What Kind of Knowledge Does Locally Linear Embedding Extract?

Michael W. Trosset

Department of Statistics

Indiana University

This research was supported by a grant from the Office of Naval Research.

Manifold Learning

Motivated by the conceit that $y_1, \ldots, y_n \in \Re^q$ lie on a low-dimensional manifold, techniques for *manifold learning* attempt to localize structure in the ambient space, then exploit properties of manifolds to construct a low-dimensional representation.

Seminal papers include

- Tenenbaum, de Silva & Langford (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323.

  Isomap "seeks to preserve the intrinsic geometry of the data, as captured in the geodesic manifold distances between all pairs of data points."

- Roweis & Saul (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326.

  The same weights that locally reconstruct y_i from its neighbors in the feature space should also reconstruct x_i from its neighbors in the representation space.

Both techniques use pairwise proximities to embed the data in $\Re^d$, $d \ll q$.

Embedding Proximity Data

Consider a collection of n objects, together with pairwise proximity information that we organize in the form of a symmetric matrix.

A *dissimilarity matrix* $\Delta = [\delta_{ij}]$ is a matrix that is symmetric, nonnegative ($\delta_{ij} \geq 0$), and hollow ($\delta_{ii} = 0$). The interpretation of dissimilarity demands the following monotonicity property: pair (i, j) is more dissimilar than pair (r, s) iff $\delta_{ij} > \delta_{rs}$.

The goal of *multidimensional scaling* (MDS) is to model dissimilarity with distance, e.g., to construct a configuration of points in a specified Euclidean space in such a way that points correspond to objects and pairwise interpoint distances approximate pairwise object dissimilarity.

A *similarity matrix* $\Gamma = [\gamma_{ij}]$ is a matrix that is symmetric, nonnegative, and satisfies $\gamma_{ii} \geq \gamma_{ij}$. The interpretation of similarity demands the following monotonicity property: pair (i, j) is more similar than pair (r, s) iff $\gamma_{ij} > \gamma_{rs}$.

To construct a meaningful configuration of points from similarity data, one typically transforms the observed similarity matrix to a dissimilarity matrix, then performs MDS. The choice of tranformation from similarity to dissimilarity is crucial.

Euclidean Distance Geometry

A dissimilarity matrix $\Delta = [\delta_{ij}]$ is a *Type 2 Euclidean distance matrix* (EDM-2) iff there exist $x_1, \ldots, x_n \in \Re^p$ such that $\delta_{ij} = \|x_i - x_j\|^2$. The smallest such p is the *embedding dimension* of the EDM-2.

Let I denote the $n \times n$ identity matrix, let $e = (1, \ldots, 1)^t \in \Re^n$, and let $P = I - ee^t/n$. Notice that P is symmetric and idempotent; Pv is the projection of $v \in \Re^n$ into $e^\perp$, $P\Delta P$ is the "double centering" of Δ, and $P\Delta Pe = 0$.

Theorem: A dissimilarity matrix Δ is EDM-2 with embedding dimension p iff the symmetric matrix

$$\tau(\Delta) = -\frac{1}{2} P\Delta P$$

is positive semidefinite (psd) and has rank p. Furthermore, if $\Delta = [\delta_{ij}]$ is EDM-2 and

$$\tau(\Delta) = \begin{bmatrix} x_1^t \\ \vdots \\ x_n^t \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix},$$

then $\delta_{ij} = \|x_i - x_j\|^2$.

Let

$$B = \tau(\Delta) = \left[x_i^t x_j\right] = XX^t.$$

Because $2Be = -P\Delta Pe = 0$, $(X^te)^t(X^te) = e^tBe = 0$ and therefore $Xe = 0$. Thus, τ converts squared distances to centered inner products.

Conversely, if $x_1, \ldots, x_n \in R^p$ have inner product matrix B, then they have EDM-2

$$\kappa(B) = \text{diag}(B)ee^t - 2B + ee^t\text{diag}(B).$$

Thus, κ converts inner products to squared distances.

If we restrict attention to symmetric psd B that satisfy $Be = 0$, then the linear transformations τ and κ are mutually inverse.

Classical MDS

Suppose that we want to embed fallible dissimilarities $\Delta = [\delta_{ij}]$ in $\Re^d$.

If the dissimilarity matrix $\Delta_2 = [\delta_{ij}^2]$ is *not* EDM-2 with embedding dimension $\leq d$, then $B = \tau(\Delta_2)$ is not psd with rank $\leq d$. Hence, we cannot factor B to obtain a d-dimensional configuration of points. Classical MDS (CMDS) circumvents this difficulty by replacing B with $\bar{B}$, the nearest symmetric psd matrix with rank $\leq d$.

Let $\lambda_1 \geq \cdots \geq \lambda_n$ denote the eigenvalues of B and let $v_1, \ldots, v_n$ denote the corresponding eigenvectors. Let $\sigma_i^2 = \max(\lambda_i, 0)$ for $i = 1, \ldots, d$. Then

$$\bar{B} = \sum_{i=1}^{d} \sigma_i^2 v_i v_i^t = \begin{bmatrix} \sigma_1 v_1 & \cdots & \sigma_d v_d \end{bmatrix} \begin{bmatrix} \sigma_1 v_1^t \\ \vdots \\ \sigma_d v_d^t \end{bmatrix}$$

produces a d-dimensional configuration of points whose principal components are eigenvectors of $\tau(\Delta_2)$.

Locally Linear Embedding

Given user-specified neighbors of each $y_i \in \Re^q$ (typically, the K nearest neighbors), LLE attempts to characterize local geometry by linear coefficients that reconstruct each y_i from its neighbors. Let $N(i)$ denote the indices of the neighbors of y_i.

1. Minimize
$$\mathsf{E}(C) = \sum_{i=1}^{n} \left\| y_i - \sum_{j \in N(i)} c_{ij} y_j \right\|^2,$$
subject to $c_{ij} = 0$ if $j \notin N(i)$ and $Ce = e$.

2. Minimize
$$\Phi(X) = \sum_{i=1}^{n} \left\| x_i - \sum_{j \in N(i)} c_{ij} x_j \right\|^2,$$
subject to $X^t e = 0$ and $X^t X = nI$.

   The d columns of the minimizer are the eigenvectors that correspond to the d smallest positive eigenvalues of $M = (I - C)^t (I - C)$.

Example

Applying LLE with $K = 4$ to

y_1	y_2	y_3	y_4	y_5	y_6	y_7
1	2	3	4	5	6	7
ϵ	$-\epsilon$	ϵ	$-\epsilon$	ϵ	$-\epsilon$	ϵ
$-\epsilon$	0	0	2ϵ	0	0	$-\epsilon$

results in reconstruction coefficients

$$6C = \begin{bmatrix} 0 & 3 & 9 & -3 & -3 & 0 & 0 \\ 12 & 0 & -18 & 6 & 6 & 0 & 0 \\ 4 & -2 & 0 & 2 & 2 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 2 & 2 & 0 & -2 & 4 \\ 0 & 0 & 6 & 6 & -18 & 0 & 12 \\ 0 & 0 & -3 & -3 & 9 & 3 & 0 \end{bmatrix}.$$

The following have unique solutions. Coefficients for y_7, y_6, y_5 follow by symmetry.

y_1	$$\begin{array}{rcrcrcrcr} 2c_{12} & + & 3c_{13} & + & 4c_{14} & + & 5c_{15} & = & 1 \\ -\epsilon c_{12} & + & \epsilon c_{13} & - & \epsilon c_{14} & + & \epsilon c_{15} & = & \epsilon \\ & & & & 2\epsilon c_{14} & & & = & -\epsilon \\ c_{12} & + & c_{13} & + & c_{14} & + & c_{15} & = & 1 \end{array}$$
y_2	$$\begin{array}{rcrcrcrcr} c_{21} & + & 3c_{23} & + & 4c_{24} & + & 5c_{25} & = & 2 \\ \epsilon c_{21} & + & \epsilon c_{23} & - & \epsilon c_{24} & + & \epsilon c_{25} & = & -\epsilon \\ -\epsilon c_{21} & & & & 2\epsilon c_{24} & & & = & 0 \\ c_{21} & + & c_{23} & + & c_{24} & + & c_{25} & = & 1 \end{array}$$
y_3	$$\begin{array}{rcrcrcrcl} c_{31} & + & 2c_{32} & + & 4c_{34} & + & 5c_{35} & = & 3 \\ \epsilon c_{31} & - & \epsilon c_{32} & - & \epsilon c_{34} & + & \epsilon c_{35} & = & \epsilon \\ -\epsilon c_{31} & & & + & 2\epsilon c_{34} & & & = & \epsilon \\ c_{31} & + & c_{32} & + & c_{34} & + & c_{35} & = & 1 \end{array}$$

For y_4, the least squares problem

$$\begin{array}{ll} \text{minimize} & (2c_{42}+3c_{43}+5c_{45}+6c_{46}-4)^2 + \\ & (-\epsilon c_{42}+\epsilon c_{43}+\epsilon c_{45}-\epsilon c_{46}+\epsilon)^2+(2\epsilon)^2 \\ \text{subject to} & c_{42}+c_{43}+c_{45}+c_{46}=1 \end{array}$$

has infinitely many solutions, but only one is symmetric ($c_{43}=c_{45}$, $c_{42}=c_{46}$).

Gaussian elimination reveals that

$$6(I-C)=\begin{bmatrix} 6 & -3 & -9 & 3 & 3 & 0 & 0 \\ -12 & 6 & 18 & -6 & -6 & 0 & 0 \\ -4 & 2 & 6 & -2 & -2 & 0 & 0 \\ 0 & -3 & 0 & 6 & 0 & -3 & 0 \\ 0 & 0 & -2 & -2 & 6 & 2 & -4 \\ 0 & 0 & -6 & -6 & 18 & 6 & -12 \\ 0 & 0 & 3 & 3 & -9 & -3 & 6 \end{bmatrix}$$

has rank 3; hence, $M=(I-C)^t(I-C)$ has 4 zero and 3 strictly positive eigenvalues.

The positive eigenvalues of $36M$ are

$$\lambda_1 = 1029 > \lambda_2 = \frac{545 + \sqrt{209217}}{2} \doteq 501 > \lambda_3 = \frac{545 - \sqrt{209217}}{2} \doteq 44.$$

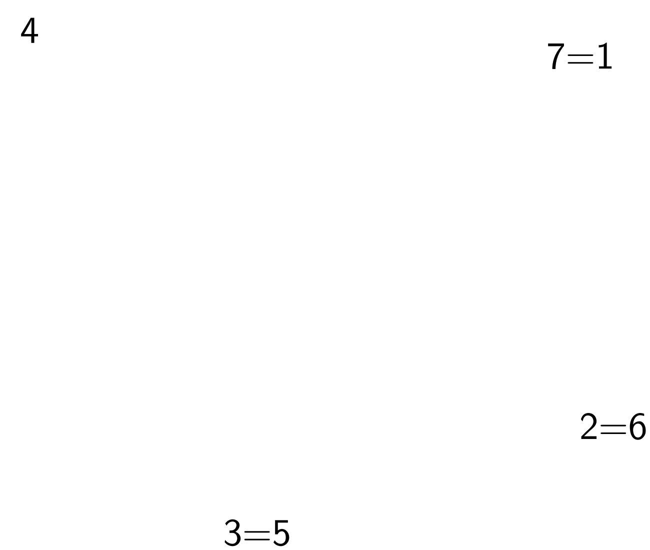
The corresponding eigenvectors are

$$\begin{aligned} v_1 &\propto (-2, 1, 4, 0, -4, -1, 2)^t\,, \\ v_2 &\propto (1, \alpha, -1, -2\alpha, -1, \alpha, 1)^t\,, \\ v_3 &\propto (1, \beta, -1, -2\beta, -1, \beta, 1)^t\,, \end{aligned}$$

where

$$\alpha = \frac{392 - \lambda_2}{194} \doteq -0.6 \quad \text{and} \quad \beta = \frac{392 - \lambda_3}{194} \doteq 1.8.$$

The 2-dimensional representation constructed by LLE is:



Laplacian Eigenmaps

Let G be a weighted graph with n vertices and edge weights stored in the symmetric $n \times n$ matrix W. Let T denote the diagonal $n \times n$ matrix whose diagonal is We, the row/column sums of W. The *Laplacian* of G is the symmetric matrix $B = T - W$. Notice that $Be = 0$ and that B is diagonally dominant, hence psd. A Laplacian eigenmap constructs a configuration from the eigenvectors that correspond to the d smallest positive eigenvalues ($v_{n-1}, \ldots, v_{n-d}$ if G is connected).

A popular rationale was suggested by Fiedler (1973) for $d = 1$. First, notice that

$$\begin{aligned}
\sum_{i \sim j} w_{ij} \left[q(i) - q(j)\right]^2 &= -\frac{1}{2} \sum_{i,j} b_{ij} \left[q(i) - q(j)\right]^2 \\
&= -\frac{1}{2} \sum_{i,j} b_{ij} q(i)^2 + \sum_{i,j} q(i) b_{ij} q(j) - \frac{1}{2} \sum_{i,j} b_{ij} q(j)^2 \\
&= -\frac{1}{2} \sum_{i} q(i)^2 \left(\sum_{j} b_{ij} \right) + q^t B q - \frac{1}{2} \sum_{j} q(j)^2 \left(\sum_{i} b_{ij} \right) \\
&= q^t B q.
\end{aligned}$$

Requiring $q^t q = 1$,

$$\begin{aligned}
v_1 &= \operatorname*{argmax} \; q^t B q = \operatorname*{argmin}_{q \perp v_n, \ldots, v_2} q^t B q \\
v_2 &= \operatorname*{argmax}_{q \perp v_1} q^t B q = \operatorname*{argmin}_{q \perp v_n, \ldots, v_3} q^t B q \\
&\vdots \\
v_{n-1} &= \operatorname*{argmax}_{q \perp v_1, \ldots, v_{n-2}} q^t B q = \operatorname*{argmin}_{q \perp v_n} q^t B q \\
e/\sqrt{n} = v_n &= \operatorname*{argmax}_{q \perp v_1, \ldots, v_{n-1}} q^t B q = \operatorname*{argmin} \; q^t B q
\end{aligned}$$

Hence. . .

The eigenvectors used in Laplacian eigenmaps are the eigenvectors that *minimize*

$$F(q) = q^t B q = \sum_{i \sim j} w_{ij} \left[q(i) - q(j) \right]^2 .$$

This quantity has been interpreted as a clustering criterion that encourages placing pairs of vertices with large edge weights close together.

For this rationale to make sense, w_{ij} must measure similarity.

Popular examples of $W = [w_{ij}]$ include:

1. W is an adjacency matrix, i.e.,
   $w_{ij} = 1$ if vertices i and j are connected and $w_{ij} = 0$ otherwise.

2. $w_{ij} = \exp(-\theta \|y_i - y_j\|^2)$

Eigenvector Scaling

It is not clear how to scale the principal components of a Laplacian eigenmap when $d > 1$.

Suppose that we construct the configuration matrix $X = \left[\begin{array}{ccc} \alpha_1 q_1 & \cdots & \alpha_d q_d \end{array} \right]$. Then

$$\begin{aligned} \sum_{i \sim j} w_{ij} \left\| x_i - x_j \right\|^2 &= \sum_{i \sim j} w_{ij} \sum_{k=1}^{d} \alpha_k^2 \left[q_k(i) - q_k(j) \right]^2 \\ &= \sum_{k=1}^{d} \alpha_k^2 \sum_{i \sim j} w_{ij} \left[q_k(i) - q_k(j) \right]^2 \\ &= \alpha_1^2 F\left(q_1\right) + \cdots + \alpha_d^2 F\left(q_d\right). \end{aligned}$$

Hence, assuming that $q_1, \ldots, q_d$ are orthonormal. . .

- $(v_{n-1}, \ldots, v_{n-d})$ minimizes the unweighted clustering criterion

$$F(q_1) + \cdots + F(q_d);$$

- $(\sigma_{n-1} v_{n-1}, \ldots, \sigma_{n-d} v_{n-d})$ minimizes the weighted clustering criterion

$$\sigma_{n-1}^2 F(q_1) + \cdots + \sigma_{n-d}^2 F(q_d);$$

- $(v_{n-1}/\sigma_{n-1}, \ldots, v_{n-d}/\sigma_{n-d})$ minimizes the weighted clustering criterion

$$\frac{1}{\sigma_{n-1}^2} F(q_1) + \cdots + \frac{1}{\sigma_{n-d}^2} F(q_d).$$

Recall that $\sigma_{n-1}^2 \leq \cdots \leq \sigma_{n-d}^2$.

Reciprocal Laplacian Eigenmaps

Let $\Gamma = [\gamma_{ij}]$ be a similarity matrix and let G be the weighted graph with edge weights γ_{ij}.

Let $L = \text{diag}(\Gamma e) - \Gamma$ be the Laplacian of G. Recall that L is symmetric, psd, and $Le = 0$.

Let $0 < \sigma_1^2 \leq \cdots \leq \sigma_r^2$ denote the strictly positive eigenvalues of L, let $v_1, \ldots, v_r$ denote the corresponding eigenvectors, and let

$$X = \left[\begin{array}{c|c|c} \frac{v_1}{\sigma_1} & \cdots & \frac{v_d}{\sigma_d} \end{array} \right].$$

If Δ_2 is such that $\tau(\Delta_2) = L^\dagger$, then X is constructed from Δ_2 by CMDS.

What is $\Delta_2 = \kappa(L^\dagger)$, i.e., what is the transformation from similarity to dissimilarity?

Because L is psd, so is $L^\dagger$; hence, $\Delta_2 = \kappa(L^\dagger)$ is EDM-2.

Two interpretations of dissimilarity implicit in reciprocal Laplacian eigenmaps:

1. G is an electrical circuit. Vertices are terminals. Each edge is a resistor with conductance γ_{ij}. Δ_2 is the *effective resistance* of G, i.e., δ_{ij}^2 is the potential difference between terminals i and j when a unit current source is applied.

   The resistance, δ_{ij}^2, is small when there are many paths with high conductance between terminals i and j. The Euclidean distance δ_{ij} is the *resistance distance*.

2. G is a Markov chain. Vertices are states and the transition probabilities are $p_{ij} = \gamma_{ij}/\gamma_{i+}$. Let $t(j|i)$ denote the expected number of transitions to get from state i to state j for the first time. Then

$$\delta_{ij}^2 = \frac{t(j|i) + t(i|j)}{\gamma_{++}}.$$

   The expected commute time, $\gamma_{++}\delta_{ij}^2$, is small when there are many paths with high probability between states i and j.

Comparison to Isomap

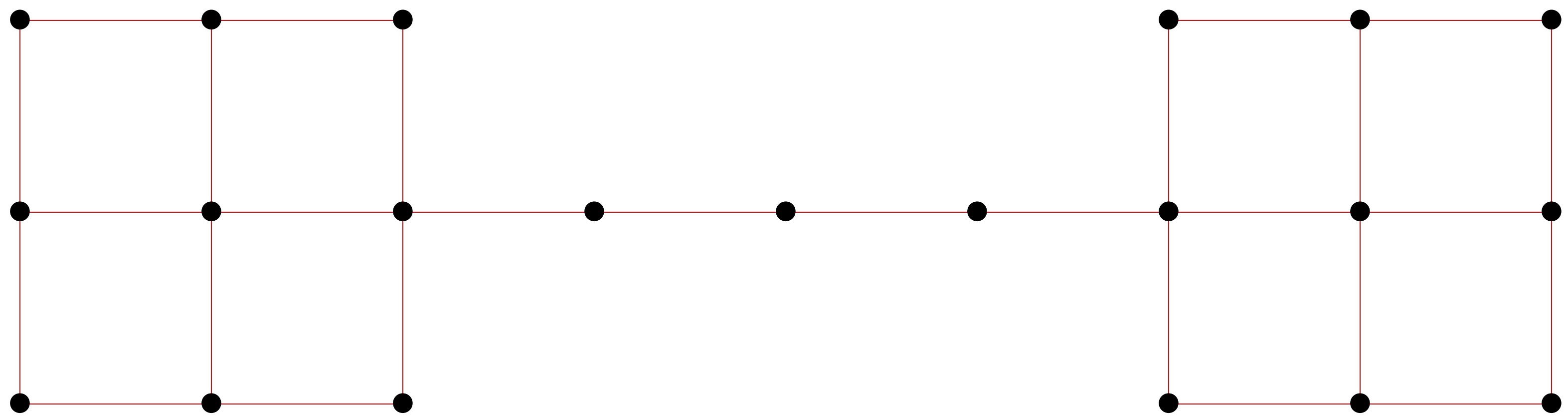
Suppose that $y_1, \ldots, y_n \in \Re^q$ lie on a manifold and fix $\epsilon > 0$.
To represent $y_1, \ldots, y_n$ as $x_1, \ldots, x_n \in \Re^d$:

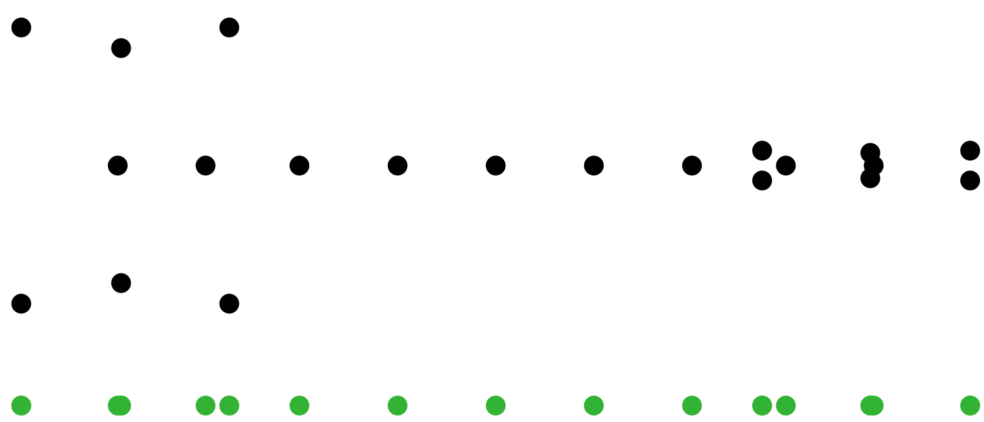
1. Let $w_{ij} = 1$ if $\|y_i - y_j\| < \epsilon$ and $w_{ij} = 0$ otherwise. Define a weighted graph G with n vertices as follows: vertices i and j are connected iff $w_{ij} > 0$, in which case edge $i \sim j$ is weighted by w_{ij}.

2. Let δ_{ij} denote the distance in G between vertices i and j.
   Isomap uses shortest path distance;
   reciprocal Laplacian eigenmaps use resistance distance.

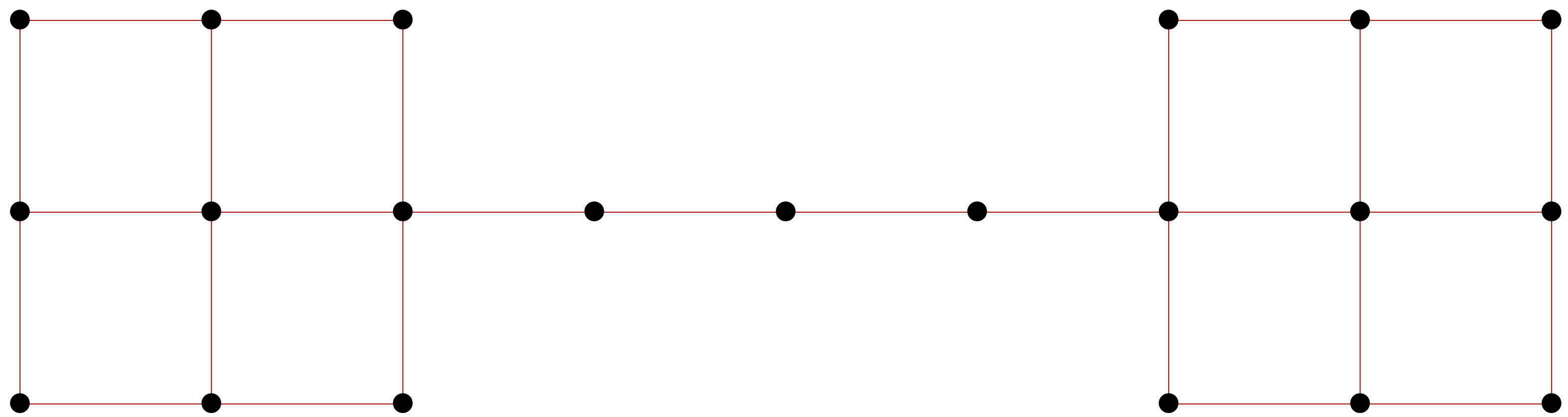
3. Embed $\Delta = [\delta_{ij}]$ in $\Re^d$ by CMDS.

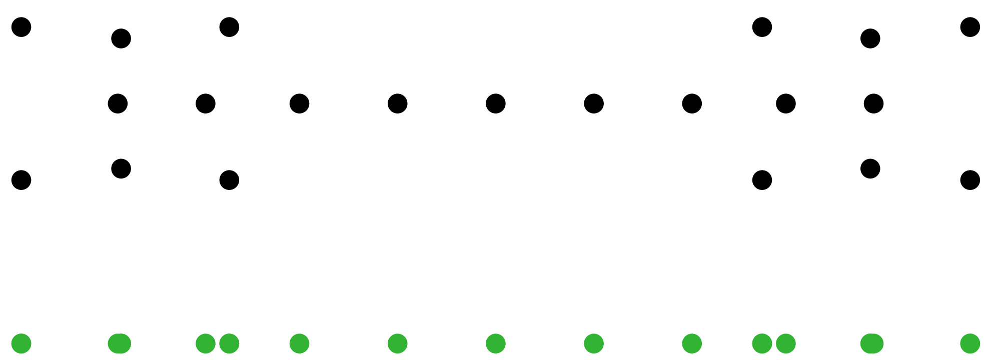
Shortest Path Distance with $\epsilon = 1$



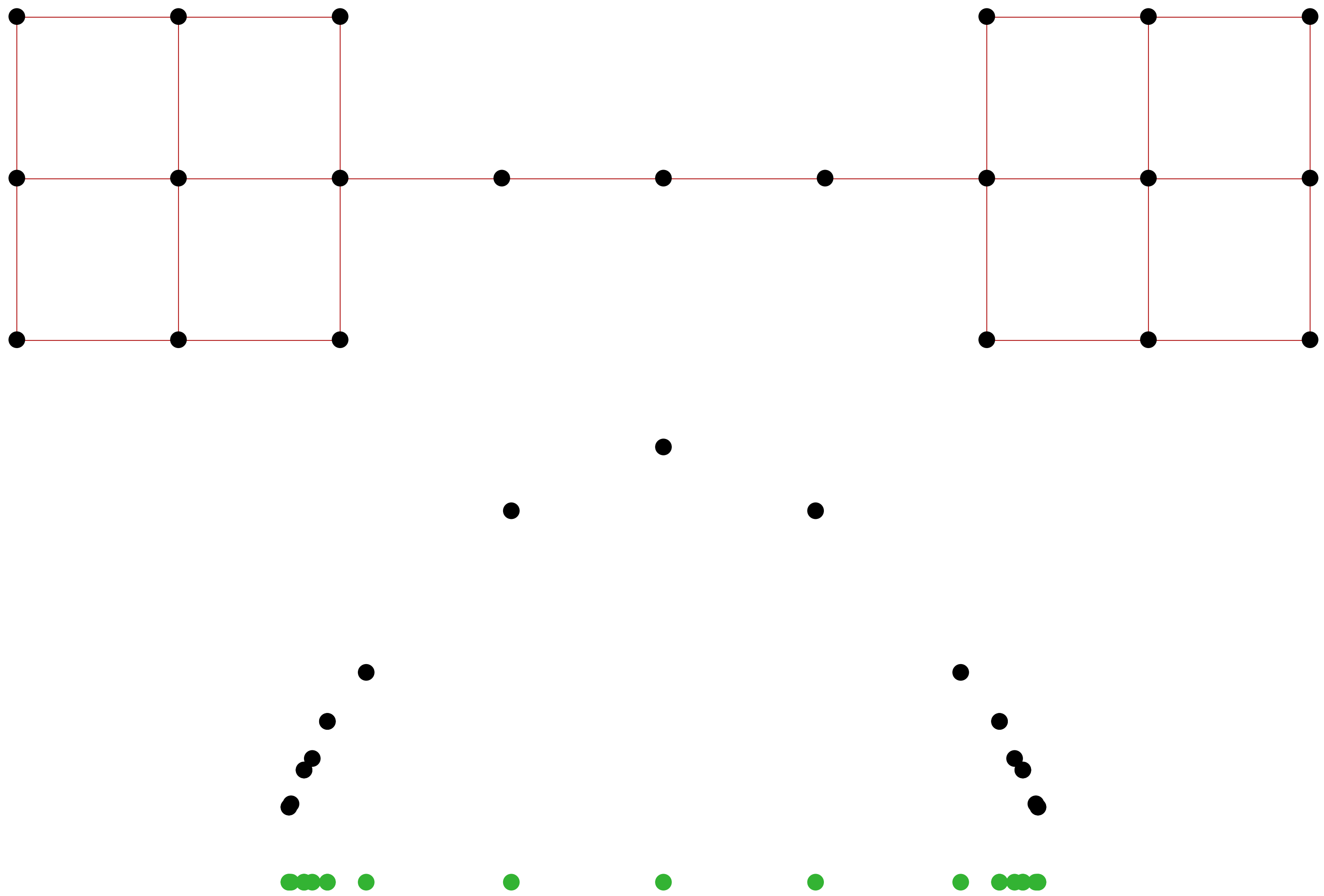


Shortest Path Distance with $\epsilon = 1$

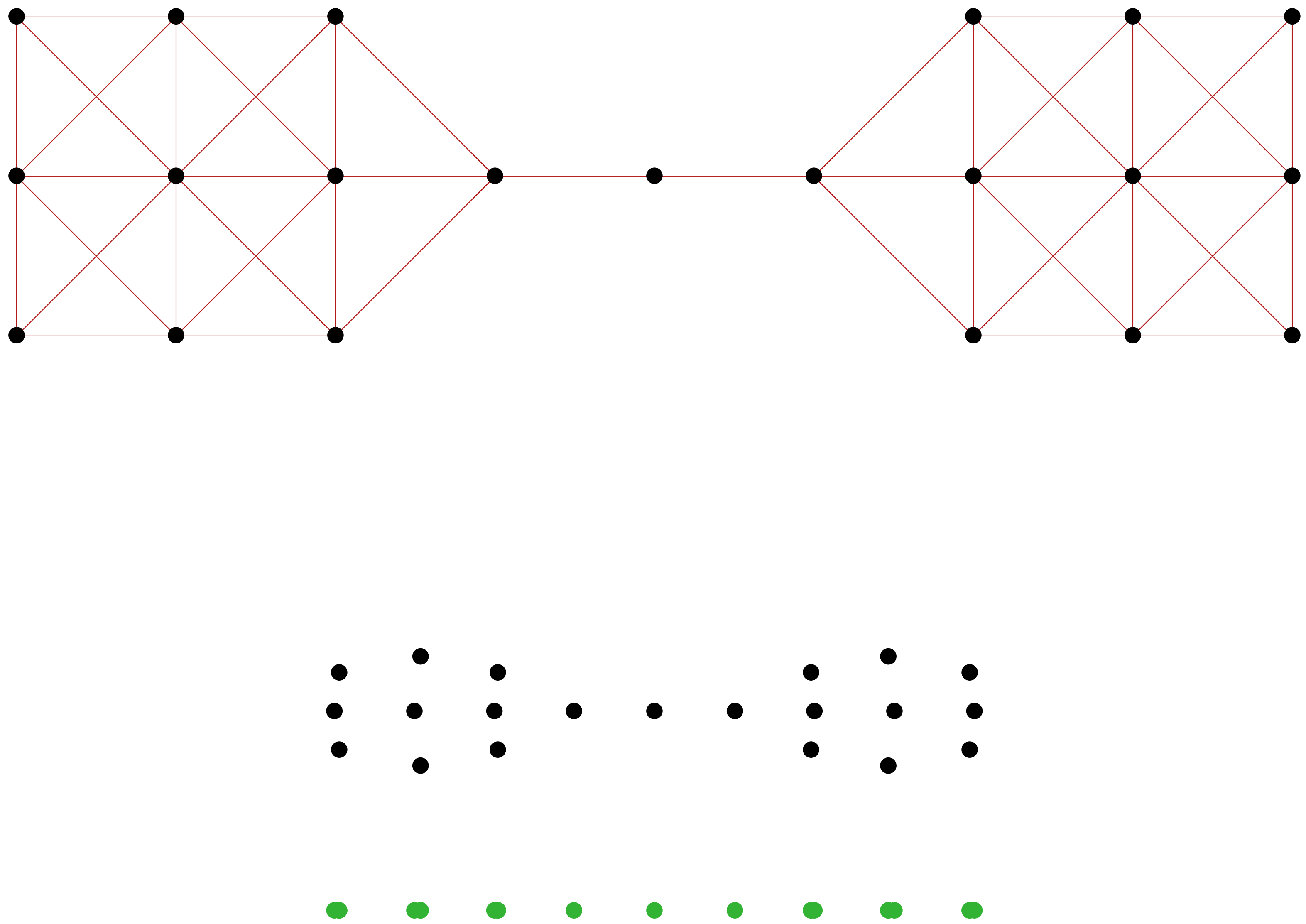




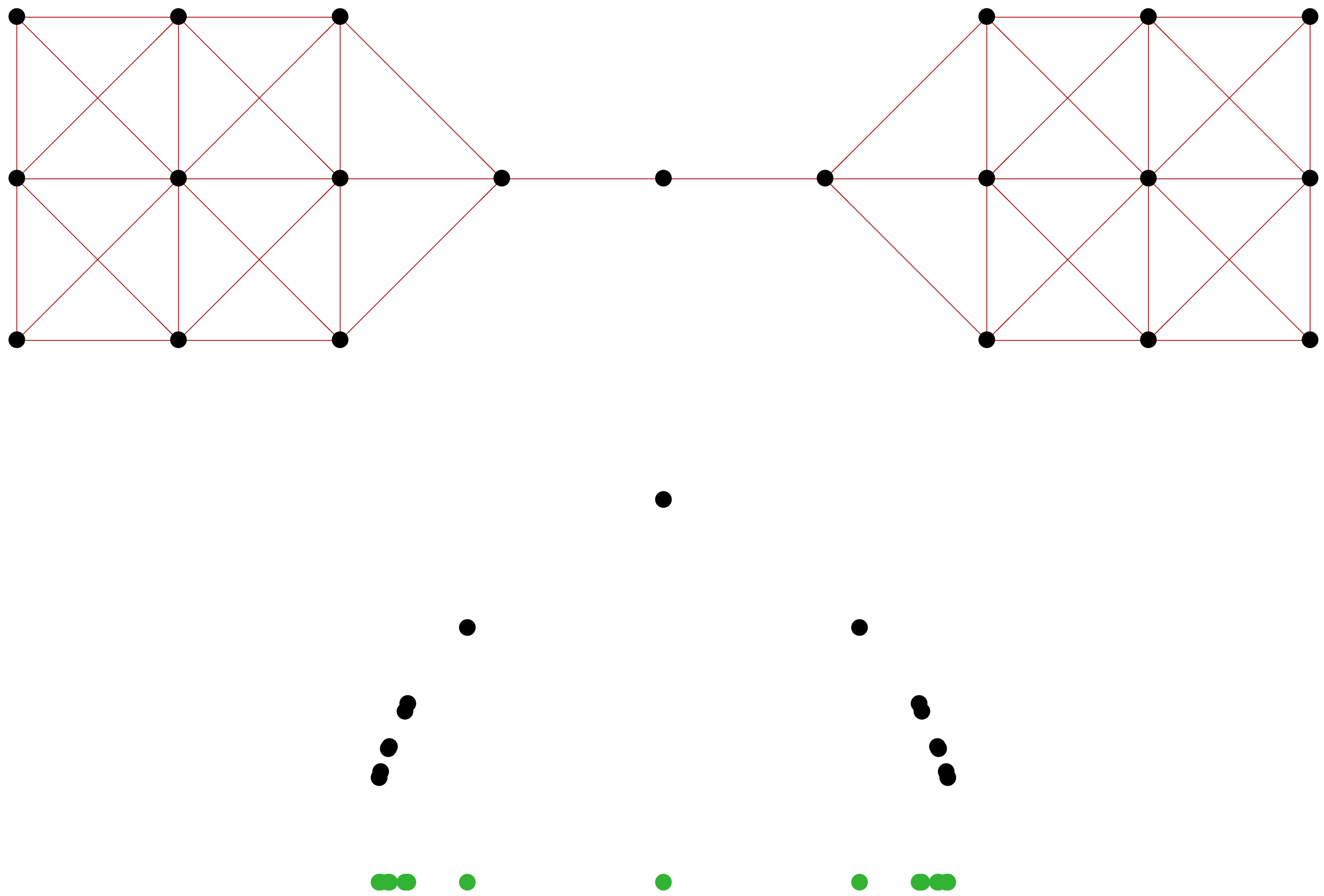
Resistance Distance with $\epsilon = 1$



Shortest Path Distance with $\epsilon = 1.5$



Resistance Distance with $\epsilon = 1.5$



LLE Redux

"LLE maps its inputs into a single global coordinate system. . . By exploiting the local symmetries of linear reconstructions, LLE is able to learn the global structure of nonlinear manifolds. . ." (Roweis & Saul)

How does LLE extract global structure from local information? Two observations:

1. Figure 3 in Roweis & Saul (2000) looks a *lot* like a Laplacian eigenmap.

2. LLE embeds by computing the *bottom* d nonzero eigenvectors of

$$M = (I - C)^t (I - C),$$

   where $Ce = e$ and $\text{diag}(C) = 0$.

Writing $M = T - W$, subject to $\text{diag}(W) = 0$ and $\text{diag}(T) = We$, we obtain $T = I + D$ and $W = C + C^t - C^t C - D$, where $D = \text{diag}(C + C^t - C^t C)$. Hence, LLE constructs an unweighted Laplacian eigenmap from the graph with edge weights W!

How do we judge similarity?

Applying LLE with $K = 4$ to

y_1	y_2	y_3	y_4	y_5	y_6	y_7
1	2	3	4	5	6	7
ϵ	$-\epsilon$	ϵ	$-\epsilon$	ϵ	$-\epsilon$	ϵ
$-\epsilon$	0	0	2ϵ	0	0	$-\epsilon$

results in

$$36W = \begin{bmatrix} - & 98 & 294 & -98 & -98 & 0 & 0 \\ 98 & - & -147 & 67 & 49 & -9 & 0 \\ 294 & -147 & - & 98 & 294 & 49 & -98 \\ -98 & 67 & 98 & - & 98 & 67 & -98 \\ -98 & 49 & 294 & 98 & - & -147 & 294 \\ 0 & -9 & 49 & 67 & -147 & - & 98 \\ 0 & 0 & -98 & -98 & 294 & 98 & - \end{bmatrix}.$$

Such examples challenge the fundamental premise of LLE, that local reconstruction coefficients preserve local Euclidean structure.

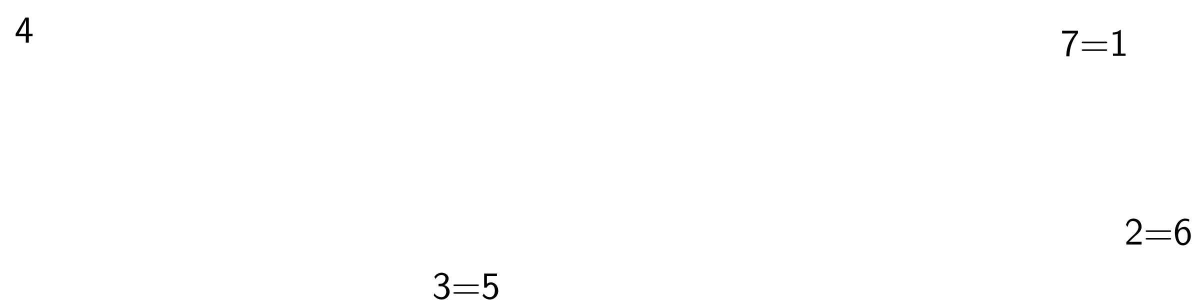
The corresponding dissimilarity matrix is:

$$\Delta = \begin{bmatrix} 0 & 0.1934 & 0.6098 & 0.9309 & 0.5875 & 0.1753 & 0.1154 \\ 0.1934 & 0 & 0.6089 & 1.0004 & 0.6197 & 0.0577 & 0.1753 \\ 0.6098 & 0.6089 & 0 & 0.4581 & 0.2309 & 0.6197 & 0.5875 \\ 0.9309 & 1.0004 & 0.4581 & 0 & 0.4581 & 1.0004 & 0.9309 \\ 0.5875 & 0.6197 & 0.2309 & 0.4581 & 0 & 0.6089 & 0.6098 \\ 0.1753 & 0.0577 & 0.6197 & 1.0004 & 0.6089 & 0 & 0.1934 \\ 0.1154 & 0.1753 & 0.5875 & 0.9309 & 0.6098 & 0.1934 & 0 \end{bmatrix}$$

Δ_2 is the 3-dimensional EDM-2 generated by

x_1	x_2	x_3	x_4	x_5	x_6	x_7
0.2727	0.3282	−0.2727	−0.6564	−0.2727	0.3282	0.2727
0.1054	−0.0584	−0.1054	0.1168	−0.1054	−0.0584	0.1054
0.0577	−0.0289	−0.0115	0.0000	0.0115	0.0289	−0.0577

The 2-dimensional representation constructed by CMDS from Δ is:



Regularization

The problem of finding optimal reconstruction coefficients,

$$\begin{aligned} \text{minimize} \qquad & \mathsf{E}(C) = \sum_{i=1}^{n} \left\| y_i - \sum_{j \in N(i)} c_{ij} y_j \right\|^2 \\ \text{subject to} \qquad & c_{ij} = 0 \text{ if } j \notin N(i) \text{ and } Ce = e, \end{aligned}$$

is separable, i.e., the reconstruction coefficients in row i of C are obtained by minimizing

$$\left\| y_i - \sum_j c_{ij} y_j \right\|^2 = \left\| \sum_j c_{ij} \left(y_i - y_j \right) \right\|^2 = c^t B c,$$

where $B = [b_{jk}] = [\langle y_i - y_j, y_i - y_k \rangle]$, subject to $e^t c = 1$. The minimizer, c_*, is computed by first solving $Bc = e$, then rescaling the solution to satisfy the constraint, i.e., $c_* = B^{-1}e/e^t B^{-1} e$. In case B is singular or ill-conditioned, Roweis & Saul regularize by instead solving $(B + \mu I)c = e$ for $\mu = 10^{-4}$.

Replacing B with I assigns reconstruction coefficient $c_{ij} = 1/K$ to each of y_i's K nearest neighbors. The subsequent construction in $\Re^d$ attempts to represent each x_i as the centroid of its K nearest neighbors. This construction regards the similarity between y_i and y_j as

$$w_{ij} = \left\{ \begin{array}{cl} 1/K & \text{if } j \in N(i) \\ 0 & \text{otherwise} \end{array} \right\} + \left\{ \begin{array}{cl} 1/K & \text{if } i \in N(j) \\ 0 & \text{otherwise} \end{array} \right\} - \frac{\#\text{ points whose nbhds include both } y_i \text{ \& } y_j}{K^2}$$

This seems plausible, but. . .

Example

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
0	0	0	0	0	0	1	-1
0	0	0	0	1	-1	0	0
ϵ	$-\epsilon$	1	-1	0	0	0	0

with $K = 2$ and local reconstruction coefficients e/K results in similarities

$$4W = \begin{bmatrix} \text{—} & -2 & 4 & 1 & 2 & 2 & 2 & 2 \\ -2 & \text{—} & 1 & 4 & 2 & 2 & 2 & 2 \\ 4 & 1 & \text{—} & 0 & 0 & 0 & 0 & 0 \\ 1 & 4 & 0 & \text{—} & 0 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & \text{—} & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & \text{—} & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 & \text{—} & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 & 0 & \text{—} \end{bmatrix}.$$