MDS miscellany - Lawrence Cayton

This document discusses some basic results on classical multidimensional scaling (MDS). MDS is a technique for finding a configuration of points in euclidean space whose interpoint (ℓ_2) distances are close to the given dissimilarities. The dissimilarities are represented in a matrix $D \in \Re^{n\times n}$, where d_{ij} is the dissimilarity between x_i and x_j. Throughout, we assume that the matrix D is symmetric and has 0's on the diagonal, but we do not assume that D is euclidean.

1 Algorithm and Fundamentals

In this section, we will review the algorithm and motivation for MDS. In particular, we will see that when the dissimilarities are euclidean, MDS finds a configuration whose interpoint distances match the given dissimilarities perfectly. The algorithm follows.

`mds(`D`)`

1. Calculate A by setting $a_{ij} := -\frac{1}{2}d_{ij}^2$.
2. Set $B := HAH$, where $H = (I - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}})$.
3. Compute the spectral decomposition of B: $B = U\Lambda U^{\mathrm{T}}$.
4. Form U_+ and Λ_+ by setting the negative eigenvalues and corresponding eigenvectors of B to zeros.
5. Return $X := U_+\Lambda_+^{1/2}$.

Interpoint distances are invariant under translations and rotations of a configuration. Thus, to adequately constrain the search space for a configuration yielding interpoint euclidean distances that match the given dissimlarities, we look only for *mean-centered* configurations —*i.e.*, we demand that

$$\sum_k X_{kj} = 0, \text{ for all } j.$$

Let us consider what happens if D is a euclidean distance matrix. We show that the configuration X returned by MDS has interpoint distances identical to those in D. First, assuming that B is positive semidefinite, B has no negative eigenvalues, so $U_+ = U$ and $\Lambda_+ = \Lambda$. Then $B = XX^{\mathrm{T}}$, so B is the inner product matrix for X. Next we rewrite B in terms of its definition (HAH) :

$$b_{ij} = -1/2\left(d_{ij}^2 - 1/n\sum_k d_{kj}^2 - 1/n\sum_k d_{ik}^2 + 1/n^2\sum_{k,l} d_{kl}^2\right) \quad (1)$$

$$= -1/2\left(d_{ij}^2 - \overline{d_i^2} - \overline{d_j^2} + \overline{d^2}\right). \quad (2)$$

Using this expansion, we calculate the distance between two points in X:

$$\begin{aligned}\|x_i - x_j\|^2 &= \|x_i\|^2 + \|x_j\|^2 - 2x_i \cdot x_j \\ &= b_{ii} + b_{jj} - 2b_{ij} \\ &= d_{ij}^2 \text{ after substituting in (2).}\end{aligned}$$

We assumed that B is positive semidefinite in the preceding analysis so that B is given precisely by its spectral decomposition; if B is not positive semidefinite, step 4 of the MDS procedure alters the matrix. We now show that this assumption is correct.

Theorem 1. *D is a euclidean distance matrix if and only if B is positive semidefinite.*

Proof. If D is a euclidean distance matrix, there exist points $x_1, \ldots, x_n \in \Re^d$ such that $\|x_i - x_j\|^2 = d_{ij}^2$ for all i and j. Moreover, since we are only interested in matching the distances, we may assume that these points are mean centered —*i.e.* $\sum_i x_{ij} = 0$, for all j. Expanding the norm, we have that $d_{ij}^2 = {x_i}^{\mathrm{T}} x_i + {x_j}^{\mathrm{T}} x_j - 2{x_i}^{\mathrm{T}} x_j$, so

$$ {x_i}^{\mathrm{T}} x_j = -1/2 \left(d_{ij}^2 - {x_i}^{\mathrm{T}} x_i - {x_j}^{\mathrm{T}} x_j \right). \tag{3}$$

We show that this inner product can be written in using only terms of D. Using the expansion for d_{ij}^2, we have that

$$1/n \sum_i d_{ij}^2 = 1/n \left(\sum_i {x_i}^{\mathrm{T}} x_i + \sum_i {x_j}^{\mathrm{T}} x_j - 2 \sum_i {x_i}^{\mathrm{T}} x_j \right). \tag{4}$$

Because of the centering, the last term drops out, leaving

$$1/n \sum_i d_{ij}^2 = 1/n \sum_i {x_i}^{\mathrm{T}} x_i + {x_j}^{\mathrm{T}} x_j. \tag{5}$$

Similarly,

$$\begin{aligned}1/n \sum_j d_{ij}^2 &= 1/n \sum_j {x_j}^{\mathrm{T}} x_i + {x_i}^{\mathrm{T}} x_i, \text{ and} & (6)\\ 1/n^2 \sum_{i,j} d_{ij}^2 &= 2/n \sum_i {x_i}^{\mathrm{T}} x_i. & (7)\end{aligned}$$

Subtracting (7) from the sum of (5) and (6) leaves ${x_j}^{\mathrm{T}} x_j + {x_i}^{\mathrm{T}} x_i$ on the right hand side, so we may rewrite (3) as

$$ {x_i}^{\mathrm{T}} x_j = -1/2 \left(d_{ij}^2 - 1/n \sum_i d_{ij}^2 - 1/n \sum_j d_{ij}^2 + 1/n^2 \sum_{i,j} d_{ij}^2 \right).$$

However, this is precisely the definition of B, so B is the inner product matrix for $\{x_i\}$. Then $B = XX^{\mathrm{T}}$, where the ith row of X is x_i, implying that it is positive semidefinite.

We already proved the reverse direction: If B is positive semidefinite, then $B = XX^{\mathrm{T}}$, and the interpoint distances of X match D precisely, so D must be euclidean.

$\square$

So, if D is a euclidean distance matrix, classical MDS will produce a configuration with interpoint distances given exactly by D. If D is not a euclidean distance matrix, however, the procedure will introduce some distortion. This distortion will be analyzed shortly.

2 Quick MDS facts

We now demonstrate four quick facts that come in useful when analyzing classical MDS. These appear mostly for reference.

Fact 1: B is mean-centered.

Proof. Since $B = HAH$ and H and A are both symmetric, B is symmetric. So, it suffices to show that B has mean-centered columns.

$$\begin{aligned}\sum_i b_{ij} &= \sum_i -1/2\left(d_{ij}^2 - \overline{d_i^2} - \overline{d_j^2} + \overline{d^2}\right) \\ &= -1/2\left(n\overline{d_j^2} - n\overline{d^2} - n\overline{d_j^2} + n\overline{d^2}\right) = 0.\end{aligned}$$

$\square$

Fact 2: B has the (eigenvector, eigenvalue) pair $(\mathbf{1}, 0)$.

Proof. Since all the rows of B are mean centered, $B_i \cdot \mathbf{1} = 0$, where B_i is the ith row of B. But then $B\mathbf{1} = \mathbf{0}$, or $B\mathbf{1} = 0\mathbf{1}$. $\square$

Fact 3: If u is an eigenvector of B not equal to $c \cdot \mathbf{1}$ (some $c \in \Re$), $\sum_i u_i = 0$.

Proof. Suppose that $Bu = \lambda u$. Then,

$$\sum_i (\lambda u)_i = \sum_i (Bu)_i = \sum_i B_i \cdot u = u \cdot \left(\sum_i B_i\right) = 0.$$

Thus if $\lambda \neq 0$, $\sum_i u_i = 0$. $\square$

Fact 4: $B^* = U_+\Lambda_+^{1/2}U_+$ is mean centered.

Proof. Suppose that $\lambda_{r+1}, \ldots \lambda_n$ are the negative eigenvalues of B. Then, the kth column of B^* is

$$B^*_k = \sum_{i=1}^{r} \lambda_i u_{ik} {u_i}^{\mathrm{T}}.$$

Summing over the elements of B^*_k, we get

$$\begin{aligned}
\sum_{j=1}^{n} B^*_{kj} &= \sum_{j=1}^{n} \sum_{i=1}^{r} (\lambda_i u_{ik} {u_i}^{\mathrm{T}})_j \\
&= \sum_{i=1}^{n} \lambda_i u_{ik} \sum_{j} u_{ij} \\
&= 0, \text{ by fact 3.}
\end{aligned}$$

□

3 Error Analysis

When D is a euclidean distance matrix, classical MDS finds a configuration of points matching D perfectly. When D is not euclidean, there is no hope of finding such a configuration. We presently analyze the error induced by MDS in this situation. We still demand that D is symmetric and has zeros on the diagonal.

First, we show that MDS minimizes $\|H(D-D^*)H\|$, where D^* is the distance matrix for X. Throughout, $\|\cdot\|$ denotes the Frobenius norm.

Lemma 1. *Let B be a real symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, where $\lambda_{r+1}, \ldots, \lambda_n < 0$. Then, for any $n \times n$ positive semidefinite matrix B^*,*

$$\|B - B^*\|^2 \geq \sum_{i=r+1}^{n} \lambda_i^2.$$

Proof. First note that the frobenius norm of a matrix A is equal to the frobenius norm of UA if U is orthonormal. This follows since

$$\|UA\|^2 = \mathrm{tr}((UA)(UA)^{\mathrm{T}}) = \mathrm{tr}(AA^{\mathrm{T}}UU^{\mathrm{T}}) = \mathrm{tr}(AA^{\mathrm{T}}) = \|A\|^2.$$

Post-multiplication by an orthonormal matrix also leaves the frobenius norm unchanged.

Let $U\Lambda U^{\mathrm{T}}$ be the spectral decomposition of B. Then, for any positive semidefinite B^*

$$\|B - B^*\| = \|U^{\mathrm{T}}(B - B^*)U\| = \|\Lambda - U^{\mathrm{T}}B^*U\|.$$

The diagonal elements of the second term are given by $(U^{\mathrm{T}}B^*U)_{ii} = {u_i}^{\mathrm{T}}B^*u_i$, where u_i is the i-th eigenvector of B. Since B^* is positive semidefinite, ${u_i}^{\mathrm{T}}B^*u_i$

must be nonnegative. Thus, by disregarding the off-diagonal elements we obtain

$$\|\Lambda - U^{\mathrm{T}} B^* U\|^2 \geq \sum_{i=1}^{n} (\lambda_i - {u_i}^{\mathrm{T}} B^* u_i)^2 \geq \sum_{i=r+1}^{n} \lambda_i^2.$$

□

Theorem 2. *The configuration of points given by MDS, X, minimizes the error $\|H(D - D^*)H\|$, where D^* is the distance matrix of X.*

Proof. If D is not euclidean, then B will have negative eigenvalues $\lambda_{r+1}, \ldots, \lambda_n$. Disregarding those negative eigenvalues (step 4) to form B^*, we induce error

$$\|B - B^*\|^2 = \|U \Lambda U^{\mathrm{T}} - U_+ \Lambda_+ {U_+}^{\mathrm{T}}\|^2 = \sum_{i=r+1}^{n} \lambda_i^2.$$

By the previous lemma, this distortion is minimal. □

Though the preceeding theorem shows that the distortion is minimized in some sense, we are more interested in comparing D and D^* directly, since the goal of MDS is to create a configuration with interpoint distances matching D.

Theorem 3. *Let D^* be the interpoint distances of the configuration determined by classical MDS on D and suppose that $\lambda_{r+1}, \ldots, \lambda_n$ are the negative eigenvalues of B. Then,*

$$\sum_{i,j} \left| d_{ij}^2 - {d_{ij}^*}^2 \right| = 2n \sum_{i=r+1}^{n} |\lambda_i|.$$

Proof. We rewrite the distances using the matrices B and B^*:

$$\sum_{i,j} \left| d_{ij}^2 - {d_{ij}^*}^2 \right| = \sum_{i,j} \left| b_{ii} + b_{jj} - 2b_{ij} - (b_{ii}^* + b_{jj}^* - 2b_{ij}^*) \right|.$$

Let $\mathbf{x} \in \Re^n$ be the vector composed of all zeros except for $x_i = -1$ and $x_j = 1$. Then, the rhs of the above term (inside the sum) equals $\mathbf{x}^{\mathrm{T}}(B - B^*)\mathbf{x}$. Recall that B^* is obtained from B by computing the spectral decomposition of B and setting all negative eigenvalues and their corresponding eigenvectors to zeros. As a result, $B - B^*$ is negative semidefinite, so $\mathbf{x}^{\mathrm{T}}(B - B^*)\mathbf{x} \leq 0$ for all $\mathbf{x}$. Because there are no positive terms, we may move the absolute value signs outside of the sum. So, we have

$$\begin{aligned} \sum_{i,j} \left| d_{ij}^2 - {d_{ij}^*}^2 \right| &= \left| \sum_{i,j} \left[2(b_{ij} - b_{ij}^*) - (b_{ii} - b_{ii}^*) - (b_{jj} - b_{jj}^*) \right] \right| \\ &= \left| 2 \sum_{i,j} (b_{ij} - b_{ij}^*) - n\mathrm{tr}(B - B^*) - n\mathrm{tr}(B - B*) \right|. \end{aligned}$$

But, B and B^* are centered, so the leftmost term is equal to zero, leaving $|2n\mathrm{tr}(B - B^*)|$. The trace of a matrix equals the sum of its eigenvalues, so

$$|2n\mathrm{tr}(B - B^*)| = 2n \sum_{i=r+1}^{n} |\lambda_i|,$$

where $\lambda_{r+1}, \ldots, \lambda_n$ are the negative eigenvalues of B. $\square$

The preceding theorem gives a useful characterization of the error induced by classical MDS. We now give an example for which classical MDS returns a suboptimal solution (with respect to $\|D - D^*\|$).

Example. Let D be the ℓ_1 interpoint distance matrix for the unit square. This square can not be isometrically embedded in euclidean space. Figure 1 shows the embedding of D found by classical MDS. Notice that the distances between opposing points (2 units) is preserved exactly, so that the distortion is focused on distances between neigboring points. Let D^* be the squared distance matrix for for this configuration. Then,

$$\|D - D^*\| = \sqrt{8\left(1 - \sqrt{2}\right)^2},$$

since $(1 - \sqrt{2})$ is the distortion induced on any two neighboring points.

If the top point is moved down to $(0, 3/4)$, the total error is

$$\sqrt{4\left(1 - \sqrt{2}\right)^2 + 2(1/4)^2 + 4(1/4)^2},$$

which is less than the error of the MDS solution. $\square$

The previous example shows that classical MDS is not guaranteed to find a configuration that matches D best, even though that is the objective! Metric MDS, which uses an iterative procedure to directly minimize the distortion, often produces better results at the expense of significant computation time.

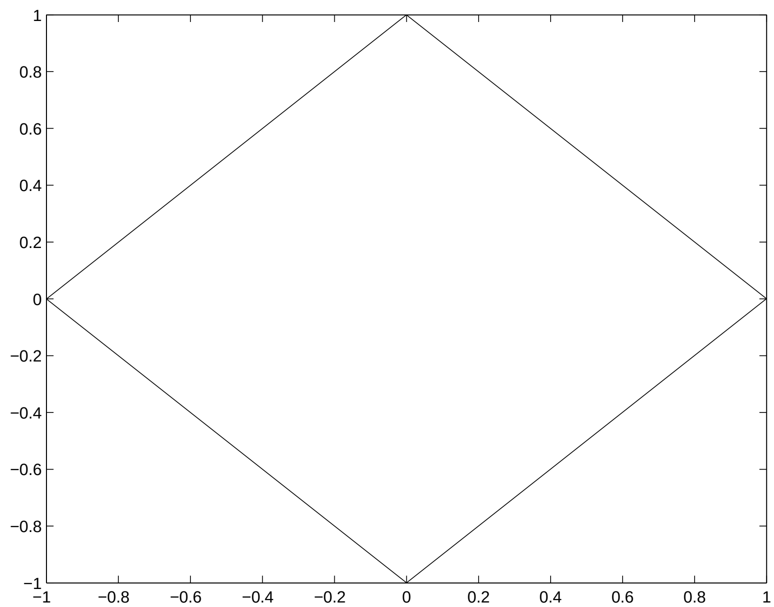

Figure 1: The classical MDS embedding of the unit square.