

论文题目： 高维数据降维方法及应用研究

高维数据降维方法及应用研究

摘 要

在这个数据大爆发的时代，数据的高效充分的利用对于我们生产水平的提高和科技的进步有巨大的帮助。而当下的数据无论是量的级别还是维数的个数都与之前的数据存在很大的不同，本文主要以数据维度的角度来分析高维数据处理的一个重要方面——数据降维。

本文首先回顾了高维数据降维方面的各种降维算法并给予简单描述，并对主成分分析(PCA)、lasso 回归、核主成分分析(KPCA)、局部线性嵌入 (LLE) 这四个算法进行详细的介绍。然后通过数据进行实验，探究各个降维算法降维后样本在降维后空间的分布特性，发现在对 Boston 房价数据集的降维过程中运用局部线性嵌入法降维后数据分布能体现数据的内在结构，而主成分分析法的结果不是很好。接着从时间复杂度和空间复杂度方面分析这四个算法的复杂度，并测试算法的运行时间进行比较，从实际的运行结果中可以看出主成分分析法的运行时间最短。

数据降维是数据处理的一个重要步骤也是机器学习的一个重要课题，数据降维在模式识别、人工智能等领域都有着举足轻重的作用，因此对于数据降维的研究是十分重要并且很有意义的。

关键词：数据降维；主成分分析；lasso 回归；核主成分分析；局部线性嵌入

Commented [I1]: 一段就行

Commented [I2]: 一段就行

Abstract

In this era of data explosion, the efficient and full use of data is of great help to the improvement of our production level and the progress of science and technology. However, the current data is very different from the previous data in terms of quantity level and number of dimensions. This article mainly analyzes an important aspect of high-dimensional data processing from the perspective of data dimensions-data dimension reduction.

This article first reviews various dimensionality reduction algorithms for high-dimensional data dimensionality reduction and gives a brief description, and gives a detailed introduction to the four algorithms of principal component analysis, lasso regression, kernel principal component analysis, and local linear embedding. Then conduct experiments through the data to explore the distribution characteristics of the samples in the dimensionality reduction space after dimensionality reduction by various dimensionality reduction algorithms, and find that in the dimensionality reduction process of the Boston housing price data set, the local linear embedding method is used to reduce the data distribution. Internal structure, and the results of principal component analysis are not very good. Then analyze the complexity of these four algorithms in terms of time complexity and space complexity, and compare the running time of the tested algorithms. From the actual running results, it can be seen that the running time of the principal component analysis method is the shortest.

Data dimensionality reduction is an important step in data processing and an important subject of machine learning. Data dimensionality reduction plays a pivotal role in the fields of pattern recognition and artificial intelligence. Therefore, the research on data dimensionality reduction is very important and meaningful .

Keywords: dimensionality reduction, PCA , lasso regression ,KPCA ,LLE

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/425031332040011341>

