Supervised Learning, HW03

Ofer Lipman, 201510435 and Daniel Shterenberg, 305199507

June 2021

1 Classical MDS

Let $\boldsymbol{R} \in \mathbb{R}^{d\times d}$ be an orthogonal matrix.

1.1

Prove that for all $\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathbb{R}^d : \|\boldsymbol{R}\boldsymbol{x}_i - \boldsymbol{R}\boldsymbol{x}_j\|_2 = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$

$$\begin{aligned}
\left\|\boldsymbol{Rx_i} - \boldsymbol{Rx_j}\right\|_2 &= \left\|R(x_i - x_j)\right\|_2 \\
&= \sqrt{\left(R\left(x_i - x_j\right)\right)^T R\left(x_i - x_j\right)} \\
&= \sqrt{\left(x_i - x_j\right)^T R^T R\left(x_i - x_j\right)} \\
&= \sqrt{\left(x_i - x_j\right)^T I_d\left(x_i - x_j\right)} \\
&= \sqrt{\left(x_i - x_j\right)^T \left(x_i - x_j\right)} \\
&= \left\|\boldsymbol{x_i} - \boldsymbol{x_j}\right\|_2
\end{aligned}$$

$\blacksquare$

2 Metric MDS

The metric MDS objective is given by:

$$\min_{\boldsymbol{Z}\in\mathbb{R}^{d\times N}} \|\boldsymbol{\Delta}_x - \boldsymbol{D}_z\|_F^2$$

where:

- $\boldsymbol{\Delta}_x[i,j] = d(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is a given matrix.
- $\boldsymbol{D}_z[i,j] = \|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2$.

Consider the surrogate function: $g\left(\boldsymbol{Z}, \tilde{\boldsymbol{Z}}\right) = \|\boldsymbol{\Delta}_x\|_F^2 + 2N\,\mathrm{Tr}\{\boldsymbol{Z}\boldsymbol{J}\boldsymbol{Z}^T\} - 4\left\langle \boldsymbol{Z}^T\widetilde{\boldsymbol{Z}}, \boldsymbol{B}\right\rangle$ where:

- $\boldsymbol{J} = \boldsymbol{I} - \frac{1}{N}\boldsymbol{1}\boldsymbol{1}^T$ is the centering matrix.
- $\boldsymbol{B} = \boldsymbol{C} - \mathrm{diag}(\boldsymbol{C}\boldsymbol{1})$
- $\boldsymbol{C}[i,j] = \begin{cases} 0 & i=j \\ -\frac{\boldsymbol{\Delta}_x[i,j]}{\widetilde{\boldsymbol{D}}_{\tilde{z}}[i,j]} & i\neq j \end{cases}$
- $\widetilde{\boldsymbol{D}}_{\tilde{z}}[i,j] = \|\tilde{\boldsymbol{z}}_i - \tilde{\boldsymbol{z}}_j\|_2$

2.1

Prove that: $\boldsymbol{B}\boldsymbol{J} = \boldsymbol{B}$

$$\begin{aligned} BJ &= (C - diag(C1))J = (C - diag(C1))(I_N - \frac{1}{N}1_N1_N^T) \\ &= \underbrace{(C - diag(C1))I_N}_{=B} - \underbrace{\frac{1}{N}(C - diag(C1))(1_N1_N^T)}_{denote\ B_1} \end{aligned}$$

We will denote $C' = C - diag(C1)$ and we will examine its values:

$$C'_{ij} = C_{ij} - diag(C1)_{ij} = \begin{cases} C_{ii} - \sum_{j=1}^N C_{ij} & i=j \\ C_{ij} & i\neq j \end{cases}$$

Lets examine a general element ij:

$$\begin{aligned} (C'1_N1_N^T)_{ij} &= \begin{pmatrix} C'_{i1} & C'_{i2} & \cdots & C'_{ii} & \cdots & C'_{iN} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \\ &= \underbrace{C'_{i1} + C'_{i2} + \cdots + C'_{iN}}_{\text{No } C'_{ii}} + C'_{ii} \\ &= C_{i1} + C_{i2} + \cdots + C_{iN} + C_{ii} - \sum_{j=1}^N C_{ij} = 0 \end{aligned}$$

And therefore

$$\boldsymbol{B_1} = \frac{1}{N}\underbrace{(C - diag(C1))}_{\boldsymbol{0}}(1_N1_N^T) = \boldsymbol{0}$$

And from this we can derive that

$$BJ = B$$

■

2.2

Show that:

$$g(\boldsymbol{Z}, \boldsymbol{Z}) = \|\boldsymbol{\Delta}_x - \boldsymbol{D}_z\|_F^2$$

Notes: (See lecture slides)

1. $\|\boldsymbol{\Delta}_x - \boldsymbol{D}_z\|_F^2 = \|\boldsymbol{\Delta}_x\|_F^2 + \|\boldsymbol{D}_z\|_F^2 - 2\langle \boldsymbol{\Delta}_x, \boldsymbol{D}_z\rangle$
2. $\|\boldsymbol{D}_z\|_F^2 = 2N\mathrm{Tr}\left\{\boldsymbol{ZJZ}^T\right\}$

Lets examine $\langle \boldsymbol{\Delta}_x, \boldsymbol{D}_z\rangle$:

$$\begin{aligned}
-\langle \boldsymbol{\Delta}_x, \boldsymbol{D}_z\rangle &= \left\langle \boldsymbol{C}, \boldsymbol{D}_z^{\circ 2}\right\rangle = \left\langle C, p\mathbf{1_N}^T - 2Z^TZ + \mathbf{1_N}p^T\right\rangle \\
&\text{Similarly to what we saw in lecture 9, } p = diag(Z^TZ) \\
&= \left\langle C, diag(Z^TZ)\mathbf{1_N}^T - 2Z^TZ + \mathbf{1_N}(diag^T(Z^TZ))\right\rangle \\
&\text{For symmetric matrix W it holds: } \langle W, Y\rangle = \left\langle W, Y^T\right\rangle \\
&= \left\langle C, 2diag(Z^TZ)\mathbf{1_N}^T - 2Z^TZ\right\rangle \\
&= 2\left\langle C, diag(Z^TZ)\mathbf{1_N}^T - Z^TZ\right\rangle \\
&= 2\left(\left\langle C, diag(Z^TZ)\mathbf{1_N}^T\right\rangle - \left\langle C, Z^TZ)\right\rangle\right) \\
&= 2\left(\left\langle C\mathbf{1_N}, diag(Z^TZ)\right\rangle - \left\langle C, Z^TZ)\right\rangle\right) \\
&\text{From question 1.8 in HW1 } \langle a, diag(X)\rangle = \langle diag(a), X\rangle \\
&= 2\left(\left\langle diag(C\mathbf{1_N}), Z^TZ\right\rangle - \left\langle C, Z^TZ\right\rangle\right) \\
&= -2\left\langle C - diag(C\mathbf{1_N}), Z^TZ\right\rangle = -2\left\langle B, Z^TZ\right\rangle
\end{aligned}$$

And now,

$$\begin{aligned}
\|\Delta_x - D_z\|_F^2 &= \|\Delta_x\|_F^2 + \|D_z\|_F^2 - 2\langle \Delta_x, D_z\rangle = \\
&= \|\Delta_x\|_F^2 + 2N\mathrm{Tr}\left\{ZJZ^T\right\} + 2\left\langle C, D_z^{\circ 2}\right\rangle \\
&= \|\Delta_x\|_F^2 + 2N\mathrm{Tr}\left\{ZJZ^T\right\} - 4\left(\left\langle B, Z^TZ\right\rangle\right) \\
&= g(Z, Z)
\end{aligned}$$

■

3 Isomap (and out of sample extension)

Consider the training set $\boldsymbol{X} \in \mathbb{R}^{D \times N_x}$ and the out of sample (test) set $\boldsymbol{Y} \in \mathbb{R}^{D \times N_y}$. Let $\widetilde{\boldsymbol{J}} \in \mathbb{R}^{N \times N}$ be the centering matrix using only the training data:

$$\widetilde{\boldsymbol{J}} = \boldsymbol{I}_N - \frac{1}{N_x} \begin{bmatrix} \boldsymbol{1}_{N_x} \\ \boldsymbol{0}_{N_y} \end{bmatrix} \boldsymbol{1}_N^T \in \mathbb{R}^{N \times N}$$

where:

- $N = N_x + N_y$
- $\begin{bmatrix} \boldsymbol{1}_{N_x} \\ \boldsymbol{0}_{N_y} \end{bmatrix} \in \mathbb{R}^N$ is the block concatenation of N_x ones and N_y zeros.

3.1

Show that:

$$\begin{bmatrix} \boldsymbol{X} & \boldsymbol{Y} \end{bmatrix} \widetilde{\boldsymbol{J}} = \begin{bmatrix} \widetilde{\boldsymbol{X}} & \widetilde{\boldsymbol{Y}} \end{bmatrix} \in \mathbb{R}^{D \times N}$$

where:

- $\begin{bmatrix} \boldsymbol{X} & \boldsymbol{Y} \end{bmatrix} \in \mathbb{R}^{D \times N}$ is the block concatenation of $\boldsymbol{X}$ and $\boldsymbol{Y}$.
- $\widetilde{\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{J} = \boldsymbol{X} - \boldsymbol{\mu}_x \boldsymbol{1}_{N_x}^T$
- $\widetilde{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{\mu}_x \boldsymbol{1}_{N_y}^T$

$\widetilde{\boldsymbol{X}}$ and $\widetilde{\boldsymbol{Y}}$ are the centered version of $\boldsymbol{X}$ and $\boldsymbol{Y}$ (when the mean is computed only using the $\boldsymbol{X}$).

Seen in class, but ok. We will notice that

$$\widetilde{\boldsymbol{J}} = \boldsymbol{I}_N - \frac{1}{N_x} \begin{bmatrix} \boldsymbol{1}_{N_x} \\ \boldsymbol{0}_{N_y} \end{bmatrix} \boldsymbol{1}_N^T = \begin{bmatrix} \boldsymbol{J}_{N_x} & -\frac{1}{N_x} \boldsymbol{1}_{N_x} \boldsymbol{1}_{N_y}^T \\ \boldsymbol{0}_{N_y} & \boldsymbol{I}_{N_y} \end{bmatrix}$$

And now we can see that:

$$\begin{aligned}
\begin{bmatrix} \boldsymbol{X} & \boldsymbol{Y} \end{bmatrix} \widetilde{\boldsymbol{J}} &= \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Y} \end{bmatrix} \begin{bmatrix} \boldsymbol{J}_{N_x} & -\frac{1}{N_x} \boldsymbol{1}_{N_x} \boldsymbol{1}_{N_y}^T \\ \boldsymbol{0}_{N_y} & \boldsymbol{I}_{N_y} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{X}\boldsymbol{J}_{N_x} & -\frac{1}{N_x} \boldsymbol{X} \boldsymbol{1}_{N_x} \boldsymbol{1}_{N_y}^T + \boldsymbol{Y} \boldsymbol{I}_{N_y} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{X}\boldsymbol{J}_{N_x} & \boldsymbol{Y} - \frac{1}{N_x} \boldsymbol{X} \boldsymbol{1}_{N_y}^T \end{bmatrix} \\
&= \begin{bmatrix} \widetilde{\boldsymbol{X}} & \widetilde{\boldsymbol{Y}} \end{bmatrix}
\end{aligned}$$

$\blacksquare$

3.2

Let

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_{xx} & \boldsymbol{D}_{xy} \\ \boldsymbol{D}_{xy}^T & \boldsymbol{A} \end{bmatrix} \in \mathbb{R}^{N \times N}$$

where:

- $\boldsymbol{D}_{xx} \in \mathbb{R}^{N_x \times N_x}, and \boldsymbol{D}_{xx}[i,j] = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2$.
- $\boldsymbol{D}_{xy} \in \mathbb{R}^{N_x \times N_y}, and \boldsymbol{D}_{xy}[i,j] = \left\|\boldsymbol{x}_i - \boldsymbol{y}_j\right\|_2^2$.
- $\boldsymbol{A} \in \mathbb{R}^{N_y \times N_y}$ is some matrix.

Show that:

$$-\frac{1}{2}\widetilde{\boldsymbol{J}}^T \boldsymbol{D} \widetilde{\boldsymbol{J}} = \begin{bmatrix} \widetilde{\boldsymbol{K}}_{xx} & \widetilde{\boldsymbol{K}}_{xy} \\ \widetilde{\boldsymbol{K}}_{xy}^T & \widetilde{\boldsymbol{A}} \end{bmatrix} \in \mathbb{R}^{N \times N}$$

for some matrix $\widetilde{\boldsymbol{A}}$ (you do not need to find it).
Hints: In the lectures, we saw that:

- $-\frac{1}{2}\boldsymbol{J}\boldsymbol{D}_{xx}\boldsymbol{J} = \widetilde{\boldsymbol{K}}_{xx} := \widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{X}}$
- $-\frac{1}{2}\boldsymbol{J}\left(\boldsymbol{D}_{xy} - \frac{1}{N_x}\boldsymbol{D}_{xx}\boldsymbol{1}_{N_x}\boldsymbol{1}_{N_y}^T\right) = \widetilde{\boldsymbol{K}}_{xy} := \widetilde{\boldsymbol{X}}^T \widetilde{\boldsymbol{Y}}$

Also seen in class, but I guess this is the point of the exercise. In the previous section we saw that

$$\widetilde{\boldsymbol{J}} = \begin{bmatrix} \boldsymbol{J}_{N_x} & -\frac{1}{N_x}\boldsymbol{1}_{N_x}\boldsymbol{1}_{N_y}^T \\ \boldsymbol{0}_{N_y} & \boldsymbol{I}_{N_y} \end{bmatrix}$$

So

$$\begin{aligned}
-\frac{1}{2}\widetilde{\boldsymbol{J}}^T \boldsymbol{D} \widetilde{\boldsymbol{J}} &= -\frac{1}{2}\begin{bmatrix} \boldsymbol{J}_{N_x} & -\frac{1}{N_x}\boldsymbol{1}_{N_x}\boldsymbol{1}_{N_y}^T \\ \boldsymbol{0}_{N_y} & \boldsymbol{I}_{N_y} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{D}_{xx} & \boldsymbol{D}_{xy} \\ \boldsymbol{D}_{xy}^T & \boldsymbol{A} \end{bmatrix} \begin{bmatrix} \boldsymbol{J}_{N_x} & -\frac{1}{N_x}\boldsymbol{1}_{N_x}\boldsymbol{1}_{N_y}^T \\ \boldsymbol{0}_{N_y} & \boldsymbol{I}_{N_y} \end{bmatrix} \\
&= -\frac{1}{2}\begin{bmatrix} \boldsymbol{J}_{N_x} & \boldsymbol{0}_{N_y} \\ -\frac{1}{N_x}\boldsymbol{1}_{N_x}\boldsymbol{1}_{N_y}^T & \boldsymbol{I}_{N_y} \end{bmatrix} \begin{bmatrix} \boldsymbol{D}_{xx} & \boldsymbol{D}_{xy} \\ \boldsymbol{D}_{xy}^T & \boldsymbol{A} \end{bmatrix} \begin{bmatrix} \boldsymbol{J}_{N_x} & -\frac{1}{N_x}\boldsymbol{1}_{N_x}\boldsymbol{1}_{N_y}^T \\ \boldsymbol{0}_{N_y} & \boldsymbol{I}_{N_y} \end{bmatrix} \\
&= -\frac{1}{2}\begin{bmatrix} \boldsymbol{J}_{N_x}\boldsymbol{D}_{xx} & \boldsymbol{J}_{N_x}\boldsymbol{D}_{xy} \\ -\frac{1}{N_x}\boldsymbol{D}_{xx}\boldsymbol{1}_{N_x}\boldsymbol{1}_{N_y}^T + \boldsymbol{I}_{N_y}\boldsymbol{D}_{xy}^T & \boldsymbol{\emptyset} \end{bmatrix} \begin{bmatrix} \boldsymbol{J}_{N_x} & -\frac{1}{N_x}\boldsymbol{1}_{N_x}\boldsymbol{1}_{N_y}^T \\ \boldsymbol{0}_{N_y} & \boldsymbol{I}_{N_y} \end{bmatrix} \\
&= -\frac{1}{2}\begin{bmatrix} \boldsymbol{J}_{N_x}\boldsymbol{D}_{xx}\boldsymbol{J}_{N_x} & \boldsymbol{J}_{N_x}\left(\boldsymbol{D}_{xy} - \frac{1}{N_x}\boldsymbol{D}_{xx}\boldsymbol{1}_{N_x}\boldsymbol{1}_{N_y}^T\right) \\ \left(\boldsymbol{D}_{xy} - \frac{1}{N_x}\boldsymbol{D}_{xx}\boldsymbol{1}_{N_x}\boldsymbol{1}_{N_y}^T\right)^T \boldsymbol{J}_{N_x} & -2\widetilde{\boldsymbol{A}} \end{bmatrix} \\
&= -\frac{1}{2}\begin{bmatrix} -2\widetilde{\boldsymbol{K}_{xx}} & -2\widetilde{\boldsymbol{K}_{xy}} \\ -2\widetilde{\boldsymbol{K}_{xy}}^T & -2\widetilde{\boldsymbol{A}} \end{bmatrix} \\
&= \begin{bmatrix} \widetilde{\boldsymbol{K}_{xx}} & \widetilde{\boldsymbol{K}_{xy}} \\ \widetilde{\boldsymbol{K}_{xy}}^T & \widetilde{\boldsymbol{A}} \end{bmatrix}
\end{aligned}$$

Where $\boldsymbol{\emptyset}, \widetilde{\boldsymbol{A}}$ are some matrices. ∎

4 Laplacian Eigenmaps

- Consider $\mathcal{X} = \left\{ \boldsymbol{x}_i \in \mathbb{R}^D \right\}_{i=1}^N$.
- Let $G = (V, E, W)$ be a weighted graph with $V = \mathcal{X}$ and:

$$\boldsymbol{W}[i,j] = \begin{cases} \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2}{2\sigma^2}\right) & \boldsymbol{x}_i \in \mathcal{N}_j \text{ or } \boldsymbol{x}_j \in \mathcal{N}_i \\ 0 & \text{else} \end{cases}$$

- $e_{ij} \in E$ if $\boldsymbol{W}[i,j] \neq 0$.
- Let $\boldsymbol{Z} \in \mathbb{R}^{d \times N}$ and $\boldsymbol{D}_z \in \mathbb{R}^{N \times N}$ such that $\boldsymbol{D}_z[i,j] = \|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2^2$ where $\boldsymbol{z}_i$ is the ith column of $\boldsymbol{Z}$.

4.1

Show that:

$$\frac{1}{2}\langle \boldsymbol{W}, \boldsymbol{D}_z \rangle = \mathrm{Tr}\left\{ \boldsymbol{Z}\boldsymbol{L}\boldsymbol{Z}^T \right\}$$

where:

- $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$ is the graph-Laplacian.
- $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{W}\mathbf{1})$ is the degree matrix.

Lets examine $\mathrm{Tr}\left\{ZLZ^T\right\}$:

$$\mathrm{Tr}\left\{ZLZ^T\right\} = \sum_{i=1}^{N} \left(ZLZ^T\right)_{ii} = \sum_{i=1}^{N} z_i L z_i^T = (*)$$

We will notice that z_i is a vector in $\mathbb{R}^N$ and as saw in the lecture, for $v \in \mathbb{R}^N, v^T L v = \frac{1}{2}\langle W, D_v \rangle$. Therefore:

$$(*) = \frac{1}{2}\sum_{i=1}^{N} \langle W, D_{z_i} \rangle = \frac{1}{2}\langle W, D_z \rangle$$

$\blacksquare$

4.2

Assume that G has two connected components, i.e. $V = V_1 \cup V_2$ such that:

$$\left\{ e_{ij} \middle| i \in V_1, j \in V_2 \right\} = \emptyset$$

Show that the graph-Laplacian $\boldsymbol{L}$ has two orthogonal eigenvectors corresponding to the zero eigenvalue. That is, exist $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathbb{R}^N$ such that:

1. $\boldsymbol{L}\boldsymbol{u}_1 = \boldsymbol{L}\boldsymbol{u}_2 = \mathbf{0}$
2. $\langle \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle = 0$.

According to the hint provided (on Whatsapp or in the lecture), we can assume without loss of generality, that the first k points of W belongs to V_1 and the last $N-k$ points belongs to V_2. (Following what we did in the lecture) We can express the weights between the vertices as 4 groups:

- $W1$ will be all the weights from $v_i \in V_1$ to $v_j \in V_1$
- $W2$ will be all the weights from $v_i \in V_2$ to $v_j \in V_2$
- $W3$ will be all the weights from $v_i \in V_1$ to $v_j \in V_2$
- $W4$ will be all the weights from $v_i \in V_2$ to $v_j \in V_1$

Because there are no edges between V_1 and V_2, we get $W_3 = \mathbf{0}_{\boldsymbol{k\times(N-k)}}$ and $W_4 = \mathbf{0}_{\boldsymbol{(N-k)\times k}}$ and we can express W as a block matrix:

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{W_1} & 0 \\ 0 & \boldsymbol{W_2} \end{bmatrix}$$

Therefore, D can be expressed as a block matrix where (As mentioned in class):

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D_1} & 0 \\ 0 & \boldsymbol{D_2} \end{bmatrix}$$

And finally, we can express $\boldsymbol{L}$ as the following diagonal block matrix:

$$\boldsymbol{L} = \begin{bmatrix} \boldsymbol{L_1} & 0 \\ 0 & \boldsymbol{L_2} \end{bmatrix}$$

where L_1 corresponds to V_1 and L_2 corresponds to V_2.
Lets denote $u_1^T = (\underbrace{1, \cdots, 1}_{k}, \underbrace{0, \cdots, 0}_{N-k})$ and $u_2^T = (\underbrace{0, \cdots, 0}_{k}, \underbrace{1, \cdots, 1}_{N-k})$ two vectors in $\mathbb{R}^N$.
Now, lets examine $\boldsymbol{Lu_1}$:

$$\begin{aligned} \boldsymbol{Lu_1} &= Du_1 - Wu_1 = diag(W)u_1 - Wu_1 \\ &= diag(W_1)1_k + diag(W_2)0_{N-k} - W_1 1_k + W_2 0_{N-k} \\ &= (diag(W_1) + W_1)1_k = 0 \end{aligned}$$

Which means that u_1 is an eigenvetor of L. Similarly u_2 is an eigenvetor of L.
The fact that $\langle u_1, u_2 \rangle = 0$ follows straight from the definition of u_1, u_2:

$$\langle u_1, u_2 \rangle = \sum_{i=1}^{N} u_{1i} * u_{2i} = \sum_{i=1}^{k} (1 * 0) + \sum_{i=k+1}^{N} (0 * 1) = 0$$

. ■