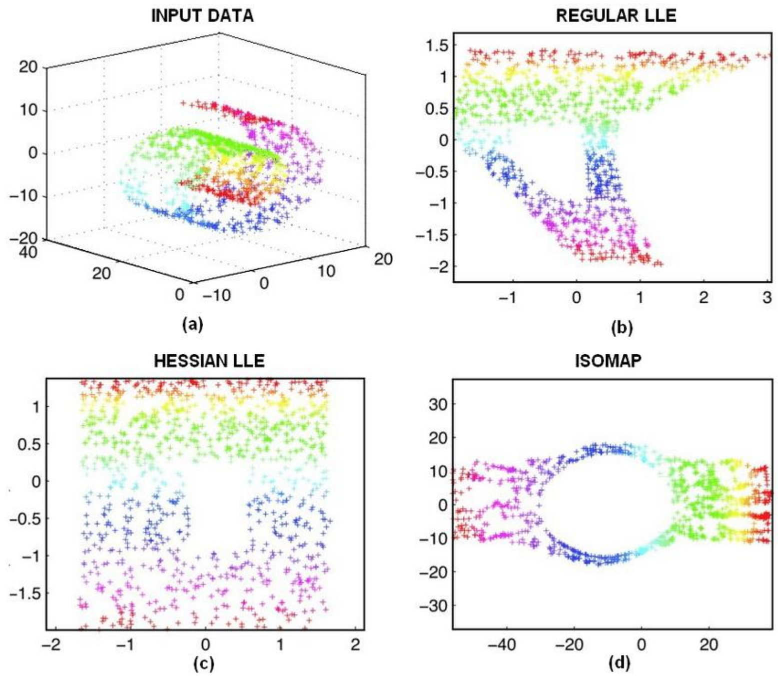
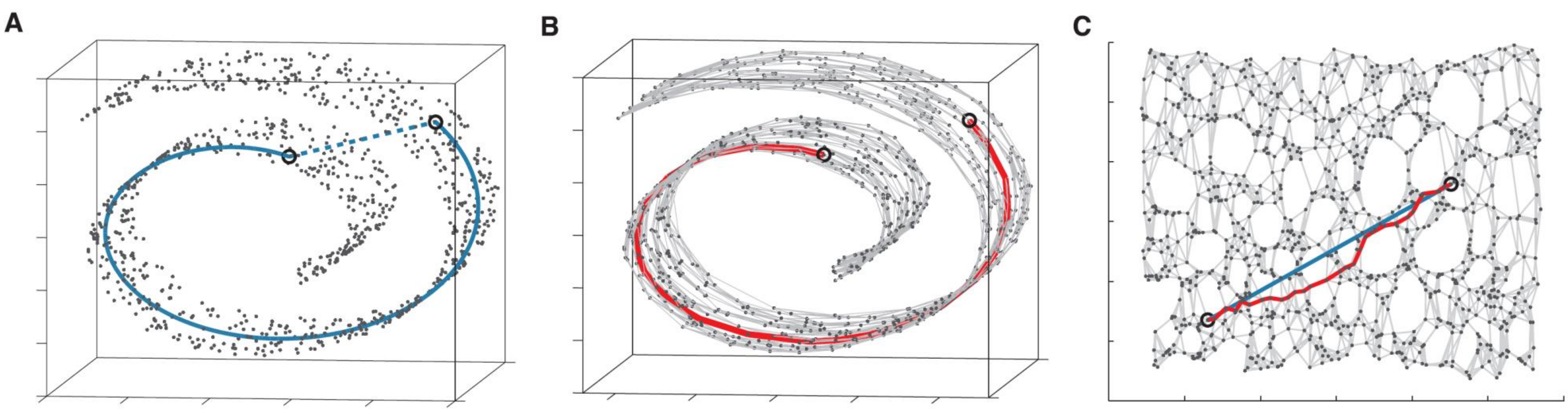


Dimensionality Reduction: ISOMAP & LLE

Pilsung Kang
School of Industrial Management Engineering
Korea University

Isometric Feature Mapping (Isomap) Tenenbaum et al. (2000)

- ISOMAP
  - ✓ Combines the major algorithmic features of PCA and MDS
    - Computational efficiency, global optimality, and asymptotic convergence guarantees
  - ✓ Builds on classical MDS but seeks to preserve the intrinsic geometry of the data, as captured in the geodesic manifold distances between all pairs of data points



Isometric Feature Mapping (Isomap)

- Isomap procedure
  - ✓ Step 1: Construct neighborhood graph
    - ε-Isomap: connect two points if they are closer than ε
    - k-Isomap: connect the point i to the point j if the i is one of the k-nearest neighbor of j
  - ✓ Step 2: Compute the shortest paths
    - Initialize $\mathrm{d_G}(i, j) = \mathrm{d_X}(i, j)$ if i and j are linked by an edge, $\mathrm{d_G}(i, j)$ = inf otherwise
    - For each value of k = 1,2, …, N in turn, replace all entries $\mathrm{d_G}(i, j)$ by

    min $\{\mathrm{d_G}(i, j), \mathrm{d_G}(i, k)+\mathrm{d_G}(k, i)$
  - ✓ Step 3: Construct d-dimensional embedding by traditional MDS

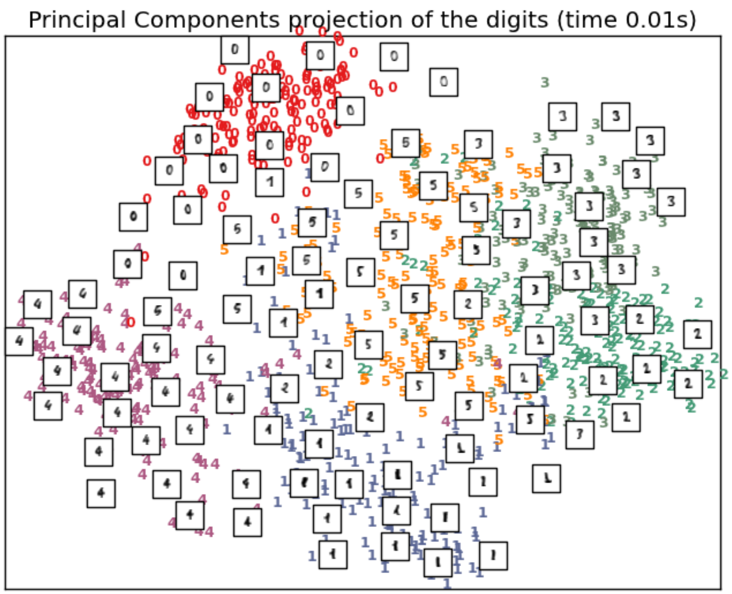
Isometric Feature Mapping (Isomap)

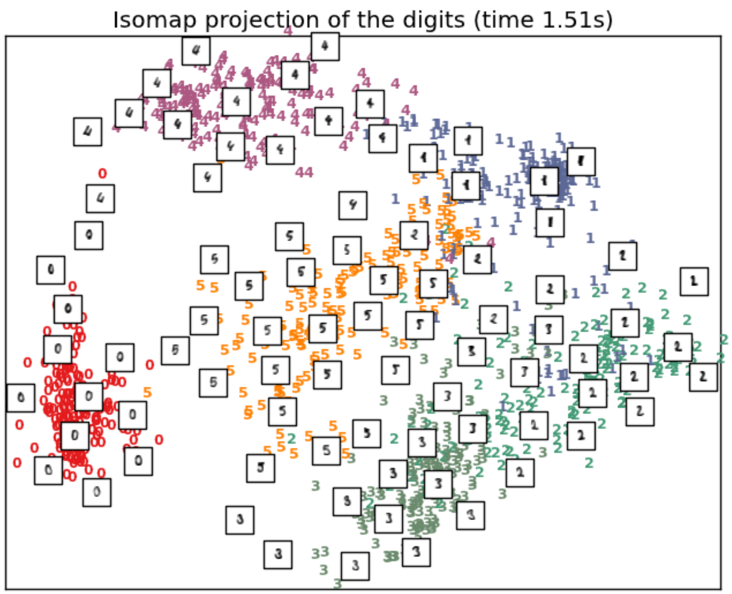
- Isomap example: Hand digit recognition

A selection from the 64-dimensional digits dataset

Isometric Feature Mapping (Isomap)

- Isomap example: Hand digit recognition

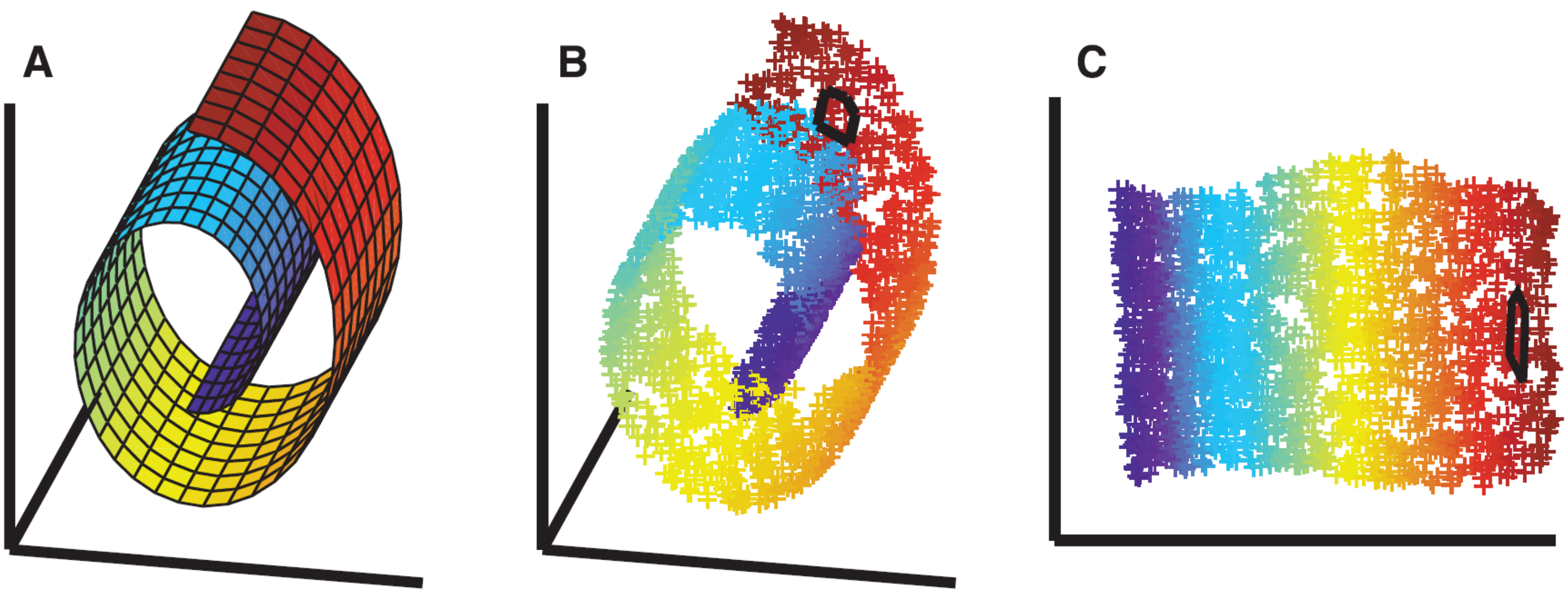




Locally Linear Embedding (LLE)

Roweis and Saul (2000)

- An eigenvector method for nonlinear dimensionality reduction
  - ✓ Simple to implement
  - ✓ Optimizations do not involve local minina
  - ✓ Capable of generating highly nonlinear embeddings
  - ✓ Map high dimensional data into a single global coordinate system of lower dimension



Locally Linear Embedding (LLE)

- LLE Procedure
  - ✓ Step 1: Compute the neighbors of each data point
  - ✓ Step 2: Compute the weight W_{ij} that best reconstruct each data point from its neighbors, minimizing the cost function by constrained linear fits

$$E(\mathbf{W}) = \sum_i \left| \mathbf{x}_i - \sum_j \mathbf{W}_{ij} \mathbf{x}_j \right|^2$$

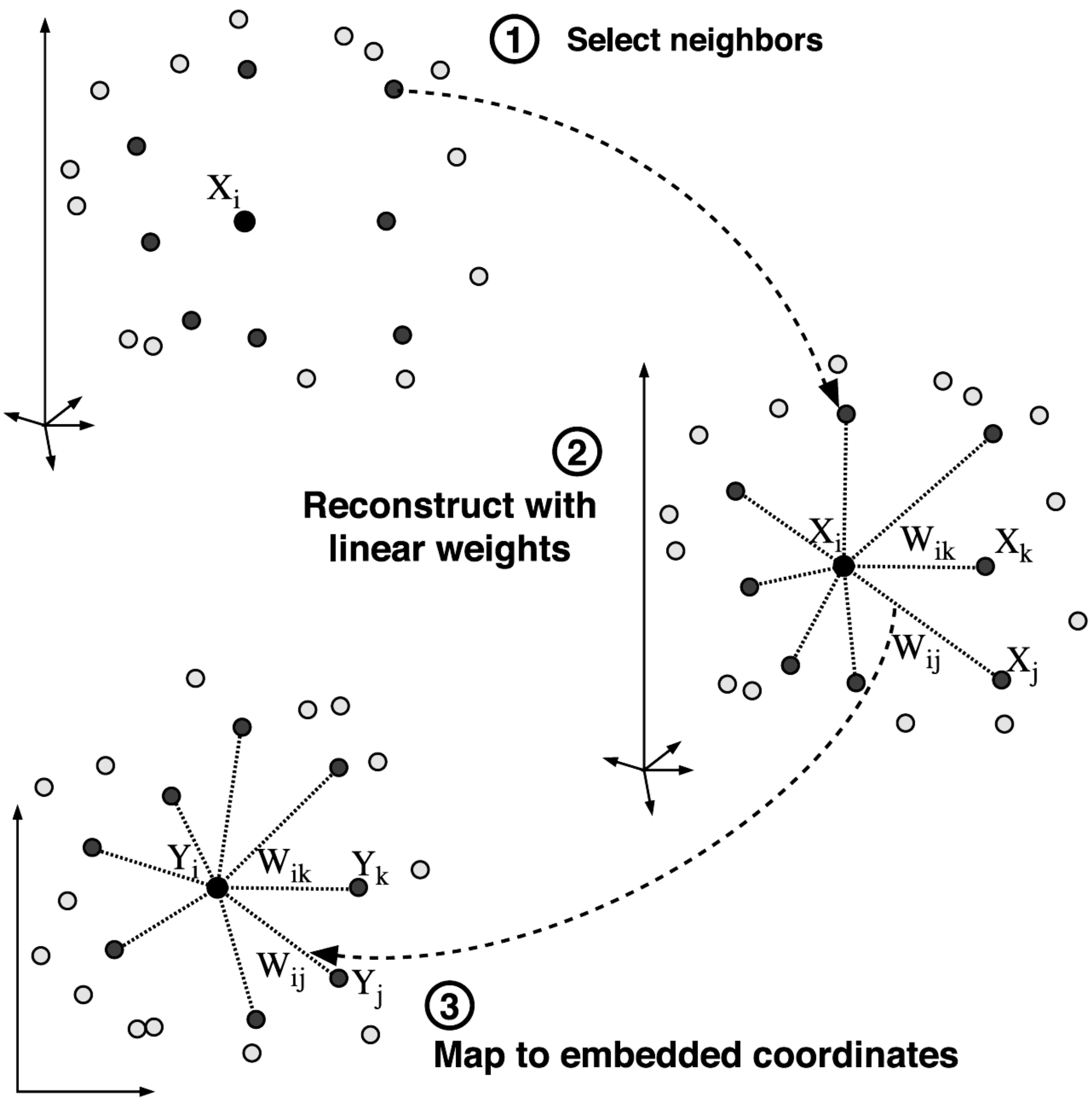
$$s.t.\ \mathbf{W}_{ij} = 0 \text{ if } \mathbf{x}_j \text{ does not belong to the neighbor of } \mathbf{x}_i$$

$$\sum_j \mathbf{W}_{ij} = 1 \text{ for all } i$$

  - ✓ Step 3: Compute the vectors best reconstructed by the weights W_{ij}, minimizing the quadratic form by its bottom nonzero eigenvectors

$$\Phi(\mathbf{W}) = \sum_i \left| \mathbf{y}_i - \sum_j \mathbf{W}_{ij} \mathbf{y}_j \right|^2$$

Locally Linear Embedding (LLE)



Weights

Invariant to rotation, re-scalings, translations of the data points due to the constraint.

Note

Although the weights W_{ij} and the vectors Y_i are computed by methods in linear algebra, the constraint that the points are only reconstructed from neighbors can result in highly nonlinear embeddings.

Locally Linear Embedding (LLE)

- LLE Procedure (cont')
  - ✓ In Step 2, the reconstruction error of $\varepsilon(\mathbf{W})$ should be zero if $\mathbf{X}_i$ is located in the convex set of its neighbors.
  - ✓ How to find Y_is in Step 3?

$$\min_{\mathbf{y}} \Phi(\mathbf{W}) = \sum_i \left| \mathbf{y}_i - \sum_j \mathbf{W}_{ij}\mathbf{y}_j \right|^2 \Rightarrow \Phi(\mathbf{W}) = \sum_{i,j} \mathbf{M}_{ij}(\mathbf{y}_i \cdot \mathbf{y}_j)$$

$$\text{where } \mathbf{M}_{ij} = \delta_{ij} - \mathbf{W}_{ij} - \mathbf{W}_{ji} + \sum_k \mathbf{W}_{ki}\mathbf{W}_{kj},\ \ \delta_{ij} = 1 \text{ if } i = j,\ \ 0 \text{ otherwise}$$

$$s.t.\ \sum_i \mathbf{y}_i = 0,\ \ \frac{1}{n}\sum_i \mathbf{y}\mathbf{y}^T = \mathbf{I}$$

Locally Linear Embedding (LLE)

- LLE Procedure (cont')
  - ✓ In Step 2, the reconstruction error of $\varepsilon(\mathbf{W})$ should be zero if $\mathbf{X}_i$ is located in the convex set of its neighbors.
  - ✓ How to find Y_is in Step 3?

$$\min_{\mathbf{y}} \Phi(\mathbf{W}) = \sum_i \left| \mathbf{y}_i - \sum_j \mathbf{W}_{ij}\mathbf{y}_j \right|^2$$

$$= \left[(\mathbf{I}-\mathbf{W})\mathbf{y} \right]^T (\mathbf{I}-\mathbf{W})\mathbf{y}$$

$$= \mathbf{y}^T(\mathbf{I}-\mathbf{W})^T(\mathbf{I}-\mathbf{W})\mathbf{y}$$

$$= \mathbf{y}^T\mathbf{M}\mathbf{y}$$

Locally Linear Embedding (LLE)

- LLE Procedure (cont')
  - ✓ The optimal embedding is found by computing the bottom d+1 eigenvectors of the matrix M (Rayleitz-Ritz theorem)
  - ✓ The bottom eigenvector is the unit vector with all equal components
  - ✓ Discarding this eigenvector enforces the constraint that the embeddings have zero mean

---

The **Rayleigh–Ritz method** allows for the computation of Ritz pairs $(\tilde{\lambda}_i, \tilde{\mathbf{x}}_i)$ which approximate the solutions to the eigenvalue problem [1]

$$A\mathbf{x} = \lambda\mathbf{x}$$

Where $A \in \mathbb{C}^{N \times N}$.

The procedure is as follows:[2]

1. Compute an orthonormal basis $V \in \mathbb{C}^{N \times m}$ approximating the eigenspace corresponding to m eigenvectors
2. Compute $R \leftarrow V^* A V$
3. Compute the eigenvalues of R solving $R\mathbf{v}_i = \tilde{\lambda}_i \mathbf{v}_i$
4. Form the ritz pairs $(\tilde{\lambda}_i, \tilde{\mathbf{x}}_i) = (\tilde{\lambda}_i, V\mathbf{v}_i)$

One can always compute the accuracy of such an approximation via $\|A\tilde{\mathbf{x}}_i - \tilde{\lambda}_i \tilde{\mathbf{x}}_i\|$

If a Krylov subspace is used and A is a general matrix, then this is the Arnoldi Algorithm.

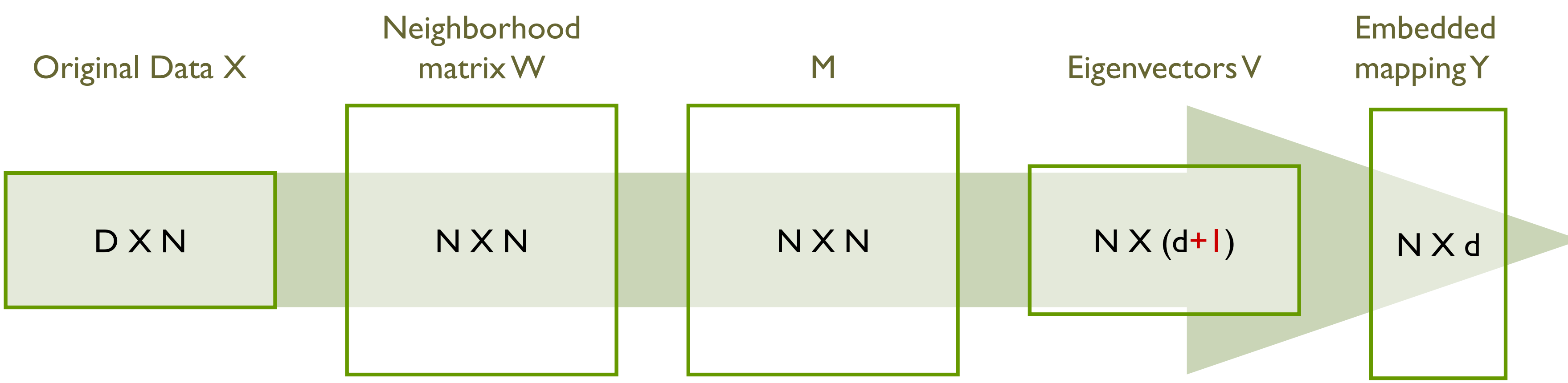
Locally Linear Embedding (LLE)

- LLE Procedure (cont')

Compute embedding coordinates y using weights W

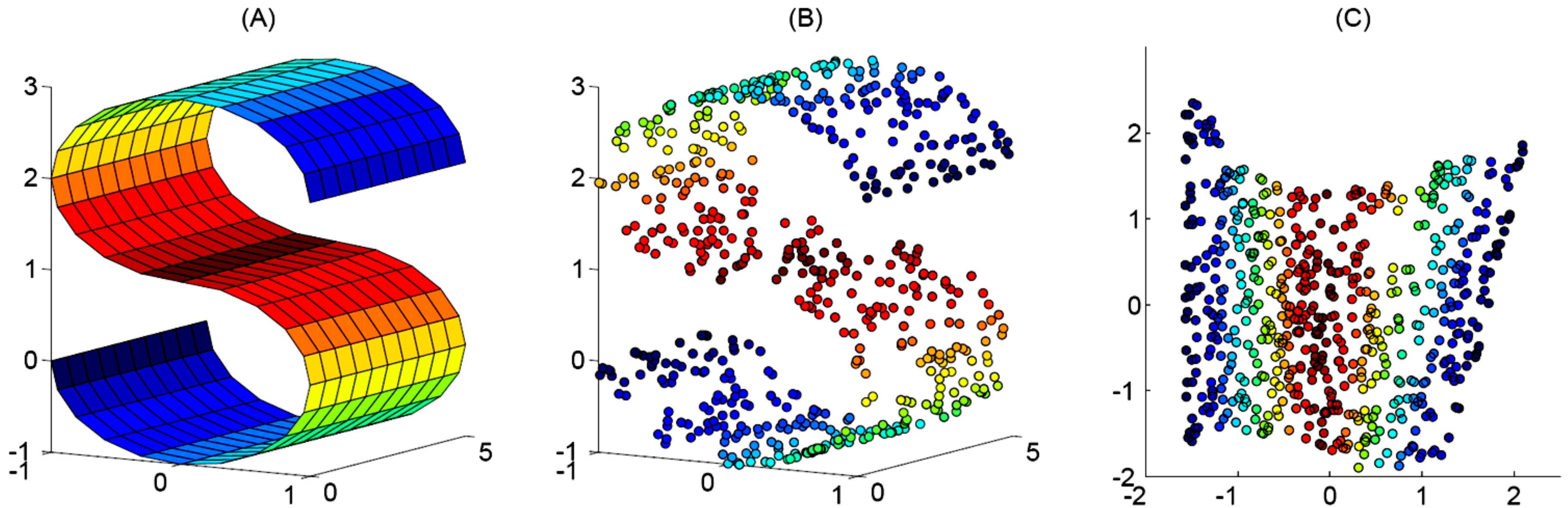
- create sparse matrix M = (I-W)'*(I-W)
- find bottom d+1 eigenvectors of M (corresponding to the d+1 smallest eigenvalues)
- set the qth ROW of Y to be the q+1 smallest eigenvector (discard the bottom eigenvector [1,1,1,1...] with eigenvalue zero) …(from LLE Homepage)

Matrix transition during LLE process



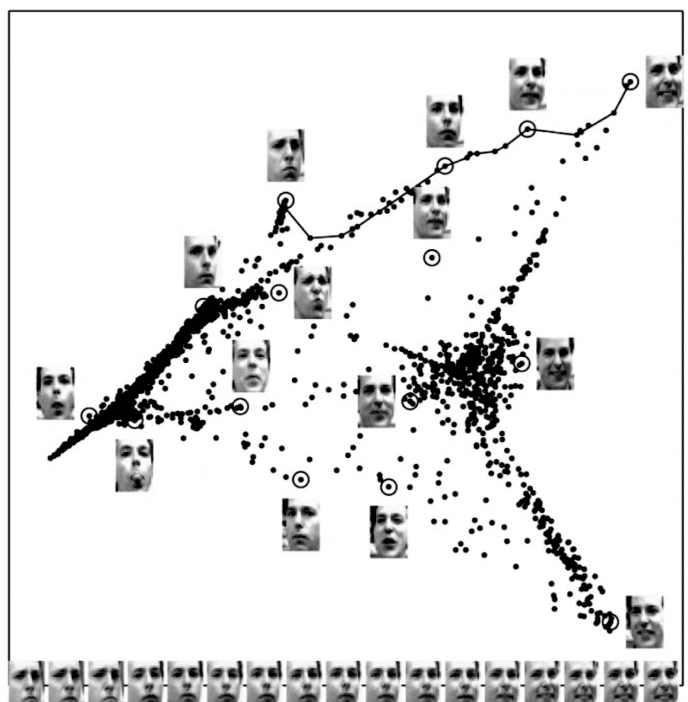
Locally Linear Embedding (LLE)

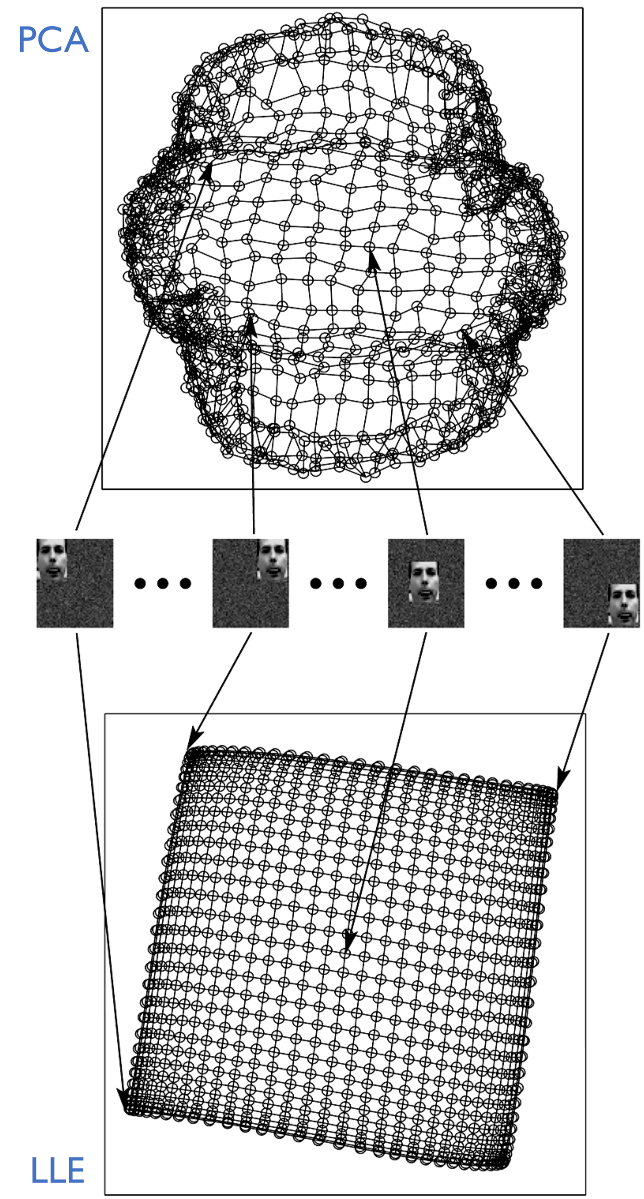
- LLE Example 1
  - ✓ N=600, k=12



Locally Linear Embedding (LLE)

- LLE Example 2
  - ✓ N = 961 grayscale images
  - ✓ Each image containing 28 X 20 face superimposed on 59 X 51 background of noise
  - ✓ D = 3009 K = 4

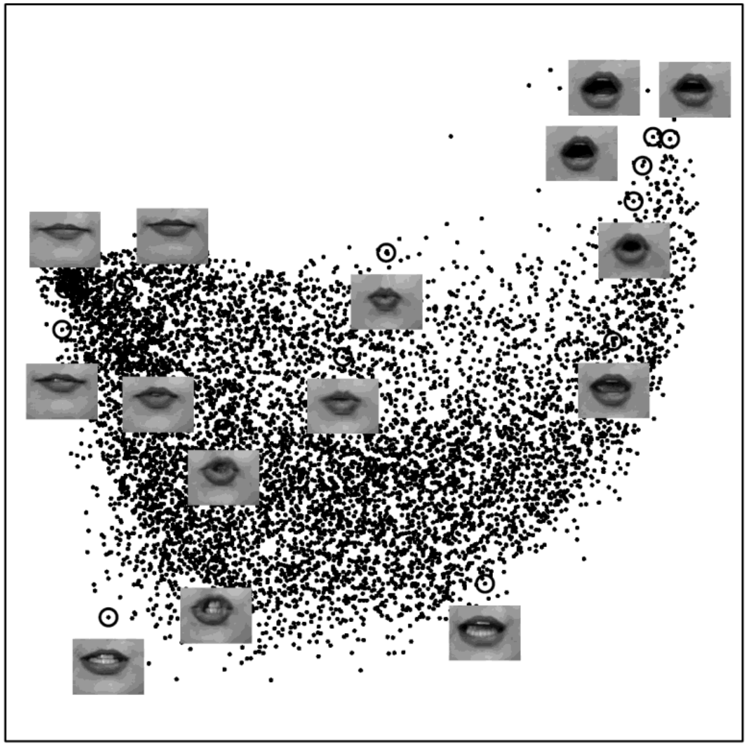




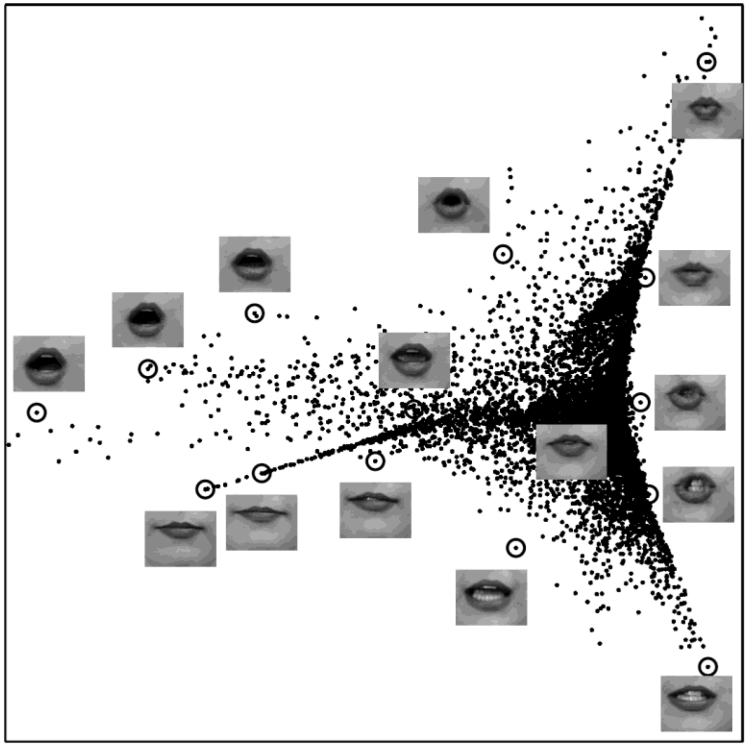
Locally Linear Embedding (LLE)

- LLE Example 3
  - ✓ Images of lips used in the animation talking head
  - ✓ N = 8588 RGB images of lips at 108 X 84 resolution
  - ✓ D = 27,216, K = 16

PCA

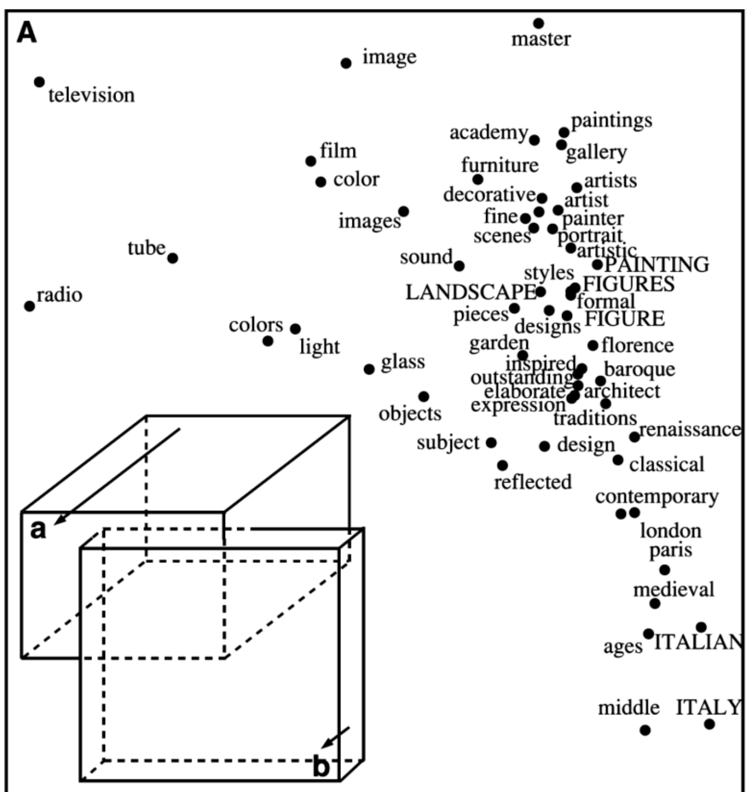


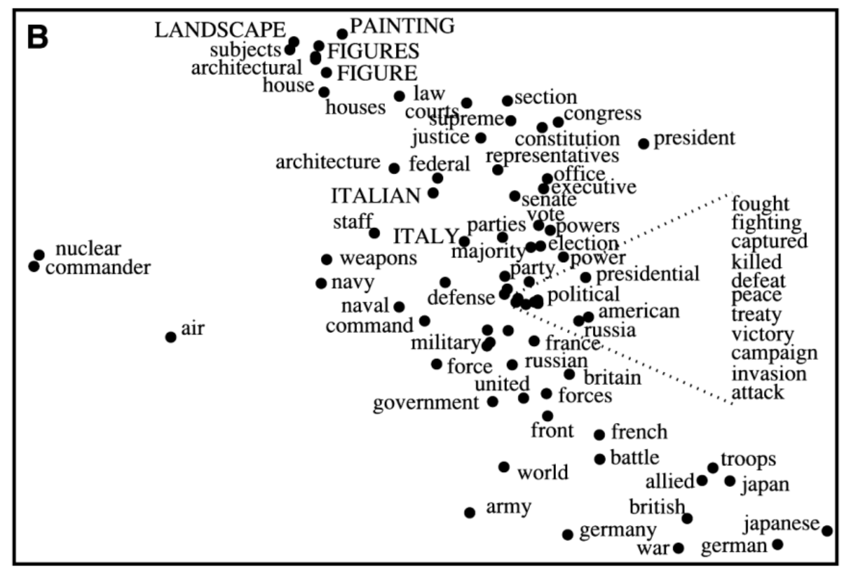
LLE



Locally Linear Embedding (LLE)

- LLE Example 3
  - ✓ Arranging words in a continuous semantic space
  - ✓ Each word was initially represented by a high-dimensional vector that counted the number of times it appeared in different encyclopedia articles






Q
&
A

References

Research Papers

- Roweis, S. and Saul, L. (2000). Nonlinear dimensionality reduction by locally linear embedding. SCIENCE 290:2323-2326.
- Tenenbaum, J.B., de Silva, V., and Langford, J.C. (2000). A global geometric framework for nonlinear dimensionality reduction. SCIENCE 290:2319-2323

Other materials

- Figure in the title page: https://www.researchgate.net/figure/Comparison-of-the-results-of-LLE-Hessian-LLE-and-ISOMAP-for-the-Swiss-roll-input-data_fig10_260755521