Unsupervised Learning Methods
Problem Set IV –
MDS, Isomap, Laplacian-Eigenmaps, and T-SNE

Due: 28.06.2022

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 MDS

Classical MDS
Consider the following inner product:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_\phi := \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{y}) \rangle$$

for some suitable $\phi : \mathbb{R}^D \to \mathbb{R}^M$, where $\langle \cdot, \cdot \rangle$ is the standard Euclidean inner product, i.e. $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \boldsymbol{a}^T \boldsymbol{b}$. Consider:

1. The induced norm:
$$\|\boldsymbol{x}\|_\phi := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle_\phi}$$

2. The induced metric:
$$d_\phi(\boldsymbol{x}, \boldsymbol{y}) := \|\boldsymbol{x} - \boldsymbol{y}\|_\phi$$

Consider a training set $\{\boldsymbol{x}_i\}_{i=1}^N$ and let $\boldsymbol{D}_\phi \in \mathbb{R}^{N \times N}$ where $\boldsymbol{D}_\phi[i,j] = d_\phi^2(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

1.1

Show that
$$-\frac{1}{2}\boldsymbol{J}\boldsymbol{D}_\phi\boldsymbol{J} = \boldsymbol{J}\boldsymbol{K}_\phi\boldsymbol{J}$$

where:

1. $\boldsymbol{J} = \boldsymbol{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^T \in \mathbb{R}^{N \times N}$

2. $\boldsymbol{K}_\phi := \boldsymbol{\Phi}^T\boldsymbol{\Phi}$, $\boldsymbol{\Phi}$ is given by:
$$\boldsymbol{\Phi} = \begin{bmatrix} | & | & & | \\ \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_N \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{M \times N}$$

   and $\boldsymbol{\phi}_i = \phi(\boldsymbol{x}_i)$.

Steps:

1. Show that ϕ must be linear, namely:
$$\phi(\alpha\boldsymbol{x} + \beta\boldsymbol{y}) = \alpha\phi(\boldsymbol{x}) + \beta\phi(\boldsymbol{y})$$

   Hint: Consider $\langle \alpha\boldsymbol{x} + \beta\boldsymbol{y}, \boldsymbol{z} \rangle_\phi$ and recall that this is true for all $\boldsymbol{z}$.

2. Show that:
$$\begin{aligned} d_\phi^2(\boldsymbol{x}, \boldsymbol{y}) &= \|\phi(\boldsymbol{x}) - \phi(\boldsymbol{y})\|_2^2 \\ &= \|\phi(\boldsymbol{x})\|_2^2 - 2\langle \phi(\boldsymbol{x}), \phi(\boldsymbol{y}) \rangle + \|\phi(\boldsymbol{y})\|_2^2 \end{aligned}$$

3. Repeat\use the lecture notes to conclude that $-\frac{1}{2}\boldsymbol{J}\boldsymbol{D}_\phi\boldsymbol{J} = \boldsymbol{J}\boldsymbol{K}_\phi\boldsymbol{J}$.

Solution:
Type your solution here...

Consider a training set $\{\boldsymbol{x}_i\}_{i=1}^{N}$ and let $\boldsymbol{D} \in \mathbb{R}^{N \times N}$ where $\boldsymbol{D}[i,j] = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2$.

1.2

Show that $\boldsymbol{v}^T \boldsymbol{D} \boldsymbol{v} \leq 0$ for any $\boldsymbol{v}$ such that $\langle \boldsymbol{v}, \boldsymbol{1} \rangle = 0$.

Solution:
Type your solution here...

MM (Majorization Minimization\Maximization)
Consider:

$$\boldsymbol{Y}[i,j] = \begin{cases} \boldsymbol{X}[i,j] & \boldsymbol{M}[i,j] = 1 \\ 0 & \boldsymbol{M}[i,j] = 0 \end{cases}$$

In other words:

$$\boldsymbol{Y} = \boldsymbol{M} \odot \boldsymbol{X}$$

where $\boldsymbol{M} \in \{0,1\}^{M \times N}$ is a binary mask matrix.
Given $\boldsymbol{Y} \in \mathbb{R}^{M \times N}$, the low-rank matrix completion objective is given by:

$$\begin{cases} \min_{\boldsymbol{X}} \|\boldsymbol{M} \odot (\boldsymbol{Y} - \boldsymbol{X})\|_F^2 \\ \text{s.t.} \\ \text{rank}(\boldsymbol{X}) \leq d \end{cases}$$

Consider the following function:

$$g(\boldsymbol{X}, \boldsymbol{Z}) := \|\boldsymbol{X} - \boldsymbol{Z} + \boldsymbol{M} \odot (\boldsymbol{Z} - \boldsymbol{Y})\|_F^2$$

1.3

Show that g surrogates the objective $f(\boldsymbol{X}) := \|\boldsymbol{M} \odot (\boldsymbol{Y} - \boldsymbol{X})\|_F^2$.
Hint:
Show that $g(\boldsymbol{X}, \boldsymbol{Z}) = \left\|\boldsymbol{M} \odot (\boldsymbol{X} - \boldsymbol{Y}) + \widetilde{\boldsymbol{M}} \odot (\boldsymbol{X} - \boldsymbol{Z})\right\|_F^2$ where $\widetilde{\boldsymbol{M}} := \boldsymbol{1}\boldsymbol{1}^T - \boldsymbol{M}$ is the complement of $\boldsymbol{M}$.

Solution:
Type your solution here...

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 distance matrix.
- $\boldsymbol{D}_z[i,j] = \|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2$.

Consider the surrogate function:

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

where:

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

1.4

Show that:

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

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. $\|\boldsymbol{D}_z\|_F^2 = 2N\mathrm{Tr}\left\{\boldsymbol{Z}\boldsymbol{J}\boldsymbol{Z}^T\right\}$

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

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

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 $\boldsymbol{p} = \begin{bmatrix} \|\boldsymbol{z}_1\|_2^2 \\ \vdots \\ \|\boldsymbol{z}_N\|_2^2 \end{bmatrix}$.

Solution:
Type your solution here...

1.5 Implementation and applications

Solve this section in the attached notebook.

---

2 Isomap

Let $G = (V, E, W)$ be a simple, undirected, and weighted graph, and assume no negative weights\edges. Let $\boldsymbol{D} \in \mathbb{R}^{N \times N}$ be the shortest path distance matrix, where $N = |V|$.

2.1

Prove or disprove:
Necessarily exists an embedding $\left\{\boldsymbol{z}_i \in \mathbb{R}^d\right\}_{i=1}^N$ (for some $d \in \mathbb{N}$) such that (for all i, j):

$$\boldsymbol{D}[i, j] = \|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2$$

Solution: Type your solution here...

- Let $\mathcal{X} = \{\boldsymbol{x}_i\}_{i=1}^N$ be the training set.
- Let $\mathcal{Z} = \{\boldsymbol{z}_i\}_{i=1}^N$ be the representation obtained by Isomap (training encoding).
- Consider a new point $\boldsymbol{x}^\star$ where $\boldsymbol{x}^\star = \boldsymbol{x}_k$ for some $k \leq N$.
- Let $\boldsymbol{z}^\star$ be the out of sample encoding applied to $\boldsymbol{x}^\star$.

2.2

Prove of disprove:

$$\boldsymbol{z}^\star = \boldsymbol{z}_k$$

Solution: Type your solution here...

2.3 Implementation and applications

Solve this section in the attached notebook.

---

3 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:

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

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

3.1

Show that:

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

where:

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

Solution:
Type your solution here...

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

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

3.2

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

Solution:
Type your solution here...

3.3 Implementation and applications

Solve this section in the attached notebook.

4 t-SNE

The t-SNE objective is given by:

$$\min_{\boldsymbol{Z} \in \mathbb{R}^{d \times N}} \underbrace{D_{\mathrm{KL}} \left(\boldsymbol{P} || \boldsymbol{Q} \right)}_{:= f(\boldsymbol{Z})} = \min_{\boldsymbol{Z} \in \mathbb{R}^{d \times N}} \sum_{i=1}^{N} \sum_{j=1}^{N} p_{ij} \log \left(\frac{p_{ij}}{q_{ij}} \right)$$

See the definitions of $\boldsymbol{P}$ and $\boldsymbol{Q}$ in the lecture notes (do not get confused by the SNE definitions). The goal of this question is to compute the gradient of the objective:

$$\nabla f \left(\boldsymbol{Z} \right) = ?$$

Let us break this task into several **smaller** steps.

4.1

Show that $f(\boldsymbol{Z}) = D_{\text{KL}}(\boldsymbol{P}||\boldsymbol{Q})$ can be written as:

$$f(\boldsymbol{Z}) = C - \langle \boldsymbol{P}, \log[\boldsymbol{Q}] \rangle$$

where C is some constant (the entropy of $\boldsymbol{P}$).

Solution:
Type your solution here...

- Reminder:

$$\boldsymbol{Q}[i,j] = \frac{1}{B} \begin{cases} 0 & i = j \\ \left(1 + \|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2^2\right)^{-1} & i \neq j \end{cases}$$

- Let $\boldsymbol{D}_z \in \mathbb{R}^{N \times N}$ such that $\boldsymbol{D}_z[i,j] = \|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2^2$.

- Let $\boldsymbol{S} = \left(\boldsymbol{1}\boldsymbol{1}^T + \boldsymbol{D}_z\right)^{\circ -1} \in \mathbb{R}^{N \times N}$, that is:

$$\boldsymbol{S}[i,j] = (1 + \boldsymbol{D}_z[i,j])^{-1}$$

Solution:
Type your solution here...

4.2

Show that:

1. $$B = \boldsymbol{1}^T (\boldsymbol{S} - \boldsymbol{I}) \boldsymbol{1} \in \mathbb{R}$$

2. $$\boldsymbol{Q} = B^{-1} (\boldsymbol{S} - \boldsymbol{I}) \in \mathbb{R}^{N \times N}$$

Solution:
Type your solution here...

4.3

Show that:

$$-\left\langle \boldsymbol{P}, \log\left[\boldsymbol{Q}\right]\right\rangle = \log\left(B\right) + \left\langle \boldsymbol{P}, \log\left[\mathbf{1}\mathbf{1}^T + \boldsymbol{D}_z\right]\right\rangle$$

Hints:

- $\boldsymbol{P}\left[i,i\right] =?$
- $\mathbf{1}^T\boldsymbol{P}\mathbf{1} =?$

Solution:
Type your solution here...

Let:

$$f\left(\boldsymbol{Z}\right) = C + \underbrace{\log\left(B\right)}_{(*)} + \underbrace{\left\langle \boldsymbol{P}, \log\left[\mathbf{1}\mathbf{1}^T + \boldsymbol{D}_z\right]\right\rangle}_{(**)}$$

4.4

Show that:

1. $$\nabla_{\boldsymbol{Z}} \underbrace{\left\langle \boldsymbol{P}, \log\left[\mathbf{1}\mathbf{1}^T + \boldsymbol{D}_z\right]\right\rangle}_{(**)} \left[\boldsymbol{H}\right] = \left\langle \boldsymbol{S} \circ \boldsymbol{P}, \nabla \boldsymbol{D}_z\left[\boldsymbol{H}\right]\right\rangle$$

2. $$\nabla_{\boldsymbol{Z}} \underbrace{\log\left(B\right)}_{(*)} \left[\boldsymbol{H}\right] = -\left\langle \boldsymbol{S} \circ \boldsymbol{Q}, \nabla \boldsymbol{D}_z\left[\boldsymbol{H}\right]\right\rangle$$

   Hints:

- $\nabla \boldsymbol{S}\left[\boldsymbol{H}\right] = \nabla\left(\mathbf{1}\mathbf{1}^T + \boldsymbol{D}_z\right)^{\circ -1}\left[\boldsymbol{H}\right] = -\left(\mathbf{1}\mathbf{1}^T + \boldsymbol{D}_z\right)^{\circ -2} \circ \nabla\left(\boldsymbol{D}_z\right)\left[\boldsymbol{H}\right] = -\boldsymbol{S} \circ \boldsymbol{S} \circ \nabla\left(\boldsymbol{D}_z\right)\left[\boldsymbol{H}\right]$
- $\boldsymbol{Q} = B^{-1}\left(\boldsymbol{S} - \boldsymbol{I}\right)$

Solution:
Type your solution here...

4.5

- Combine all previous results and write the gradient of the objective:

$$\nabla f(\boldsymbol{Z}) = ?$$

- Use $\boldsymbol{A} := (\boldsymbol{P} - \boldsymbol{Q}) \circ \boldsymbol{S}$ to simplify your answer.
- What can you say about the gradient $\nabla f(\boldsymbol{Z})$ when $\boldsymbol{P} = \boldsymbol{Q}$?

Hint: Use the lecture notes.

Solution:
Type your solution here...

4.6 Implementation and applications

⌨ Solve this section in the attached notebook. ⌨