

1 Setup

Suppose we have high-dimensional data $Z \in \mathbb{R}^{n \times q}$ with MDS embedding $X \in \mathbb{R}^{n \times p}$ for $p \ll q$. Given an out-of-sample point $w \in \mathbb{R}^q$, we want to compare various techniques for embedding w into our p -dimensional embedding. We will consider the following techniques:

1. Minimize absolute error in squared distances:

$$\min_{y \in \mathbb{R}^q} \sum_{i=1}^n |d^2(w, z_i) - d^2(y, x_i)|$$

2. Minimize squared error in distances:

$$\min_{y \in \mathbb{R}^q} \sum_{i=1}^n (d(w, z_i) - d(y, x_i))^2$$

3. Trosset's MDS extension:

$$\min_{y \in \mathbb{R}^q} \left\| \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \\ y \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_n & y \end{bmatrix} - \tau_w(A_2) \right\|^2$$

4. Trosset's approximate solution:

$$y = (X^T X)^{-1} X^T b$$

5. Bengio's MDS extension using

$$\tilde{K}(a, b) = -\frac{1}{2} \left(d^2(a, b) - \frac{1}{2} \sum_{i=1}^n d^2(a, z_i) - \frac{1}{2} \sum_{i=1}^n d^2(b, z_i) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d^2(z_i, z_j) \right)$$

Note we include (1) and (2) because while (2) may seem more natural, CMDS minimizes a cost function of the same form as (1).

2 Trosset Approx = Bengio (= PCA when using Euclidean distance)

It turns out that Trosset's approximate solution and Bengio's solution are actually equal, regardless of the chosen dissimilarity measure. When the dissimilarity measure is Euclidean, both of these methods are equivalent to extending the PCA projection of the original data. To see why this is true, let's dissect the Bengio algorithm. Given the squared dissimilarity matrix D_2 of the original data, the algorithm uses the eigendecomposition of the double-centered matrix

$$M = -\frac{1}{2} H D_2 H.$$

The embedding is then given by

$$y = \left(\frac{1}{\sqrt{\lambda_1}} v_1^T \tilde{w}, \dots, \frac{1}{\sqrt{\lambda_q}} v_q^T \tilde{w} \right), \text{ where } \tilde{w} = \begin{bmatrix} \tilde{K}(w, z_1) \\ \vdots \\ \tilde{K}(w, z_n) \end{bmatrix}$$

and (λ_i, v_i) are the eigenpairs of M . For Trosset's approximate solution, first note X is the MDS embedding of Z , so

$$X = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_q \\ | & & | \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_q} \end{bmatrix}.$$

Hence,

$$y = (X^T X)^{-1} X^T b = \left(\frac{1}{\sqrt{\lambda_1}} v_1^T b, \dots, \frac{1}{\sqrt{\lambda_q}} v_q^T b \right).$$

It turns out that $\tilde{w} = b$ for all out-of-sample points $w \in \mathbb{R}^p$. This gives a clear interpretation of the matrix

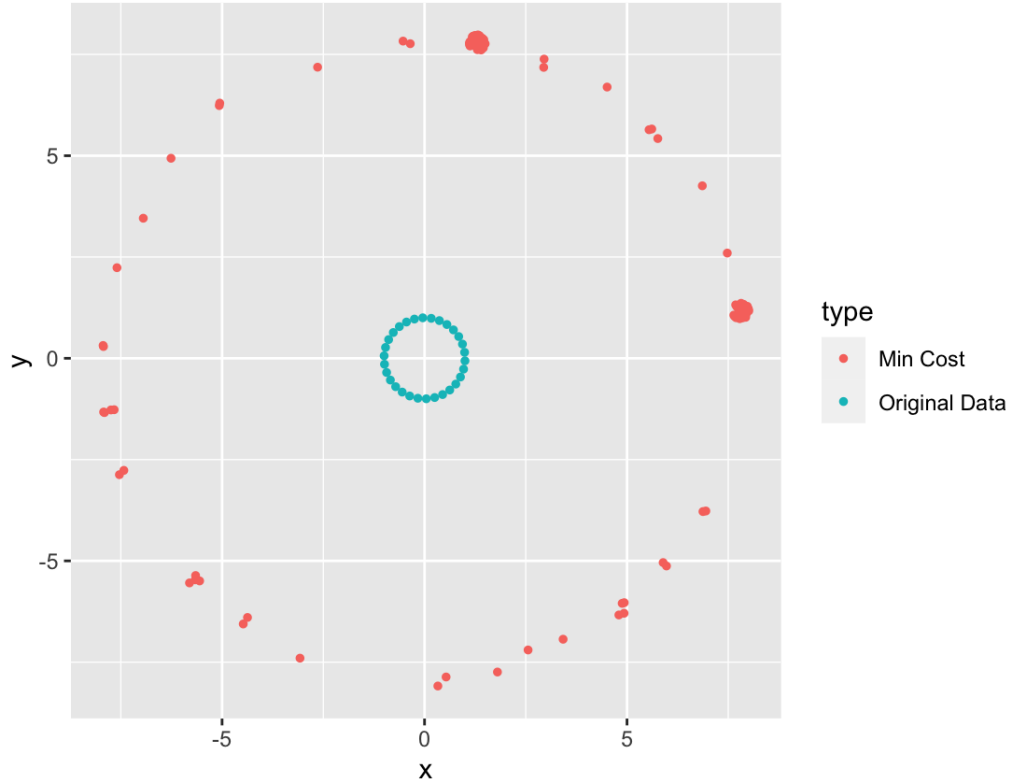
$$B = \tau_w(A_2) = \begin{bmatrix} \tau_e(D_2) & b \\ b^T & \beta \end{bmatrix}$$

for arbitrary squared dissimilarity matrices A_2 . B is the Gram matrix for points z_1, \dots, z_n, w using kernel

$$\tilde{K}(a, b) = -\frac{1}{2} \left(d^2(a, b) - \frac{1}{n} \sum_{i=1}^n d^2(a, z_i) - \frac{1}{n} \sum_{i=1}^n d^2(b, z_i) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d^2(z_i, z_j) \right).$$

3 Min Cost Clustering

When using absolute error in squared distances as the embedding cost function, we see a clustering of embedded points:



This phenomenon is related to the L_1 regularization used in LASSO regression. It's well-known that the L_1 regularization term in LASSO regression forces some of the coefficient estimates to 0. The same is occurring

here. The absolute cost function is embedding points so that $|d^2(w, z_i) - d^2(y, x_i)| = 0$ for some i . The points in each cluster all have $|d^2(w, z_i) - d^2(y, x_i)| = 0$ for the same i . For example in the configuration above, each point in the large cluster towards the top satisfies $|d^2(w, z_i) - d^2(y, x_i)| = 0$ for exactly $i = 2, 16$. Each point in the large cluster to the right satisfies $|d^2(w, z_i) - d^2(y, x_i)| = 0$ for exactly $i = 10, 26$.