

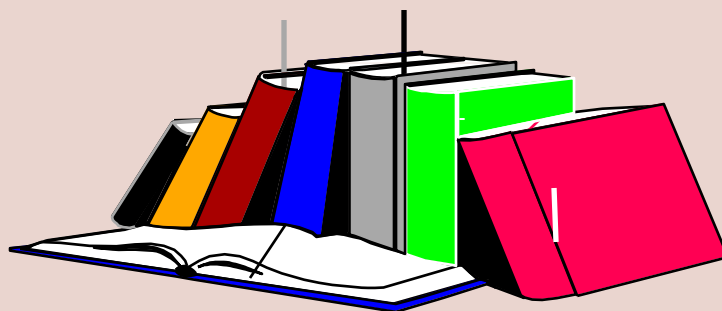
第10章 主成分分析

本章教学目标:

掌握主成分分析的基本概念、基本原理及其分析应用的基本步骤;

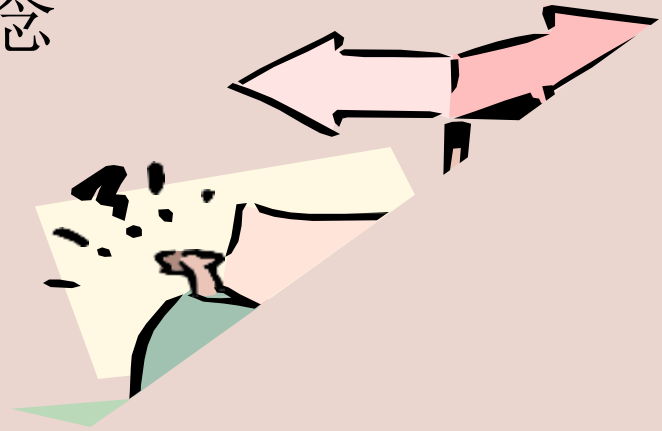
掌握使用软件进行主成分分析;

能应用主成分分析方法解决实际问题



本章主要内容:

- § 10.1 主成分分析的概念
- § 10.2 主成分计算
- § 10.3 样本主成分
- § 10.4 主成分的选取
- § 10.5 案例分析
- § 10.6 上机实现



本章内容重点:

主成分分析的概念; 主成分的选取; 上机实现

§ 10.1 主成分的概念

在实证数据分析研究中，人们为了尽可能完整地搜集信息，对于每个样本往往要观测它的很多项指标，少者四、五项，多则几十项，这些指标之间通常不是相互独立而是相关的。因此，从统计分析或推断的角度来说，人们总是希望能把大量的原始指标组合成较少的几个综合指标，从而使分析简化。

例如，一个人的身材需要用好多项指标才能完整地描述，诸如身高、臂长、腿长、肩宽、胸围、腰围、臀围等，但人们购买衣服时一般只用长度和肥瘦两个指标就够了，这里长度和肥瘦就是描述人体形状的多项指标组合而成的两个综合指标。

假设观测指标共有 p 个，分别用 x_1, x_2, \dots, x_p 表示。将这些指标综合为一个指标的方法显然有很多，但最简单的方法是将这些指标用线性组合的方法将它们组合起来。因此，可设定其综合指标的形式为这些指标的线性组合，即

$$y = a^T x = a_1 x_1 + a_2 x_2 + \dots + a_p x_p$$

我们希望构造少数几个这样的综合指标，并且这几个综合指标之间是不相关的。

其中反映原始观测指标的变动程度最大的综合指标最重要，我们称其为原始观测指标的第一主成分；而反映原始观测指标变动程度次大的综合指标，称为原始观测指标的第二主成分；反映原始观测指标变动程度第三大的综合指标，称为第三主成分；……，

§ 10.2 主成分计算

设 p 个指标 (随机变量) 是 p 维随机变量, 其协方差矩阵为

$$\Sigma_{p \times p} = \text{COV}(\mathbf{X}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}$$

式中, 协方差 $\sigma_{ij} = E\{[X_i - E(X)]$

限制 $\mathbf{a}^{(1)\text{T}} \mathbf{a}^{(1)} = 1$ 求 $\text{COV}(\mathbf{a}^{(1)\text{T}}$ 的最大值。

实际上, 这就是求的 $\max_{\mathbf{a} \neq 0} \frac{\mathbf{a}^{(1)\text{T}} \Sigma \mathbf{a}^{(1)}}{\mathbf{a}^{(1)\text{T}} \mathbf{a}^{(1)}}$ 的值

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/805143343222011123>