

主成分分析法案例

汇报人：xxx

20xx-03-19



目录

contents

- 引言
- 数据预处理
- 主成分分析过程
- 结果展示与解释
- 案例应用与讨论
- 结论与总结

01

引言





背景与目的

随着大数据时代的到来，数据处理和分析变得日益重要。主成分分析（PCA）作为一种常用的数据分析方法，被广泛应用于各个领域。

背景

目的

本案例旨在通过具体实例，详细阐述PCA方法的原理、步骤和应用，以便读者更好地理解 and 掌握这一技术。



主成分分析法简介

PCA算法定义

主成分分析（PCA）是一种统计方法，通过正交变换将一组可能存在相关性的变量转换为一组线性不相关的变量，转换后的这组变量叫主成分。

PCA算法原理

PCA利用降维的思想，在损失很少信息的前提下把多个指标转化为几个综合指标的多元统计方法。通常把转化生成的综合指标称之为主成分，其中每个主成分都是原始变量的线性组合，且各个主成分之间互不相关，这就使得主成分比原始变量具有某些更优越的性能。

PCA算法应用

PCA在数据分析、机器学习、图像处理等领域具有广泛应用，如数据降维、特征提取、异常检测等。



案例选择与数据来源

案例选择

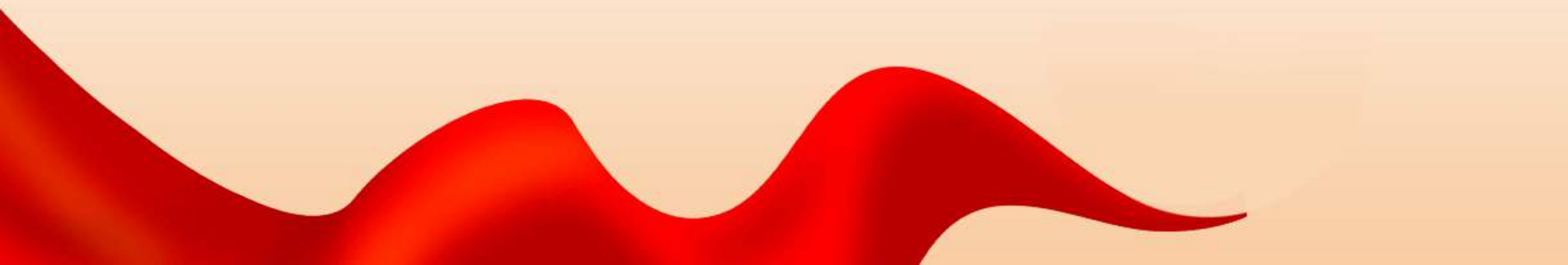
本案例选择了某电商平台的销售数据进行分析。该数据集包含了多种商品的销售量、价格、评价等信息，是一个具有多维特征的数据集。

数据来源

数据来源于该电商平台的公开数据接口，经过预处理和清洗后，得到了规范化的数据格式，便于进行后续的PCA分析。

02

数据预处理

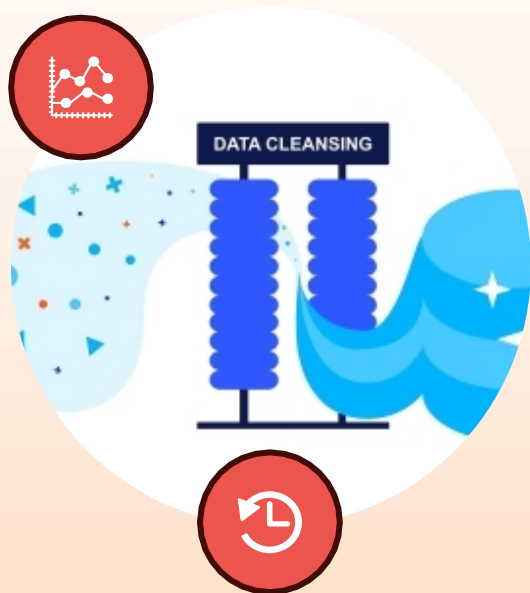




数据清洗与整理

去除重复数据

检查数据集中是否存在重复的行或记录，并予以删除。

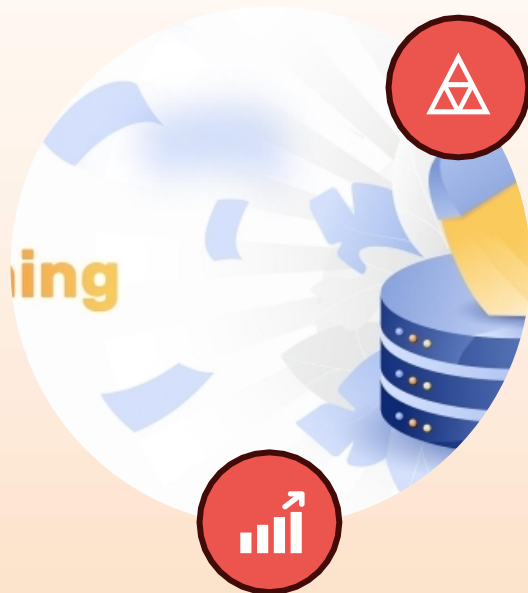


纠正错误数据

对数据中的错误进行识别和纠正，如拼写错误、逻辑错误等。

数据类型转换

将非数值型数据转换为数值型数据，以便进行后续的数学计算。

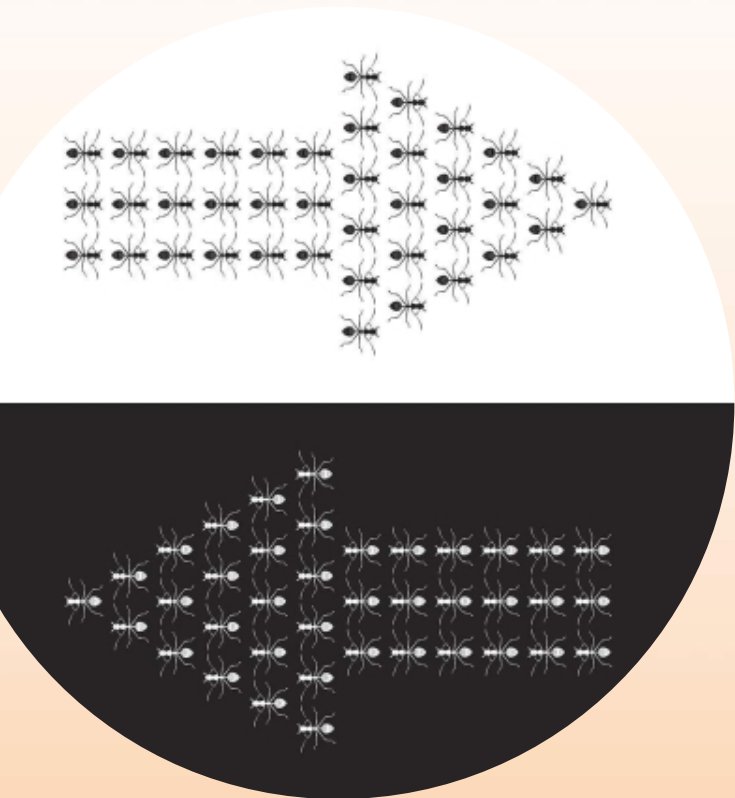


数据排序与分组

根据需要对数据进行排序或分组，以便更好地理解和分析数据。



缺失值处理



删除缺失值

将含有缺失值的行或列从数据集中删除。

填充缺失值

根据数据的分布和特性，选择合适的填充方法，如均值填充、中位数填充、众数填充等。

插值法

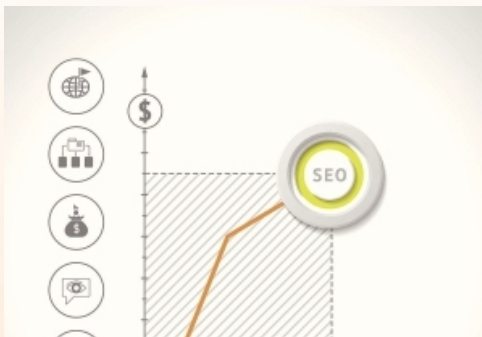
利用已知数据点估算缺失值，如线性插值、多项式插值等。

预测模型

建立预测模型来估算缺失值，如回归模型、决策树等。



异常值检测与处理



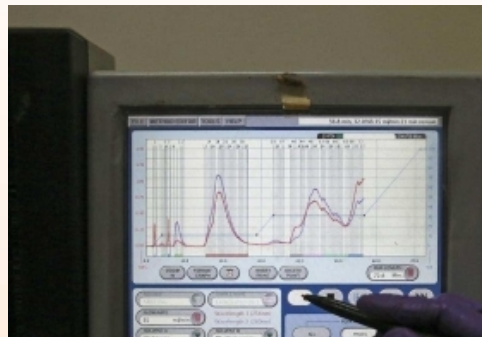
统计方法

利用统计学原理，如 3σ 原则、箱线图等，识别出数据中的异常值。



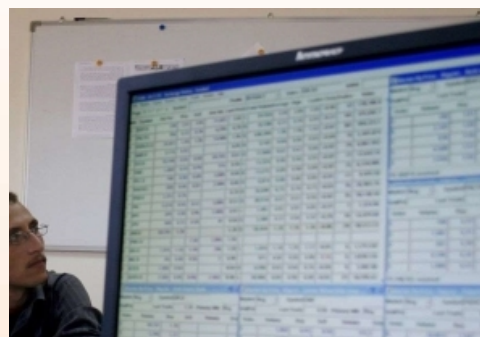
机器学习方法

利用机器学习算法，如孤立森林、DBSCAN等，检测数据中的异常值。



视觉化方法

通过绘制散点图、直方图等图形，直观地观察数据中的异常值。



处理异常值

根据实际需求，选择删除异常值、替换异常值或保留异常值并进行分析。

数据标准化与归一化

标准化

将数据转换为均值为0、标准差为1的分布，消除不同特征之间的量纲差异。



Sequence Analy

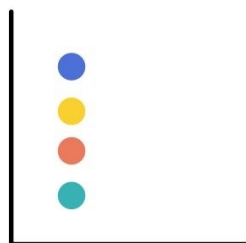


归一化

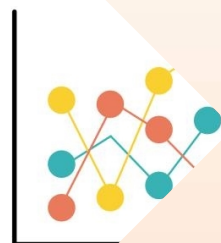
将数据缩放到[0,1]或[-1,1]的范围内，使不同特征之间具有可比性。



Time Series data



Cross-Sectional data



Network data

小数定标标准化

通过移动数据的小数点位置来进行标准化处理。

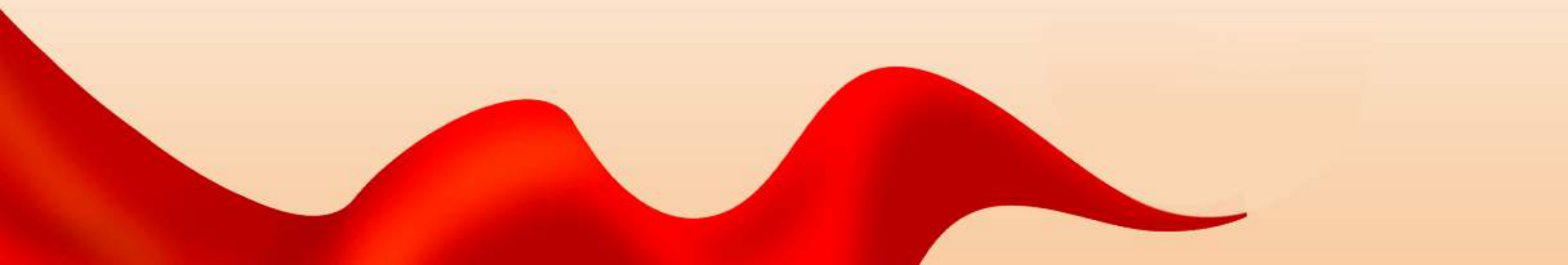


非线性归一化

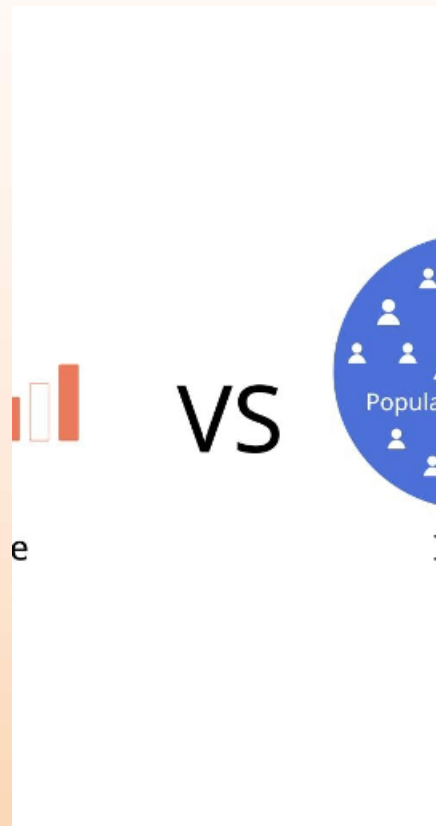
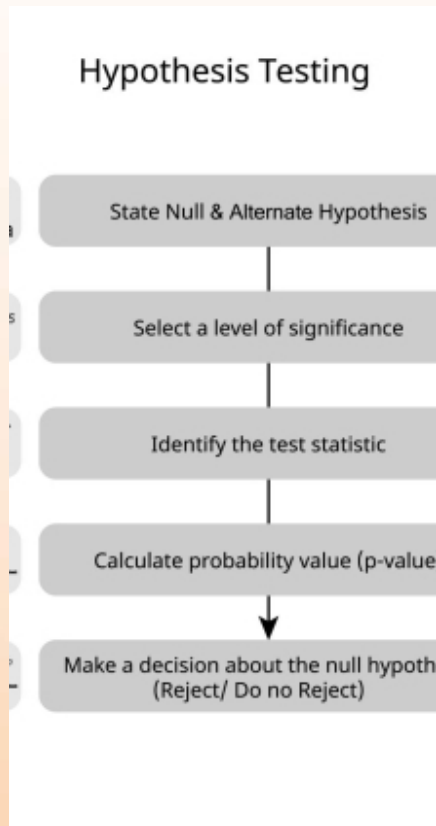
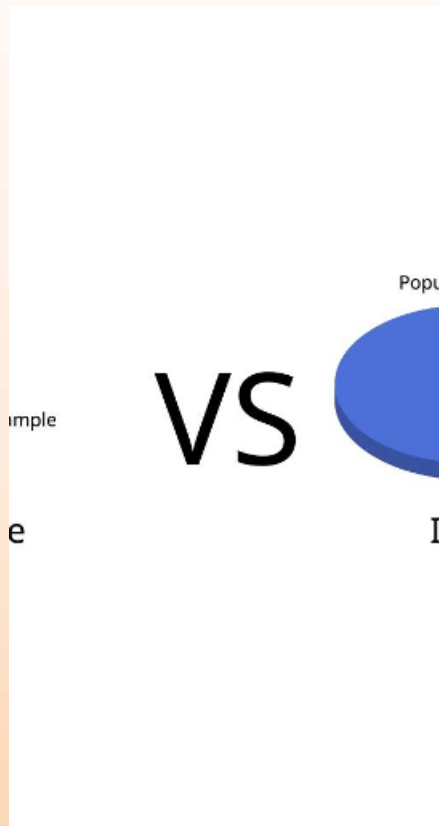
对于某些非线性分布的数据，可以采用非线性变换方法进行归一化处理。

03

主成分分析过程



协方差矩阵计算



标准化原始数据

将原始数据进行标准化处理，以消除量纲的影响。



计算协方差矩阵

根据标准化后的数据计算协方差矩阵，以反映各指标之间的相关程度。



特征值与特征向量求解

求解特征方程

通过求解协方差矩阵的特征方程，得到特征值和特征向量。

特征值排序

将特征值按大小进行排序，以便确定主成分的顺序。



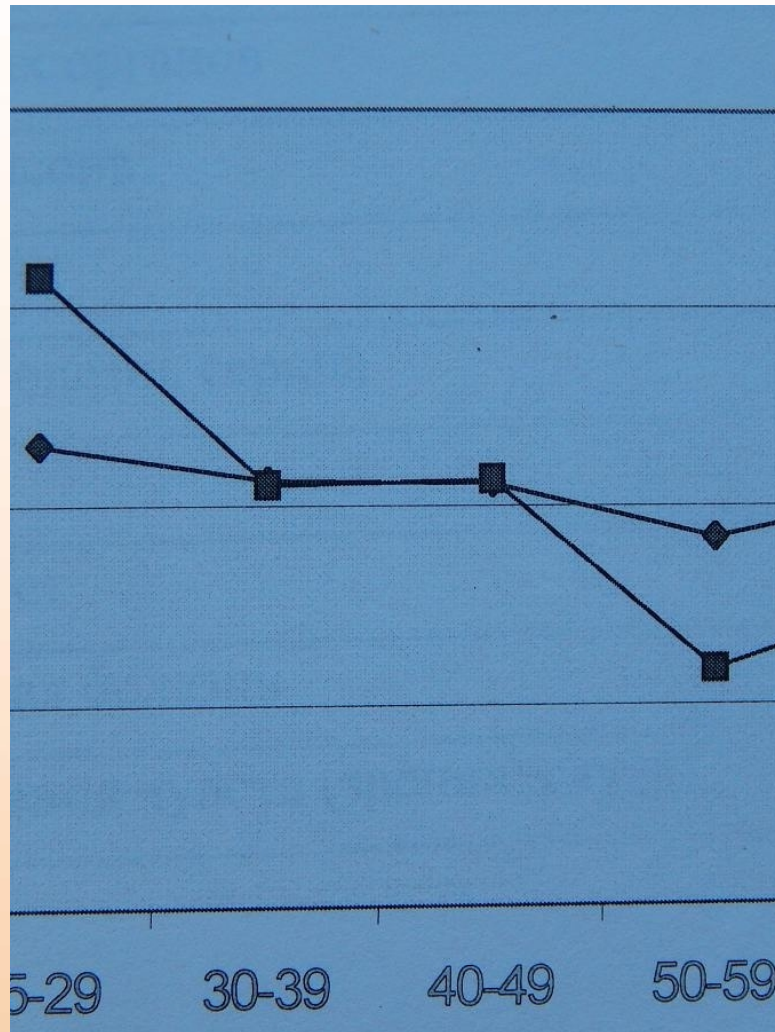
主成分个数确定

累计贡献率原则

根据特征值的累计贡献率来确定主成分个数，通常选取累计贡献率达到85%以上的前几个主成分。

碎石图判断

通过绘制碎石图来直观判断主成分个数，当曲线变得平缓时，说明后续的主成分贡献较小，可以考虑舍去。



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/625202243212011243>