Unsupervised Learning Methods
Problem Set III –
MDS, Isomap and Laplacian-Eigenmaps

Due: 18.06.2021

Guidelines

- Answer all questions (PDF + Jupyter notebook).
- You must type your solution manual (handwriting is not allowed).
- Submission in pairs (use the forum if needed).
- You **may** submit the entire solution in a single ipynb file (or in PDF + ipynb files).
- You **may** (and should) use the forum if you have any questions.
- Good luck!

1 Classical MDS

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

1.1

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

$$\|\boldsymbol{R}\boldsymbol{x}_i - \boldsymbol{R}\boldsymbol{x}_j\|_2 = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$$

1.2 Implementation and applications

Solve this section in the attached notebook.

---

2 Metric MDS

The metric MDS objective is given by:

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

where:

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

Consider the surrogate function:

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

where:

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

2.1

Prove that:

$$\boldsymbol{B}\boldsymbol{J} = \boldsymbol{B}$$

2.2

Show that:

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

Notes: (See lecture slides)

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

Hint:
For $\widetilde{\boldsymbol{Z}} = \boldsymbol{Z}$ we have:

$$\left\langle\boldsymbol{\Delta}_x, \boldsymbol{D}_z\right\rangle = -\left\langle\boldsymbol{C}, \boldsymbol{D}_z^{\circ 2}\right\rangle$$

where $\boldsymbol{D}_z^{\circ 2}[i,j] = \boldsymbol{p}\mathbf{1}^T - 2\boldsymbol{Z}^T\boldsymbol{Z} + \mathbf{1}\boldsymbol{p}^T$ and $\boldsymbol{p} = \begin{bmatrix} \left\|\boldsymbol{z}_1\right\|_2^2 \\ \vdots \\ \left\|\boldsymbol{z}_N\right\|_2^2 \end{bmatrix}$.

2.3 Implementation and applications

Solve this section in the attached notebook.

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:

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

where:

- $N = N_x + N_y$
- $\begin{bmatrix} \mathbf{1}_{N_x} \\ \mathbf{0}_{N_y} \end{bmatrix} \in \mathbb{R}^N$ 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 $\boldsymbol{Y}$.
- $\widetilde{\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{J} = \boldsymbol{X} - \boldsymbol{\mu}_x \mathbf{1}_{N_x}^T$
- $\widetilde{\boldsymbol{Y}} = \boldsymbol{X} - \boldsymbol{\mu}_x \mathbf{1}_{N_y}^T$

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

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:

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

3.3 Implementation and applications

Solve this section in the attached notebook.

---

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.

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

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

4.2

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. $\langle \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle = 0$

4.3 Implementation and applications

Solve this section in the attached notebook.