Nonlinear Dimensionality Reduction

Fall 2005

Ahmed Elgammal
Dept of Computer Science
Rutgers University

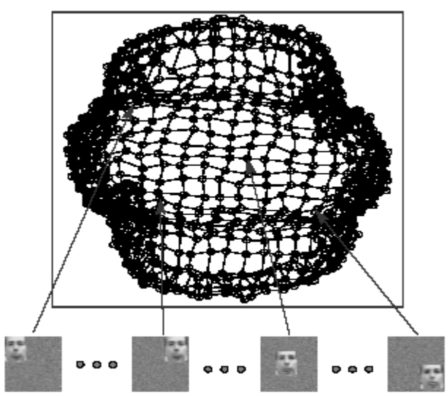
Outline

- Intuition of Nonlinear Dimensionality Reduction(NLDR)
- ISOMAP
- LLE

PCA & MDS

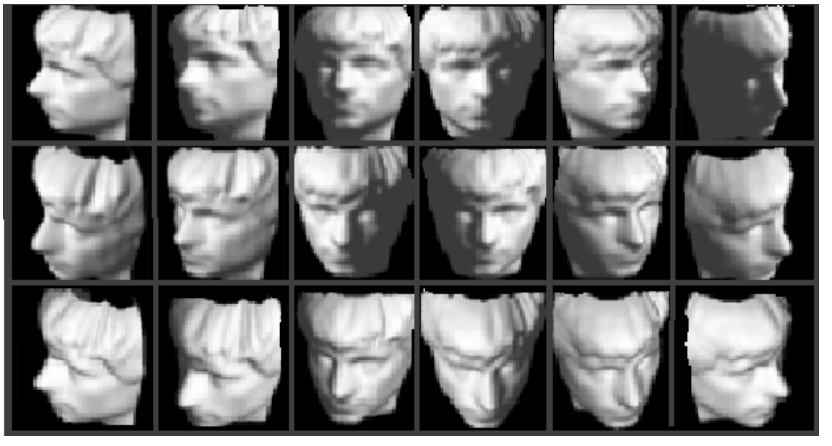
- Principle Component Analysis:PCA
  - Linear projections of greatest variance from the top eigenvectors of the data covariance matrix
- Multidimensional Scaling:MDS
  - Low dimensional embedding that best preserves pairwise distances between data points
- Modeling of linear variabilities in high dimensional data
- No local minima
- Find linear subspace and **cannot deal properly with data lying on nonlinear manifolds**

Problem in PCA and MDS

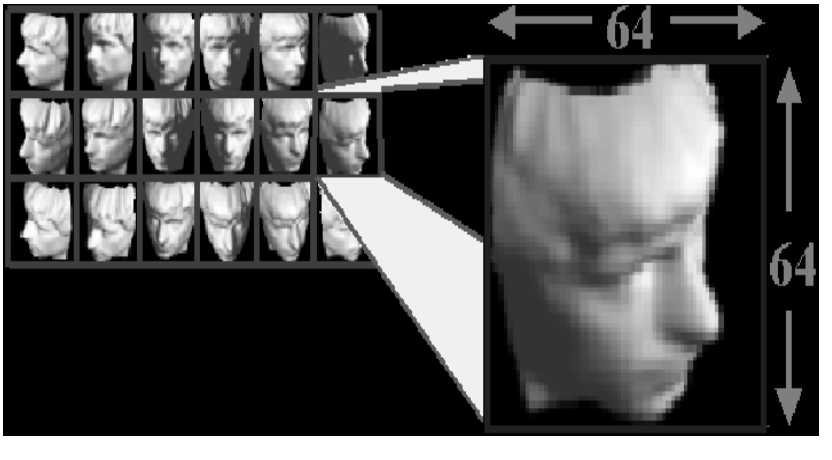


- Create distortion in local and global geometry
  - Map faraway data point to nearby point

Intuition: how does your brain store these pictures?



Brain Representation

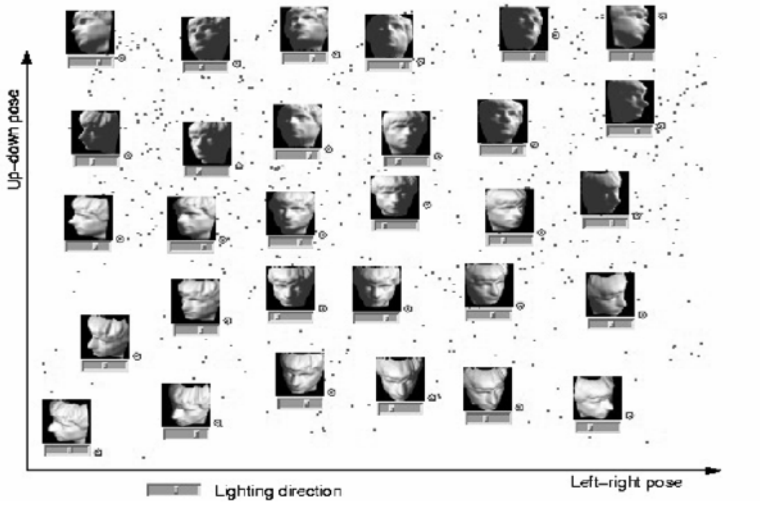


Brain Representation

- Every pixel?
- Or perceptually meaningful structure?
  - Up-down pose
  - Left-right pose
  - Lighting direction

So, your brain successfully reduced the high-dimensional inputs to an intrinsically 3-dimensional manifold!

The computed manifold



Model for NLDR

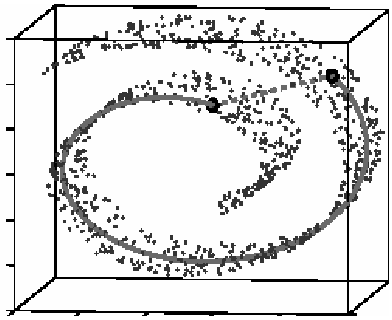
- R^N is the *observation space*
- Y is d-dimensional *feature space*
- A smooth embedding f: $R^N \rightarrow Y$ (N>>d)

Isomap: Algorithm

Step	Name	Description
1 $O(DN^2)$	Construct neighborhood graph, **G**	Compute matrix $\mathbf{D}_G=\{d_X(i,j)\}$ $d_x(i,j)$ = Euclidean distance between *neighbors*
2 $\mathbf{O(DN^2)}$	Compute shortest paths between ***all*** pairs	Compute matrix $\mathbf{D}_G=\{d_G(i,j)\}$ $d_G(i,j)$ = approx geodesic dist.
3 $O(dN^2)$	Construct d-dimensional coordinate vectors, $\mathbf{y}_i$	Apply MDS to $\mathbf{D}_G$ instead of $\mathbf{D}_X$

Sample points

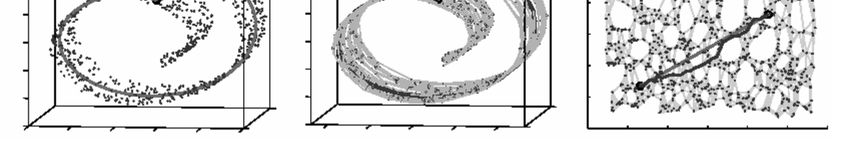
- Altogether there are 20,000 points in the "Swiss roll" data set. We sample 1000 out of 20,000.



Construct neighborhood graph G

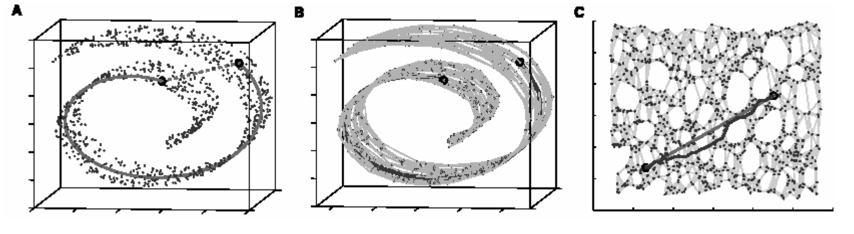
K- nearest neighborhood (K=7)

D_G is 1000 by 1000 (Euclidean) distance matrix of two neighbors (figure A)



Compute all-points shortest path in G

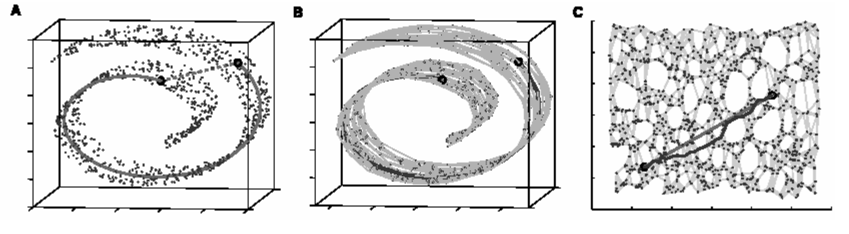
Now D_G is 1000 by 1000 geodesic distance matrix of two arbitrary points along the manifold(figure B)



Use MDS to embed graph in R^d

Find a d-dimensional Euclidean space Y(Figure c) to minimize the cost function:

$$E = \|\tau(D_G) - \tau(D_Y)\|_{L^2}$$



Linear Approach-classical MDS

$E = \|\tau(D_G) - \tau(D_Y)\|_{L^2}$

$\tau(D_G) = -\frac{1}{2} H D_G^2 H$

$\tau(D_Y) = -\frac{1}{2} H D_Y^2 H$

$D_G^2 = \left[D_{ij}^2\right]$

$H = I - \frac{1}{N}\underline{1*1}$

$\min \| \tau(D_G) - \tau(D_Y) \| = \min \| -0.5 H (D_G^2 - D_Y^2) H \|^2$

$= \min \| X^T X - Y^T Y \|^2 == \min trace(X^T X - Y^T Y)^2$

Theorem: For any squared distance matrix D_G^2 ,there exists of points x_i and, x_j, such that $d_{ij}^2 = (x_i - x_j)'(x_i - x_j)$

So $D_G^2 = X^T X$

Solution

$$Y = \Lambda_{Y(d)}^{1/2} V \Rightarrow \mathbf{Y} = \begin{bmatrix} | & | & & | \\ \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_N \\ | & | & & | \end{bmatrix}_{dxN} = \begin{bmatrix} - & \sqrt{\lambda_1}\mathbf{v}_1 & - \\ - & \sqrt{\lambda_2}\mathbf{v}_2 & - \\ & \vdots & \\ - & \sqrt{\lambda_d}\mathbf{v}_d & - \end{bmatrix}_{dxN}$$

Y lies in R^d and consists of N points correspondent to the N original points in input space.

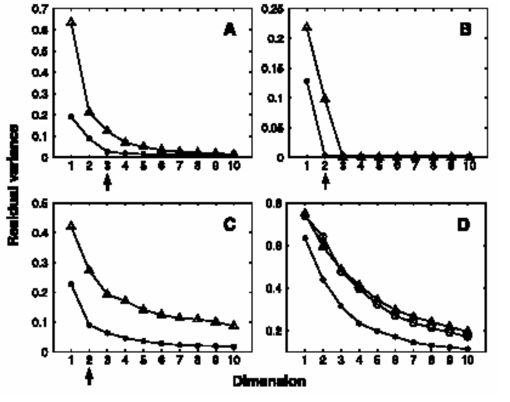
Isomap: Advantages

- Nonlinear
- Globally optimal
  - Still produces globally optimal low-dimensional Euclidean representation even though input space is highly folded, twisted, or curved.
- Guarantee asymptotically to recover the true dimensionality.

Isomap: Disadvantages

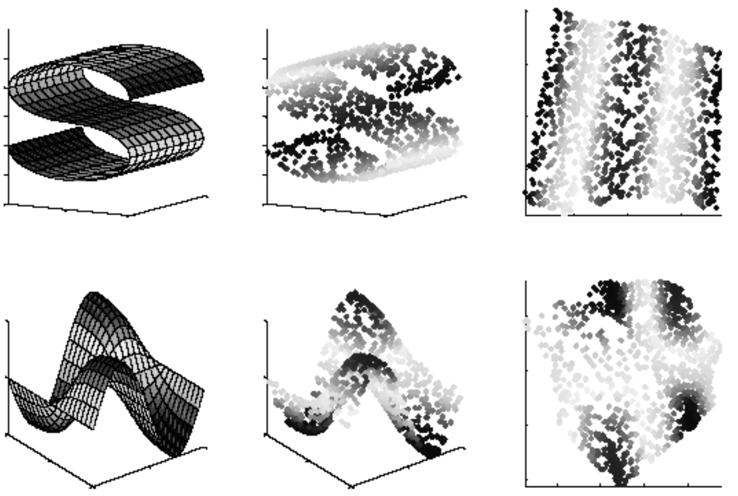
- Guaranteed asymptotically to recover geometric structure of nonlinear manifolds
  – As N increases, pairwise distances provide better approximations to geodesics by "hugging surface" more closely
  – Graph discreteness overestimates $d_M(i,j)$
- K must be high to avoid "linear shortcuts" near regions of high surface curvature
- Mapping novel test images to manifold space

PCA, MD vs ISOMAP



LLE(Locally Linear Embedding)

- Property
  - Preserving the local configurations of nearest neighbours
- LLE
  - Local: only neighbours contribute to each reconstruction
  - Linear: reconstructions are confined to linear subspace
- Assumption
  - Well-sampled data->locally linear patch of the manifold
  - d-dimensional manifold->2d neighbors

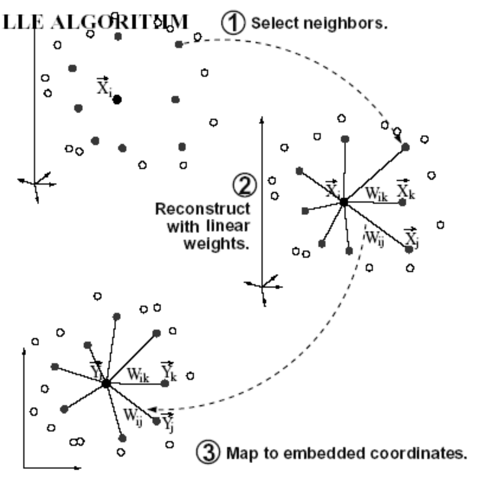


LLE Algorithm

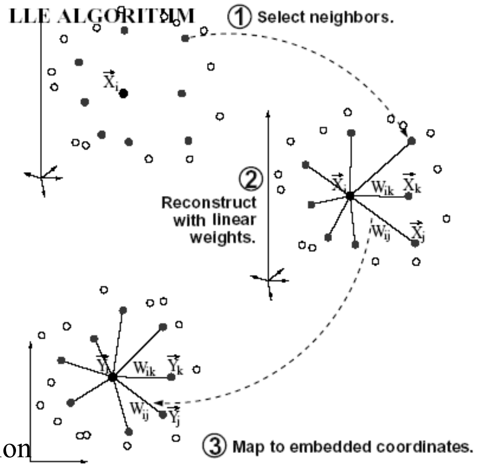
- Step1

  Neighborhood search

  - Compute the neighbours of each data point
  - K nearest neighbours per data point



LLE Algorithm



- Step2

  Constrained Least Square Fits

  – Reconstruction Error

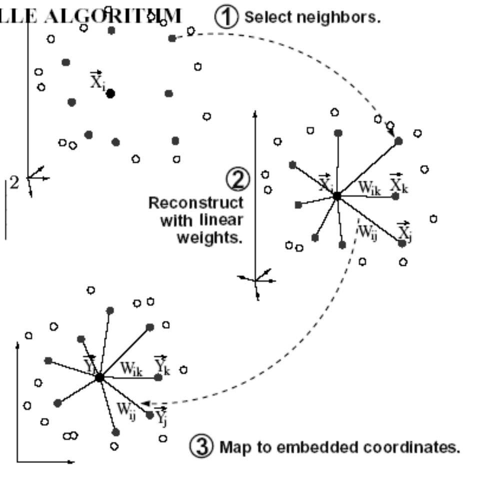
$$\mathcal{E}(W) = \sum_i \left|\vec{X}_i - \sum_j W_{ij}\vec{X}_j\right|^2$$

if not neighbour

$W_{ij} = 0$

$\sum_j W_{ij} = 1$ Invariant to rotations, rescaling, and translation of that point and its neighbour

LLE Algorithm



- Step3

  Eigenvalue Problem

  – Reconstruction

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

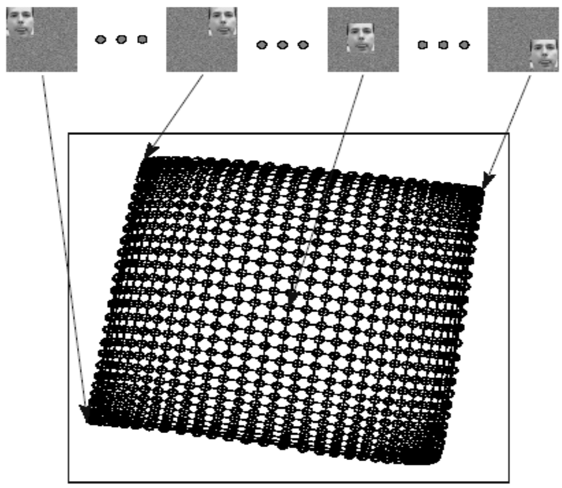
– $\sum_i \vec{Y}_i = \vec{0}.$

centred at the origin

– $\frac{1}{N}\sum_i \vec{Y}_i\vec{Y}_i^{\top} = I$

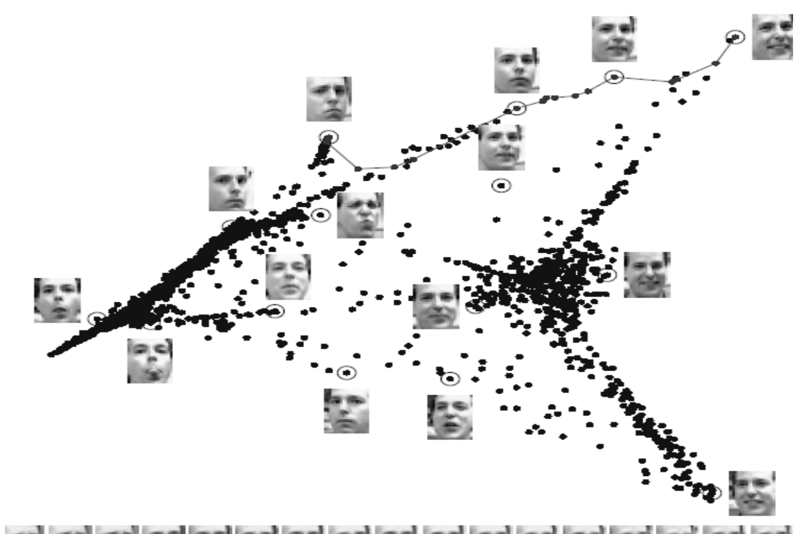
Avoid degenerate solution

Embedding by LLE



Corners faces to the corners of its two dimensional embedding

Examples



LLE advantages

- Ability to discover nonlinear manifold of arbitrary dimension
- Non-iterative
- Global optimality
- Few parameters: K, d
- O(DN2) and space efficient due to sparse matrix

LLE disadvantages

- Requires smooth, non-closed, densely sampled manifold
- Quality of manifold characterization dependent on neighborhood choice
- Sensitive to outliers

Comparisons: PCA vs LLE vs Isomap

- ***PCA: find embedding coordinate vectors that minimize distance to all data***
- ***LLE: find embedding coordinate vectors that best fit local neighborhood relationships***
- ***ISOMAP: find embedding coordinate vectors that preserve geodesic shortest distances***