Dimensionality Reduction in Non-linear Data

Janani Ravi
CO-FOUNDER, LOONYCORN
www.loonycorn.com

Overview

Manifold learning for dimensionality reduction

Implementations of manifold learning

Applying manifold learning to image data

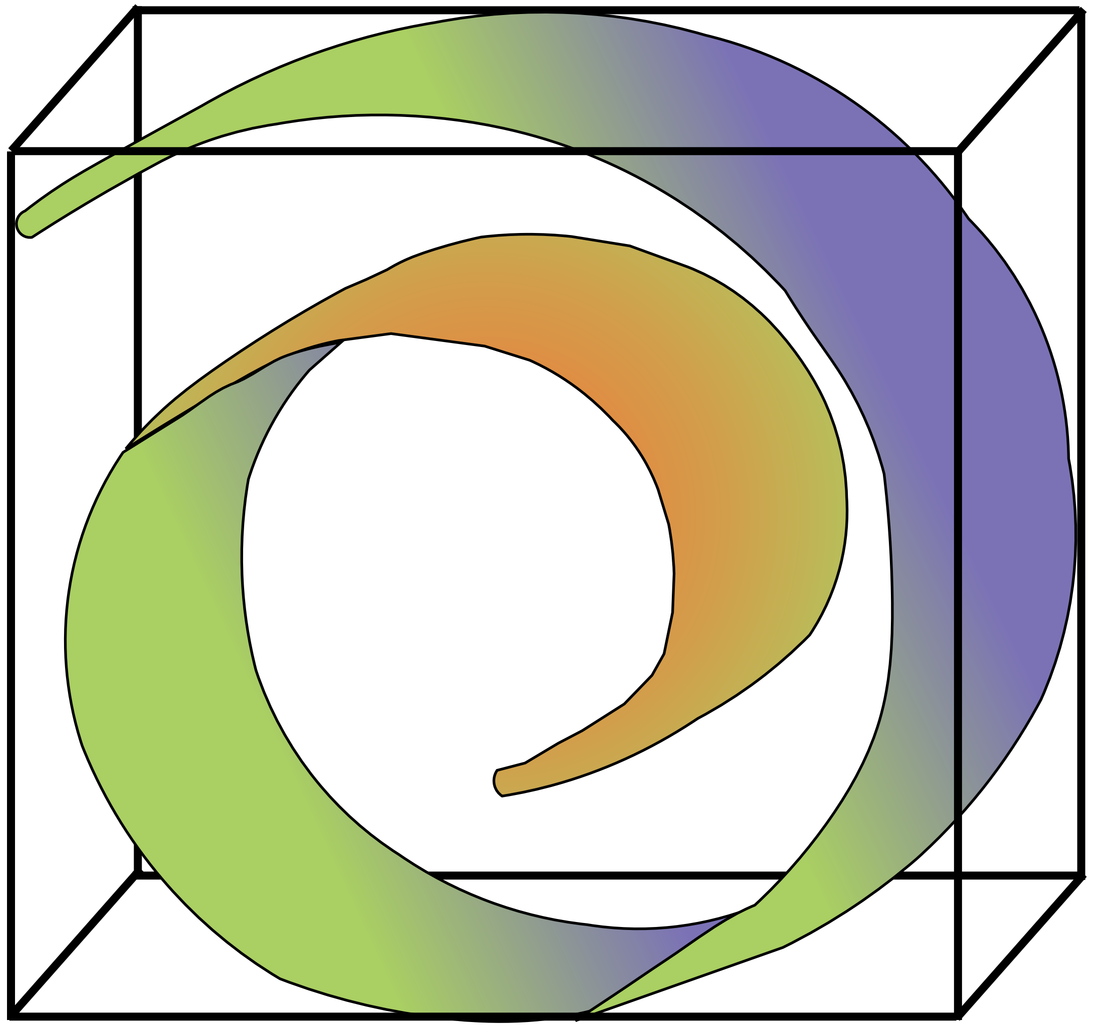
Manifold Learning

Manifold Hypothesis:
Very complicated data is often not that complicated after all

Choosing Manifold Learning

Use Case	Possible Solution
Y not linearly related to X	Manifold learning
Very high dimensionality of X (e.g. pixel counts in image data)	
Many constraints on allowable values of X-variables (sparse features)	
Three-dimensional plots of Y against pairs of X indicate manifold shape	

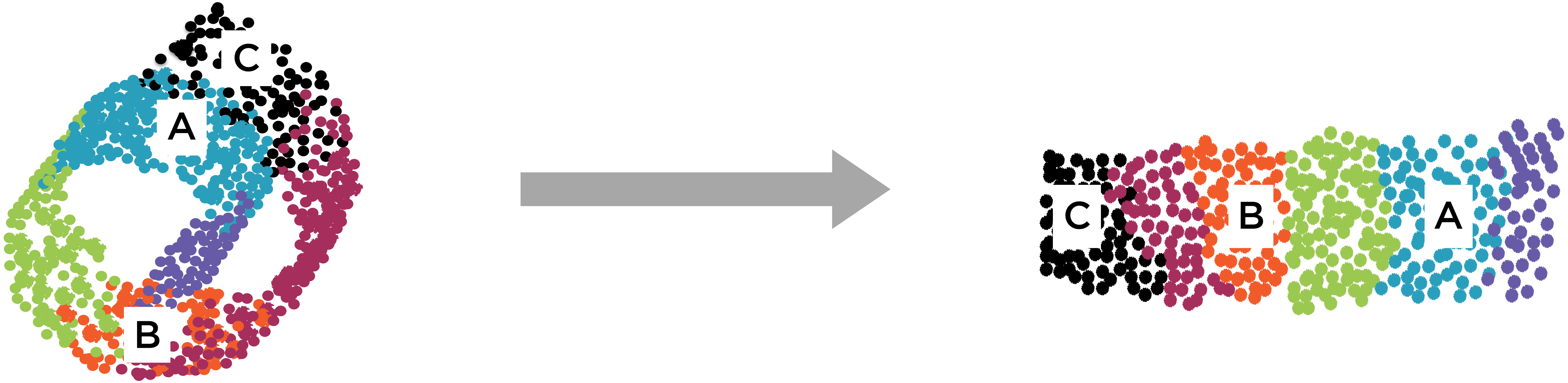
Manifold Data



Manifold Hypothesis

Many high-dimensional datasets can be easily unrolled so that they lie along a much lower dimensional manifold

Manifold Hypothesis



High-dimensional data

Manifold Learning

Low-dimensional embedding

Manifold Learning Techniques

MDS	Isomap	Kernel PCA
LLE	t-SNE	Spectral Embedding

Multidimensional Scaling (MDS)

Aims to preserve pair-wise Euclidean distances between all points while reducing dimensionality. Some intuitive similarities to MSE regression in underlying math.

Isomap

Aims to preserve pair-wise Euclidean distances between neighboring points only (not all points) while reducing dimensionality; works out equivalent to preserving geodesic distance between all points.

Locally Linear Embedding

Expresses each point as centroid (weighted average) of nearest neighbors; then tries to maintain same weights upon conversion to new dimensions.

Spectral Embeddings

Builds a graph where each point serves as a node; then fits a smooth function in lower dimensional space to pass through all nodes. Often implemented using technique called Laplacian Eigenmaps.

t-distributed Stochastic Neighbor Embedding (t-SNE)

Aims to keep similar points together and dissimilar points apart. First fits a Student-t probability distribution to the data, hence the name. Widely used in visualizing clusters.

Demo

Implement Manifold Learning techniques on S-curve data

Demo

Implement Manifold Learning techniques on simple image data using the handwritten digits dataset

Demo

Implement Manifold Learning techniques on complex image data using the Olivetti faces dataset

Summary

Manifold learning for dimensionality reduction

Implementations of manifold learning

Applying manifold learning to image data

Related Courses

Building Clustering Models with scikit-learn

Building Regression Models with scikit-learn

Building Neural Networks with scikit-learn