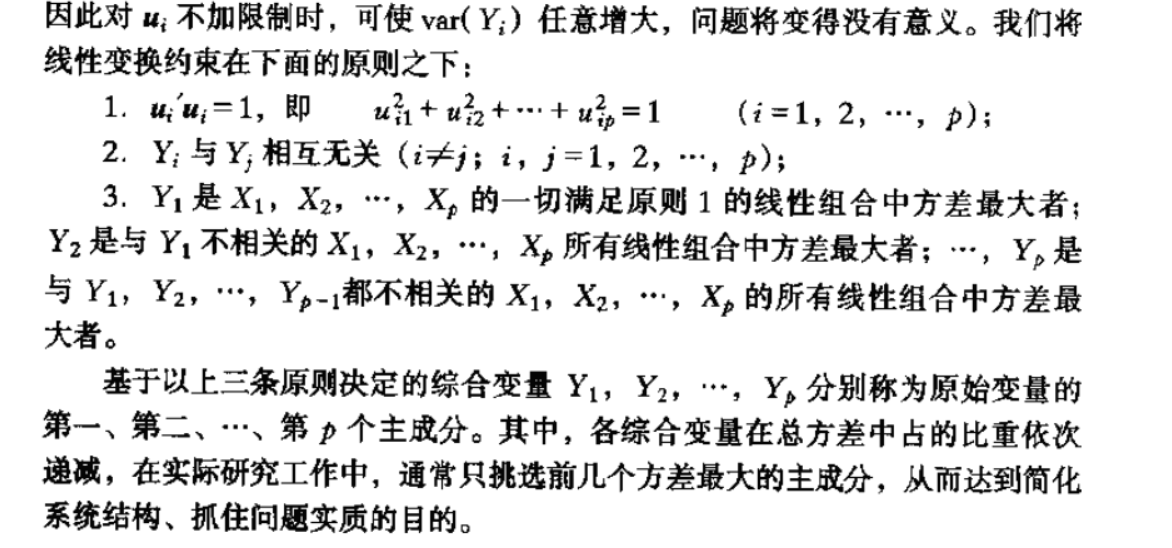
主成分分析(PCA)，主成分分析是一种降维算法，它能将多个指标转换为少数几个主成分，这些主成分是原始变量的线性组合，且彼此之间互不相关，其能反映出原始数据的大部分信息。一般来说，当研究的问题涉及到多变量且变量之间存在很强的相关性时，我们可考虑使用主成分分析的方法来对数据进行简化。

1. 算法目的：

主成分分析是把原来多个变量划为少数几个综合指标的一种统计分析方法。从数学角度来看，这是一种降维处理技术。这是要在力保数据信息丢失最小的原则下, 对高维变量空间进行降维处理; 即在保证数据信息损失最小的前提下, 经线性变换和舍弃小部分信息, 以少 数的综合变量取代原始采用的多维变量。也可用来解决多重共线性的问题

1. 算法原理：





4.算法实现步骤：

1.首先对数据进行标准化。

2.计算样本协方差矩阵。

3.计算R的特征值和特征向量。

4.计算主成分贡献率和累计贡献率。

5.计算我们所需要的值。

1. 算法优缺点

优点：作为一种降维的处理技术，它的优点是显而易见的，它可以帮我们在高维度的数据（指标太多）保留下最重要的一些特征，去除噪声和不重要的特征，从而实现提升数据处理速度

的目的。以此来为我们节省大量的时间和成本，并使结果更容易被人理解，是一种应用非常广泛的数据预处理方法。

缺点：主成分的解释其含义一般多少带有点模糊性，不像原始变量的含义，那么清楚、确切，而且主成分的解释其含义一般多少带有点模糊性，不像原始变量的含义那么清楚、确切，这些也是变量降维过程中不得不付出的代价。

参考文献

[1]王学民.应用多元分析 (第二版) [M].上海:上海财经大学出版社, 2004.

[2]王学民.对主成分分析中综合得分方法的质疑[J].统计与决策,2007(08):31-32.