10. Tasks in Machine Learning

Contents:

Dimensionality Reduction

School of Electronic and Computer Engineering
Peking University

Wang Wenmin

What is Dimensionality Reduction 什么是降维

- A longer description 较长描述
  To transform an initial very high-dimensional representation of data into a lower-dimensional representation of these data while preserving some properties of the initial representation.
  将初始的极高维数据表示转换为这些数据的低维表示，而保留原始表示的某些性质。
- A shorter description 较短描述
  To simplify inputs by mapping high-dimensional space into a lower dimensional representation.
  通过将高维空间映射到低维空间表示来简化输入。
- A very short description 极简描述
  To map inputs into a lower dimensional space.
  将输入映射到低维空间。

Contents:

Why Dimensionality Reduction 为什么降维

- Curse of dimensionality 维度灾难
  - This phenomena arises when analyzing data in high-dimensional spaces.
    当在高维空间对数据进行分析时，该现象就会发生。
- Data sparsity or irrelevant 数据稀疏或无关
  - When the dimensionality increases, the volume of the space increases so fast that the available data become sparse.
    随着维度的增加，空间的体积增长非常迅速，使得可用的数据变得稀疏。
  - Some features may be irrelevant.
    某些特征可能是无关的。
- Visualization 可视化
  - The data with two or three dimensions is easy to represent.
    二维或三维数据易于表示。

Contents:

Linear and Nonlinear 线性与非线性

- Linear Dimensionality Reduction 线性降维
  - performs a linear mapping high-dimensional input data to a lower dimensional space.

    采用某种线性方式将高维输入数据映射到低维空间。
- Nonlinear Dimensionality Reduction 非线性降维
  - performs a nonlinear mapping high-dimensional input data to a lower dimensional space.

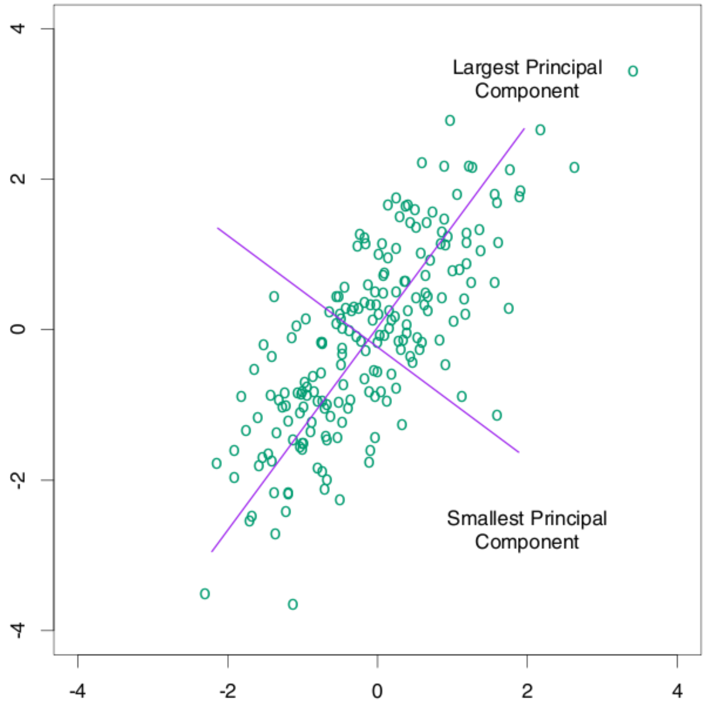
    采用某种非线性方式将高维输入数据映射到低维空间。

Typical Methods of Linear Dimensionality Reduction 线性降维的典型方法

- Principal Component Analysis (PCA) 主成分分析 (PCA)
- Linear Discriminate Analysis (LDA) 线性判别分析 (LDA)
- Multilinear subspace learning 多线性子空间学习
  - Multilinear Principal Component Analysis (MPCA) 多线性主成分分析
  - Multilinear Linear Descriminant Analysis (MLDA) 多线性线性判别分析

Example: Principal Component Analysis (PCA) 主成分分析 (PCA)

- PCA is a statistical procedure.
  PCA是一种统计过程。
- It uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.
  它采用用正交变换方法将一组可能相关变量的观测值转换成一组称为主成分的线性不相关变量值。
- The number of principal components is less than the number of original variables.
  主成分的数量小于原始变量的数量。

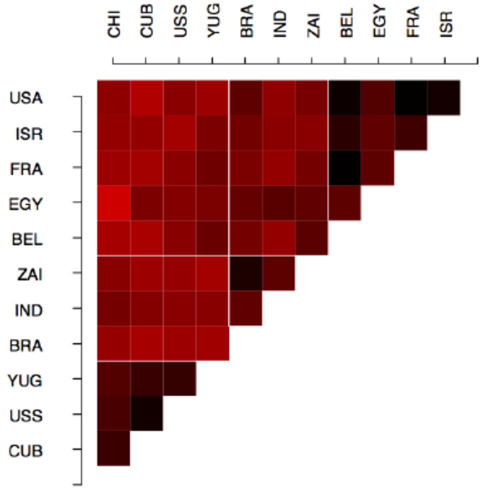


Approaches of Nonlinear Dimensionality Reduction 非线性降维的方法

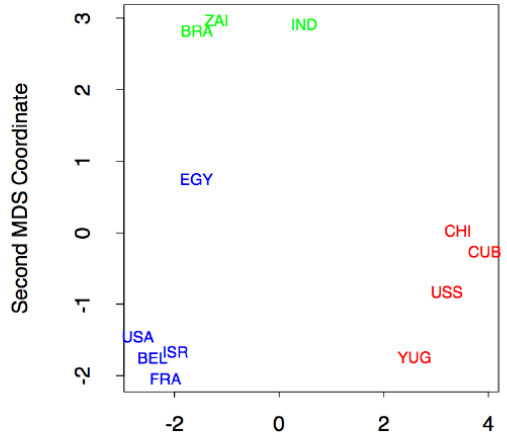
Multi-dimensional Scaling 多元尺度分析	Kernel approaches 核方法	Manifold learning approaches 流形学习方法
• Classical multidimensional scaling 经典多元尺度分析 • Metric multidimensional scaling 度量多元尺度分析 • Non-metric multidimensional scaling 非度量多元尺度分析 • Generalized multidimensional scaling 广义多元尺度分析	• Kernel Principal Component Analysis 核主成分分析 • Kernel Fisher Discriminant Analysis (KFD) 核费希尔判别分析	• Isometric feature mapping (Isomap) 等距特征映射 • Locally-linear embedding (LLE) 局部线性嵌入

Example: Multi-dimensional Scaling (MDS) 多元尺度分析(MDS)

- MDS is a set of related statistical techniques often used in data visualisation for exploring similarities or dissimilarities in data.
  MDS是用于数据可视化的一些相关统计技术，用来考察数据中的相似性和非相似性。
- An MDS algorithm takes a matrix of pair-wise distances between all points, then computes a position for each point in a low-dimensional space, suitable for 2D or 3D visualisation.
  MDS算法构建一个所有点之间的成对距离矩阵，然后在低维空间计算每个点的位置，便于二维或三维可视化。



Recorded Dissimilarity Matrix
录得的非相似矩阵

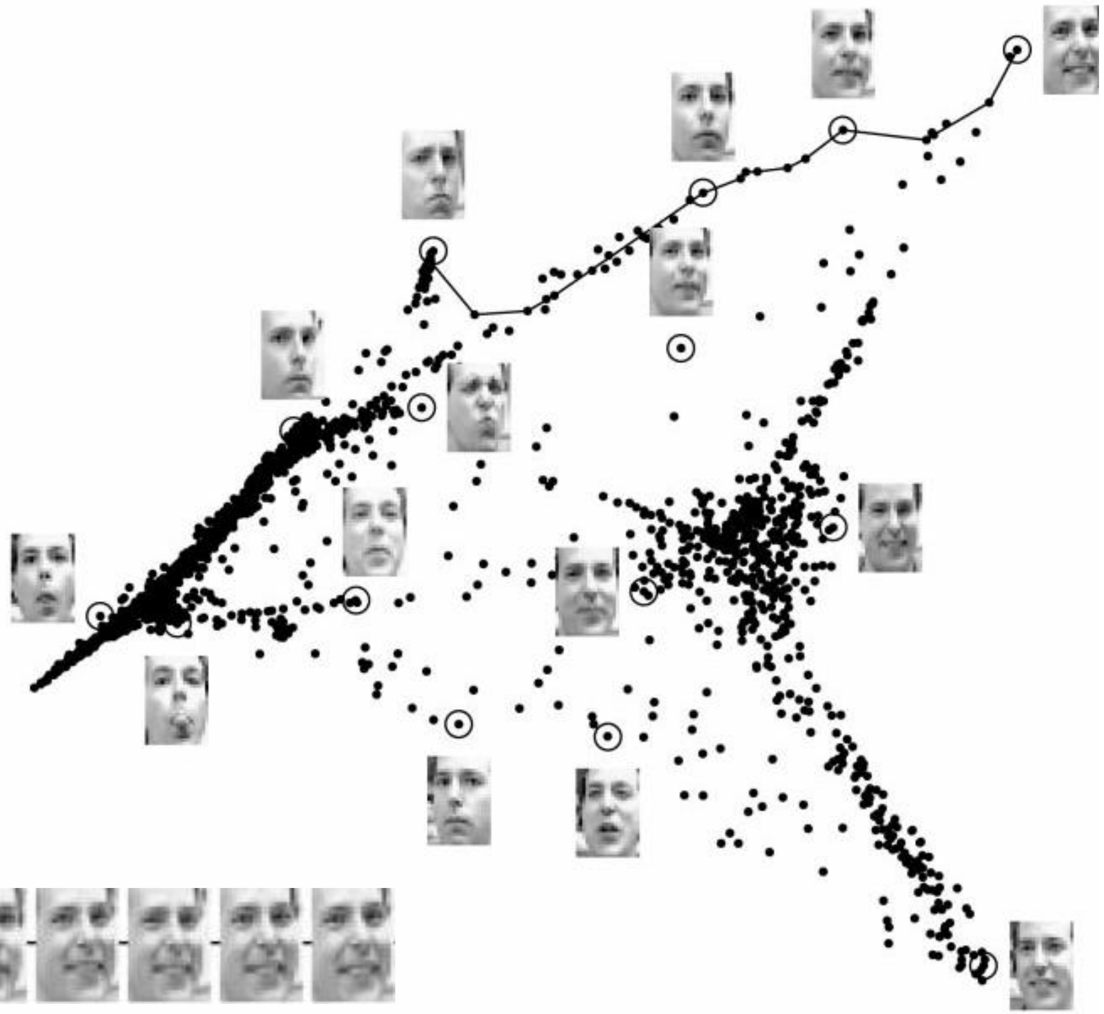


First MDS Coordinate
第一个MDS坐标

Contents:

Typical Applications of Dimensionality Reduction 降维的典型应用

- Image processing
  图像处理
- Face recognition
  人脸识别
- Handwriting recognition
  手写体识别
- Gene expression profiles
  基因表达谱
- etc.



Source: Science, vol. 290, Dec. 22, 2000.

Thank you for your attention!

AI