Chapter 1

Multidimensional Scaling: More complete proof and some insights not mentioned in class

1.1 Motive of MDS

We are given the pair-wise (Euclidean/non-Euclidean) distance matrix $\mathbf{D}^X$ of N points and we are asked to find a set of N points $\mathbf{Y} = \{\mathbf{y}_i \ for \ i \in [1, N]\}$ in a k dimensional space so that the pair-wise Euclidean distance matrix $\mathbf{D}^Y$ calculated using $\mathbf{Y}$ is the closest possible approximation of $\mathbf{D}^X$.

1.2 Notation and convention

For ease of understanding and relatively simpler linear algebra manipulations for the author, we will follow these conventions throughout the document

- Vectors are boldface small: $\mathbf{y}$
- Matrices are capital and boldface: $\mathbf{X}$
- Each column of the data matrix contains one data vector
- $\mathbf{e}$ is an $N \times 1$ column vector of all ones

1.3 The main steps of MDS

1. **Input:** $N \times N$ data matrix $\mathbf{D}^X$, dimension k
2. $\mathbf{B}^X = -\frac{1}{2}\mathbf{H}\mathbf{D}^X\mathbf{H}$, here, $\mathbf{H} = I_N - \frac{1}{N}\mathbf{e}\mathbf{e}^T$ with $\mathbf{e}$ being an $N \times 1$ column vector of all ones.
3. The k-rank SVD of centered matrix $\mathbf{B}^X = (\mathbf{U}\mathbf{D}^{1/2})(\mathbf{D}^{1/2}\mathbf{U}^T) = \mathbf{Y}^T\mathbf{Y}$, therefore, $\mathbf{U}$ is an $N \times k$ matrix and $\mathbf{D}$ is an $k \times k$ diagonal matrix with the k largest singular values on the diagonal and $\mathbf{Y} = \mathbf{D}^{1/2}\mathbf{U}^T$ is a $k \times N$ matrix.
4. The set of N k-dimensional embeddings/points are the N columns of $\mathbf{Y}$.

1.4 The insight of MDS

We start the proof of the algorithm with an assumption of a *hypothetical* set of N points $\mathbf{X} = \{\mathbf{x}_i \ for \ i \ \in [1, N]\}$ (with each $\mathbf{x}_i$ as one column of $\mathbf{X}$) in d dimensions. Note that we have neither $\mathbf{X}$ nor do we know the value of d. We are only supplied with the pair-wise Euclidean distances for $\mathbf{X}$, given as the distance matrix $\mathbf{D}^X$. Therefore, each element of $\mathbf{D}^X$ can be written as

$$(\mathbf{D}^X_{ij})^2 = (\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j) = \|\mathbf{x}_i\|^2 - 2\mathbf{x}_i^T\mathbf{x}_j + \|\mathbf{x}_j\|^2 \tag{1.1}$$

We can easily see that

$$\mathbf{D}^X = \mathbf{Z} - 2\mathbf{X}^T\mathbf{X} + \mathbf{Z}^T \tag{1.2}$$

Here, $\mathbf{Z} = \mathbf{z}\mathbf{e}^T$ and $\mathbf{z} = [\|\mathbf{x}\|_1^2 \ \|\mathbf{x}\|_2^2 \ldots \|\mathbf{x}\|_N^2]^T$ i.e. an $N \times 1$ vector with each element being the squared norm of *hypothetical* point set $\mathbf{X}$. Therefore, $\mathbf{Z}$ takes the form

$$\mathbf{Z} = \begin{bmatrix} \|\mathbf{x}_1\|^2 & \|\mathbf{x}_1\|^2 & \cdots & \|\mathbf{x}_1\|^2 \\ \|\mathbf{x}_2\|^2 & \|\mathbf{x}_2\|^2 & \cdots & \|\mathbf{x}_2\|^2 \\ \vdots & \vdots & \ddots & \vdots \\ \|\mathbf{x}_N\|^2 & \|\mathbf{x}_N\|^2 & \cdots & \|\mathbf{x}_N\|^2 \end{bmatrix} \tag{1.3}$$

Question- Compute $\frac{1}{N}\mathbf{Z}\mathbf{e}\mathbf{e}^T$ and $\frac{1}{N}\mathbf{e}\mathbf{e}^T\mathbf{Z}^T$.

Now, let's translate the mean of the set of *hypothetical* point set ($\mathbf{X}$) to the origin. Note that this operation does not change the Euclidean distance between any pairs of points but does a wonderful thing, which we will see shortly. To carry out this operation we simply compute the mean of all points contained in $\mathbf{X}$

$$\mathbf{m}^X = \frac{1}{N}\mathbf{X}\mathbf{e} \quad \text{(see it yourself)} \tag{1.4}$$

Let's define the centering operation as

$$\mathbf{H} = \mathbf{I}_N - \frac{1}{N}\mathbf{e}\mathbf{e}^T \text{ (see how } \mathbf{XH} \text{ is mean centered?)} \tag{1.5}$$

Let's now apply a "double centering" operation to equation 1.2 to get (see it yourself)

$$\begin{aligned} \mathbf{A}^X &= \mathbf{H}\mathbf{D}^X\mathbf{H} = -2\tilde{\mathbf{X}}^T\tilde{\mathbf{X}} \\ \mathbf{B}^X &= -\frac{1}{2}\mathbf{A} = -\frac{1}{2}\mathbf{H}\mathbf{D}^X\mathbf{H} = \tilde{\mathbf{X}}^T\tilde{\mathbf{X}} \end{aligned} \tag{1.6}$$

Here, $\tilde{\mathbf{X}}$ is the matrix with i^{th} column as the "mean subtracted" *hypothetical* points $\tilde{\mathbf{x}}_i$. Now we are done and the rest is just simple linear algebra results. Remember, the task was to find a *concrete* set of N points ($\mathbf{Y}$) in k dimensions so that the pairwise Euclidean distances between all the pairs in the *concrete* set $\mathbf{Y}$ is a close approximation to the pair-wise distances given to us in the matrix $\mathbf{D}^X$ i.e. we want to find $\mathbf{D}^Y$ such that

$$\mathbf{D}^Y = argmin_{rank(\mathbf{D}^Y \leq k)} \left\| \mathbf{D}^X - \mathbf{D}^Y \right\|_F^2 \tag{1.7}$$

Note that after applying the "double centering" operation to both $\mathbf{X}$ and $\mathbf{Y}$, eqn 1.7 yields

$$\mathbf{B}^Y = argmin_{rank(\mathbf{B}^Y \leq k)} \left\| \mathbf{B}^X - \mathbf{B}^Y \right\|_F^2 = \left\| \tilde{\mathbf{X}}^T\tilde{\mathbf{X}} - \tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}} \right\|_F^2 \tag{1.8}$$

The above equation is a well known optimization problem that can be solved via Singular Value Decomposition (SVD) of $\mathbf{B}^X$. The second equality is due to the "double centering" operation on the distance matrix $\mathbf{D}^Y$ as well. Therefore, now we have eliminated the translational freedom from $\mathbf{Y}$ and we can be sure that whatever embedding we get will have zero mean. This is important because otherwise the embedding points would have been arbitrary up to a translational degree of freedom (there is still a rotational degree of freedom in the points).

$$\mathbf{B}^X \approx \mathbf{U}\mathbf{D}\mathbf{U}^T = (\mathbf{U}\mathbf{D}^{1/2})(\mathbf{D}^{1/2}\mathbf{U}^T) = \tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}} \tag{1.9}$$

Here, $\mathbf{U}$ is $N \times k$ matrix and $\mathbf{D}$ is $k \times k$ diagonal matrix with k largest singular values on the diagonal and $\tilde{\mathbf{Y}} = \mathbf{D}^{1/2}\mathbf{U}^T$ is $k \times N$ matrix. Finally, we get N embedding points in k dimension as the column vectors of $\tilde{\mathbf{Y}}$ with the property that it has a zero mean and is the embedding in dimension k or less that best preserves the pair-wise Euclidean distances given to us in the form of $\mathbf{D}^X$.