STA 5106
Computational Methods in Statistics I

Department of Statistics
Florida State University

Class 8
September 19, 2019

Special Topic 1

Nonlinear Dimensionality Reduction by Locally Linear Embedding

Reference:
"Nonlinear dimensionality reduction by locally linear embedding," Roweis & Saul, **Science**, 2000.
"Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds, " Lawrence K. Saul, Sam T. Roweis, JMLR, 2003.

Nonlinear Dimensionality Methods

- **Principal Component Analysis (1901)**
- Sammon Mapping (1969)
- Kernel Principal Component Analysis (1998)
- Isomap (2000)
- **Locally-linear embedding (2000)**
- Laplacian Eigenmaps (2001)
- Local Tangent Space Alignment (2005)
- Non-linear PCA (2005)
- Local Multidimensional Scaling (2006)
- Manifold Sculpting (2008)
- … ...

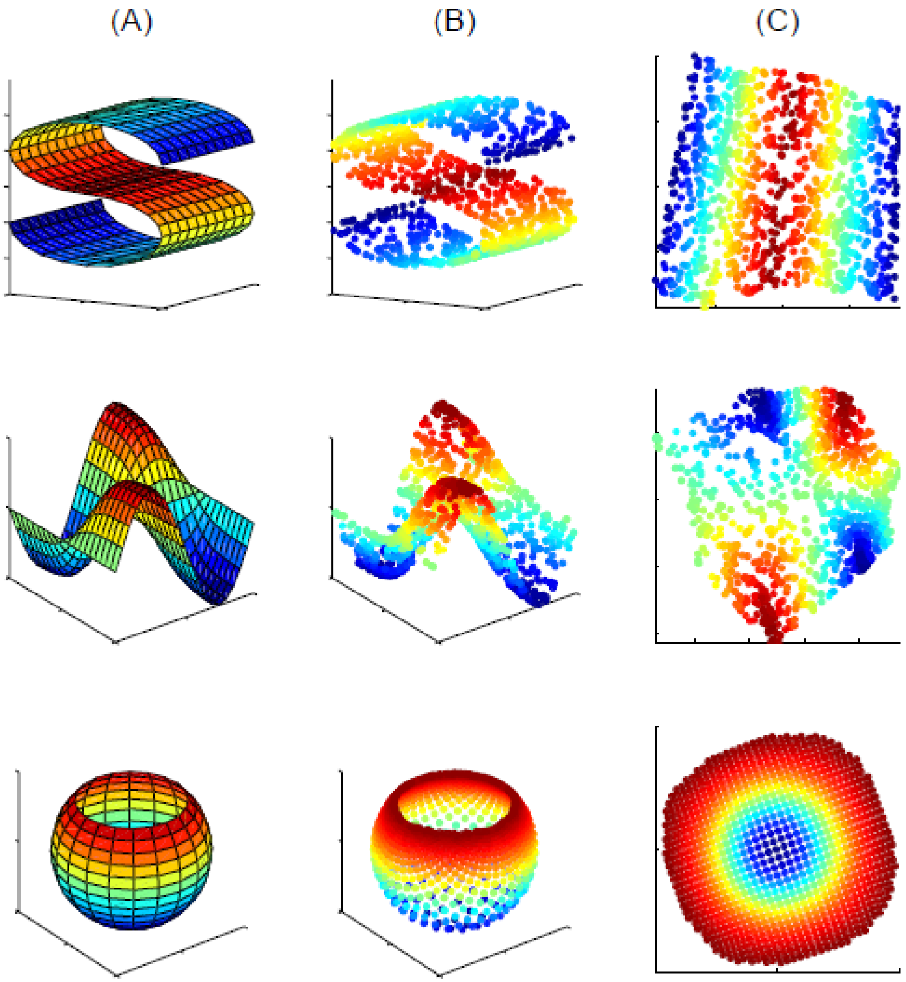
Introduction

- How do we judge similarity?
  - Our mental representations of the world are formed by processing large numbers of sensory inputs— including, for example, the pixel intensities of images, and the power spectra of sounds.

- While complex stimuli of this form can be represented by points in a high-dimensional vector space, they typically have a much more compact description.

Nonlinear Dimensionality Reduction

- An unsupervised learning algorithm must discover the global internal coordinates of the manifold without external signals that suggest how the data should be embedded in higher dimensions.

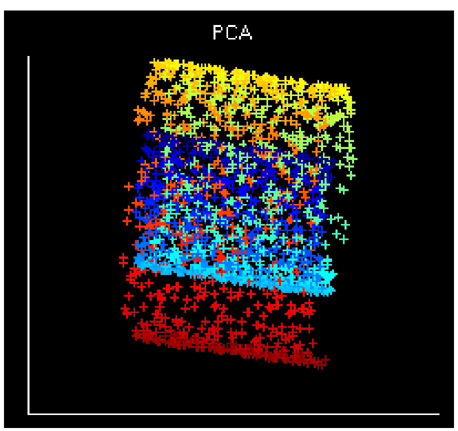


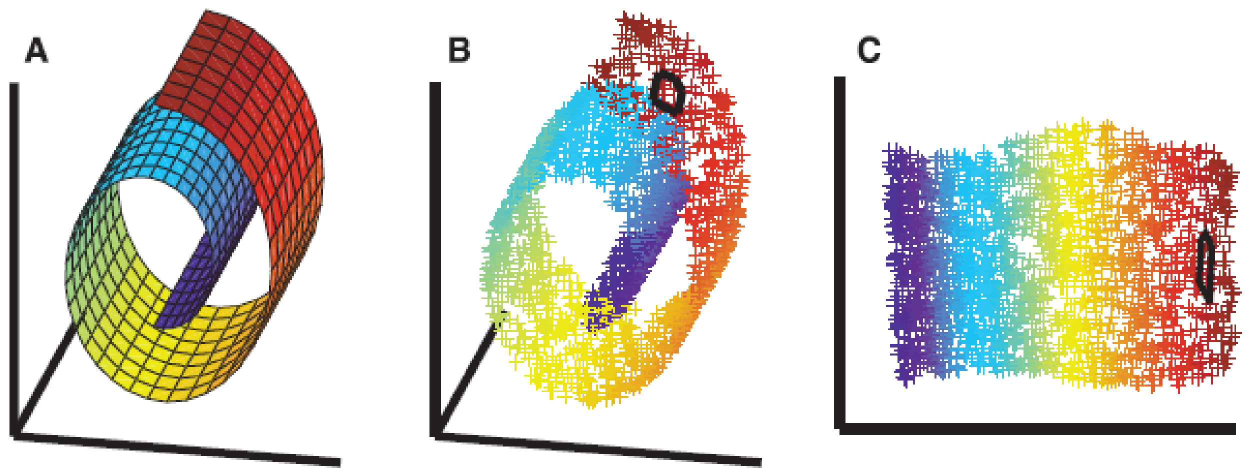
Basic Idea

- In mathematics, a **manifold** is a mathematical space that on a small enough scale resembles the Euclidean space of a specific dimension.
- The concept of manifolds allows more complicated structures to be expressed and understood in terms of the relatively well-understood properties of simpler spaces.
- **Locally Linear Embedding (LLE)** recovers global nonlinear structure from locally linear fits.
- The problem involves mapping high-dimensional inputs into a low dimensional “description” space with as many coordinates as observed modes of variability.

Basic Idea

- The color coding illustrates the neighborhood preserving mapping discovered by LLE.
- Unlike LLE, the projection of the data by principal component analysis (PCA) maps faraway data points to nearby points in the plane, failing to identify the underlying structure of the manifold.





Step 1: Neighborhood Search

- Suppose the data consist of N real-valued vectors $\vec{X}_i$ each of dimensionality D, sampled from some underlying manifold.

- We expect each data point and its neighbors to lie on or close to a locally linear patch of the manifold.

- One can identify neighbors by choosing all points within a ball of fixed radius.

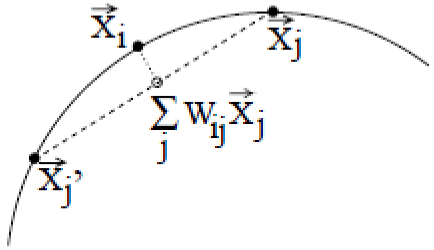
Step 2: Least Squares Fits

- The second step of LLE is to reconstruct each data point from its nearest neighbors.

- Reconstruction errors are measured by the cost function

$$\varepsilon(W) = \sum_i \left| \vec{X}_i - \sum_j W_{ij} \vec{X}_j \right|^2 \qquad (1)$$

which adds up the squared distances between all the data points and their reconstructions.



- Here the rows of the weight matrix sum to one: $\sum_j W_{ij} = 1$

Least-Squares Problem

- For a point $\vec{X}$ with K neighbors $\vec{\eta}_j$, we minimize the error

$$\left|\vec{X} - \sum_{j=1}^{K} W_j \vec{\eta}_j\right|^2 = \left|\sum_{j=1}^{K} W_j \left(\vec{X} - \vec{\eta}_j\right)\right|^2 = \sum_{jk} W_j W_k G_{jk}$$

- We have introduced the “local” Gram matrix

$$G_{jk} = \left(\vec{X} - \vec{\eta}_j\right) \cdot \left(\vec{X} - \vec{\eta}_k\right)$$

- By construction, this Gram matrix is symmetric.
- The reconstruction error can be minimized analytically using a Lagrange multiplier to enforce the constraint that $\sum_j W_j = 1$
- The optimal reconstruction weights are

$$W_j = \frac{\sum_k [G^{-1}]_{jk}}{\sum_{lm} [G^{-1}]_{lm}}$$

Step 3: Eigenvalue Problem

- Each high-dimensional observation $\vec{X}_i$ is mapped to a low-dimensional vector $\vec{Y}_i$ representing global internal coordinates on the manifold.

- This is done by choosing *d*-dimensional coordinate $\vec{Y}_i$ to minimize the embedding cost function

$$\Phi(Y) = \sum_i \left| \vec{Y}_i - \sum_j W_{ij} \vec{Y}_j \right|^2 \qquad (2)$$

- This cost function, like the previous one, is based on locally linear reconstruction errors, but here we fix the weights W_{ij} while optimizing the coordinates $\vec{Y}_i$

Algorithm: Eigenvalue Problem

- The embedding vectors $\vec{Y}_i$ are found by minimizing the cost function $\Phi(Y) = \sum_i \left| \vec{Y}_i - \sum_j W_{ij} \vec{Y}_j \right|^2$ over $\vec{Y}_i$ with fixed weights W_{ij}

- We remove this degree of freedom by requiring the coordinates to be centered on the origin $\sum_i \vec{Y}_i = \vec{0}$

- To avoid degenerate solutions, we constrain the embedding vectors to have unit covariance, with outer products that satisfy

$$\frac{1}{N} \sum_i \vec{Y}_i \vec{Y}_i^T = I_d$$

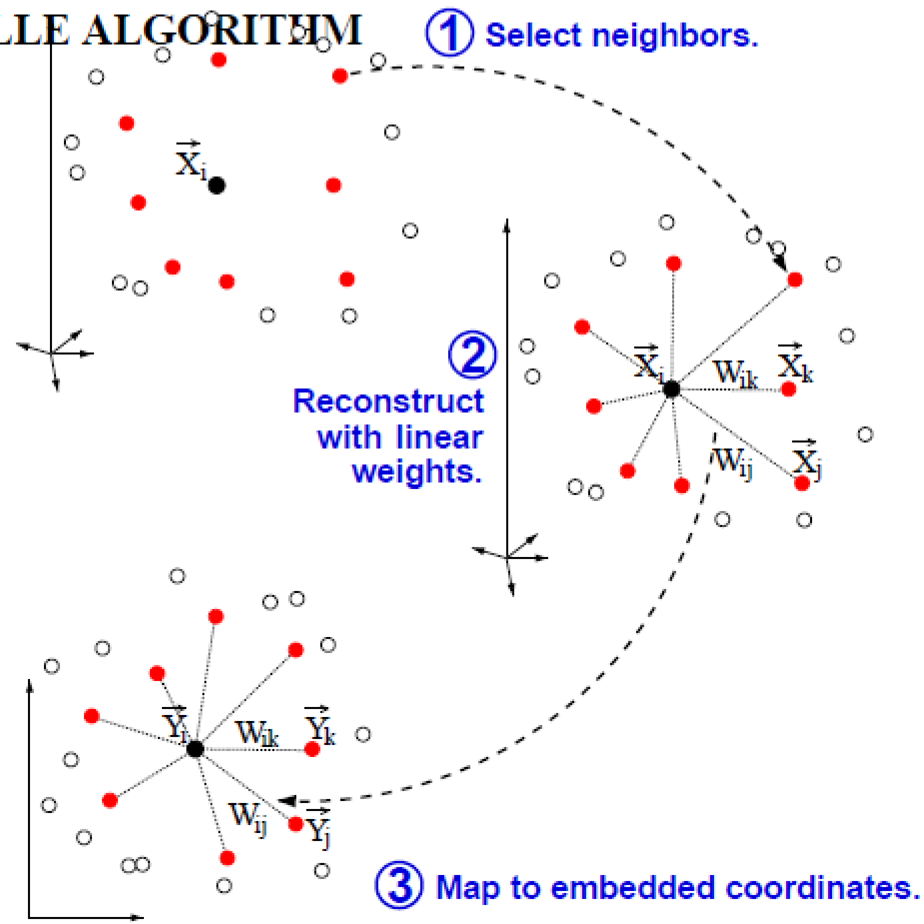
- For such implementations of LLE, the algorithm has only one free parameter: the number of neighbors, *K*.

Algorithm Summary

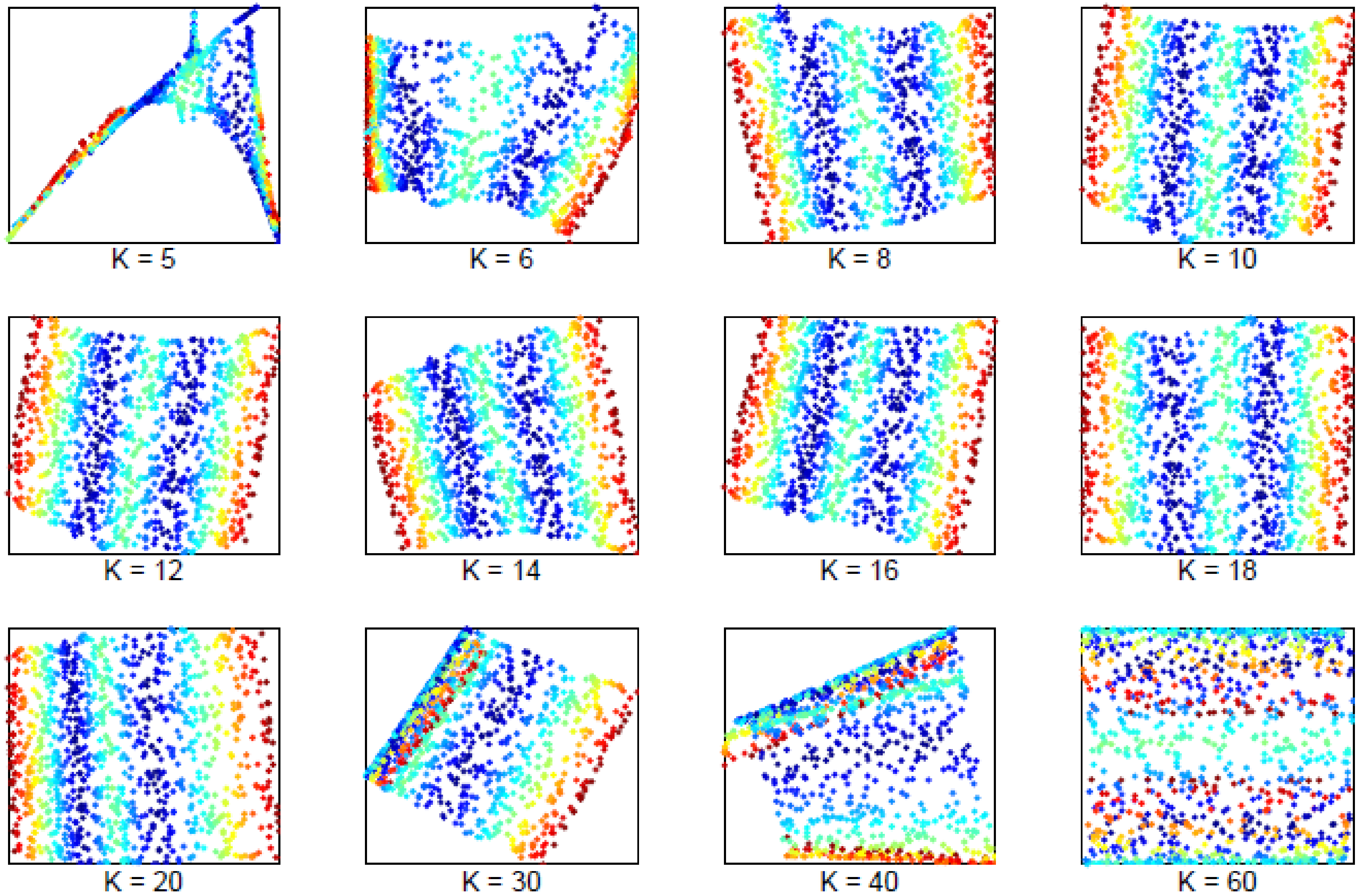
1. Compute the neighbors of each data point, $\vec{X}_i$.

2. Compute the weights W_{ij} that best reconstruct each data point $\vec{X}_i$ from its neighbors, minimizing the cost in Equation (1) by constrained linear fits.

3. Compute the vectors $\vec{Y}_i$ best reconstructed by the weights W_{ij}, minimizing the quadratic form in Equation (2) by its bottom nonzero eigenvectors.



Neighborhood Size on LLE



Toolbox in Matlab

- Website: https://lvdmaaten.github.io/drtoolbox/
- The toolbox is available to download.
- 34 dimensional reduction methods (including PCA, LLE) are there.