Unsupervised Learning Methods
Problem Set IV
t-SNE and (Geometric) Domain Adaptation

Due: 15.07.2021

Guidelines

- Answer all questions (PDF + **two** Jupyter notebooks).
- 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 ipynb file (or in PDF + ipynb files).
- You **may** (and should) use the forum if you have any questions.
- Good luck!

1 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.

1.1

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

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

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

- Reminder:

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

- Let $\boldsymbol{D}_z \in \mathbb{R}^{N\times N}$ such that $\boldsymbol{D}_z\left[i,j\right] = \left\|\boldsymbol{z}_i-\boldsymbol{z}_j\right\|_2^2$.
- Let $\boldsymbol{S} = \left(\boldsymbol{1}\boldsymbol{1}^T + \boldsymbol{D}_z\right)^{\circ -1} \in \mathbb{R}^{N\times N}$, that is:

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

1.2

Show that:

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

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

1.3

Show that:

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

Hints:

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

Let:

- $g_1(\boldsymbol{Z}) = \left\langle \boldsymbol{P}, \log\left[\mathbf{1}\mathbf{1}^T + \boldsymbol{D}_z\right]\right\rangle$
- $g_2(\boldsymbol{Z}) = \log(B)$

The following two exercises will be useful when computing the gradients of g_1 and g_2.

1.4

- Consider some symmetric matrix $\boldsymbol{M} = \boldsymbol{M}^T \in \mathbb{R}^{N\times N}$.
- Let $g_3(\boldsymbol{Z}) = \langle \boldsymbol{M}, \boldsymbol{D}_z\rangle$.

1.4.1

- Assume that $\boldsymbol{M}$ is a diagonal matrix (only in this section).
- Simplify $g_3(\boldsymbol{Z})$ and compute the gradient $\nabla g_3(\boldsymbol{Z})$.
  (No computations are required for this)

1.4.2

- Express the directional derivative as $\nabla g_3(\boldsymbol{Z})[\boldsymbol{H}] = \langle \boldsymbol{G}, \boldsymbol{H}\rangle$,
  where $\boldsymbol{G} \in \mathbb{R}^{N\times N}$ is a function of $\boldsymbol{M} \in \mathbb{R}^{N\times N}$ and $\boldsymbol{Z} \in \mathbb{R}^{d\times N}$.
- Make sure you result coincide with the diagonal case.

Hint: $\boldsymbol{D}_z = \operatorname{diag}\left(\boldsymbol{Z}^T\boldsymbol{Z}\right)\mathbf{1}^T - 2\boldsymbol{Z}^T\boldsymbol{Z} + \mathbf{1}\operatorname{diag}^T\left(\boldsymbol{Z}^T\boldsymbol{Z}\right)$.

1.5

Compute the gradient of $g_1(\boldsymbol{Z}) := \left\langle \boldsymbol{P}, \log\left[\mathbf{1}\mathbf{1}^T + \boldsymbol{D}_z\right]\right\rangle$

$$\nabla g_1(\boldsymbol{Z}) = ?$$

Hints:

- $\langle \boldsymbol{P}, \boldsymbol{S} \circ \nabla(\boldsymbol{D}_z)[\boldsymbol{H}]\rangle = \langle \boldsymbol{P} \circ \boldsymbol{S}, \nabla(\boldsymbol{D}_z)[\boldsymbol{H}]\rangle$
- $\boldsymbol{P}$ and $\boldsymbol{S}$ are symmetric (use 1.4.2).

1.6

Compute the gradient of $g_2(\boldsymbol{Z}) = \log(B)$

$$\nabla g_2(\boldsymbol{Z}) = ?$$

Hints:

- $\nabla \boldsymbol{S}[\boldsymbol{H}] = \nabla\left(\mathbf{1}\mathbf{1}^T + \boldsymbol{D}_z\right)^{\circ -1}[\boldsymbol{H}] = -\left(\mathbf{1}\mathbf{1}^T + \boldsymbol{D}_z\right)^{\circ -2} \circ \nabla(\boldsymbol{D}_z)[\boldsymbol{H}] = -\boldsymbol{S} \circ \boldsymbol{S} \circ \nabla(\boldsymbol{D}_z)[\boldsymbol{H}]$
- $\boldsymbol{Q} = B^{-1}(\boldsymbol{S} - \boldsymbol{I})$
- Use 1.4.1 and 1.4.2.

1.7

- 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 $\boldsymbol{P} = \boldsymbol{Q}$?

1.8 Implementation and applications

Solve this section in the attached notebook.

---

2 Geometric Domain Adaptation

- Let $\mathcal{M} \subset \mathbb{R}^3$ be the sphere manifold:

$$\mathcal{M} = \left\{ \boldsymbol{x}_i \in \mathbb{R}^3 \,\middle|\, \|\boldsymbol{x}_i\|_2 = 1 \right\}$$

- The geodesic $\boldsymbol{\gamma}_{\boldsymbol{x}\to\boldsymbol{y}}(t)$ is given by (assuming $\boldsymbol{x}$ and $\boldsymbol{y}$ are not antipodal):

$$\boldsymbol{\gamma}_{\boldsymbol{x}\to\boldsymbol{y}}(t) = \cos(ta)\,\boldsymbol{x} + \sin(ta)\,\boldsymbol{p}, \qquad t \in [0,1]$$

  where:

  - $a = \arccos\left(\boldsymbol{x}^T\boldsymbol{y}\right) \in \mathbb{R}$
  - $\boldsymbol{p} = \frac{\left(\boldsymbol{I} - \boldsymbol{x}\boldsymbol{x}^T\right)\boldsymbol{y}}{\|(\boldsymbol{I} - \boldsymbol{x}\boldsymbol{x}^T)\boldsymbol{y}\|_2} \in \mathbb{R}^3$

2.1

Show that:

1. $\boldsymbol{\gamma}(0) = \boldsymbol{x}$
2. $\boldsymbol{\gamma}(1) = \boldsymbol{y}$

Hints:

- $\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = ?$
- $\sin(\arccos(\alpha)) = ?$
- $\left\|\left(\boldsymbol{I} - \boldsymbol{x}\boldsymbol{x}^T\right)\boldsymbol{y}\right\|_2 = \sqrt{1 - ?}$

- Consider the symmetric matrix

$$\boldsymbol{X} = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

2.2

Prove that:

$$\boldsymbol{X} \succ 0 \iff \begin{cases} x > 0 \\ z > 0 \\ xz > y^2 \end{cases}$$

Hints:

- Show both directions at once.
- $\mathrm{Tr}\{\boldsymbol{X}\} = \lambda_1 + ?$
- $\det(\boldsymbol{X}) = \lambda_1 \cdot ?$

The (affine invariant) geodesic (squared) distance between two SPD matrices is given by:

$$d^2\left(\boldsymbol{X},\boldsymbol{Y}\right)=\left\|\log\left(\boldsymbol{X}^{-\frac{1}{2}}\boldsymbol{Y}\boldsymbol{X}^{-\frac{1}{2}}\right)\right\|_F^2$$

2.3

1. Let $\boldsymbol{A}=\boldsymbol{A}^T\in\mathbb{R}^{d\times d}$, show that:

$$\|\boldsymbol{A}\|_F^2=\sum_{i=1}^{d}\lambda_i^2\left(\boldsymbol{A}\right)$$

2. Assume $\boldsymbol{A}\succ 0$ and show that:

$$\|\log\left(\boldsymbol{A}\right)\|_F^2=\sum_{i=1}^{d}\log^2\left(\lambda_i\left(\boldsymbol{A}\right)\right)$$

   (**Hint:** What can you say about the eigenvalues of $\log\left(\boldsymbol{A}\right)$?)

3. Show that $\boldsymbol{X}^{-\frac{1}{2}}\boldsymbol{Y}\boldsymbol{X}^{-\frac{1}{2}}$ and $\boldsymbol{X}^{-1}\boldsymbol{Y}$ have the same set of eigenvalues.
   (**Hint:** HW2)

4. Conclude that (explain both equality):

$$d^2\left(\boldsymbol{X},\boldsymbol{Y}\right)=\sum_{i=1}^{d}\log^2\left(\lambda_i\left(\boldsymbol{X}^{-\frac{1}{2}}\boldsymbol{Y}\boldsymbol{X}^{-\frac{1}{2}}\right)\right)=\sum_{i=1}^{d}\log^2\left(\lambda_i\left(\boldsymbol{X}^{-1}\boldsymbol{Y}\right)\right)$$

Remark:
Instead of finding the eigenvalues of $\boldsymbol{X}^{-1}\boldsymbol{Y}$, that is:

$$\boldsymbol{X}^{-1}\boldsymbol{Y}\boldsymbol{u}=\lambda\boldsymbol{u}$$

We solve the (equivalent) task:

$$\boldsymbol{Y}\boldsymbol{u}=\lambda\boldsymbol{X}\boldsymbol{u}$$

This is known as the generalized eigen problem (with $\boldsymbol{Y}$ and $\boldsymbol{X}$),
and we do not need to invert any matrix.

2.4

Prove that:

1. For any invertible $\boldsymbol{C}\in\mathrm{GL}$:

$$d\left(\boldsymbol{C}\boldsymbol{X}\boldsymbol{C}^T,\boldsymbol{C}\boldsymbol{Y}\boldsymbol{C}^T\right)=d\left(\boldsymbol{X},\boldsymbol{Y}\right)$$

   (**Hint:** use similarity between matrices)

2.

$$d\left(\boldsymbol{X},\boldsymbol{Y}\right)=d\left(\boldsymbol{X}^{-1},\boldsymbol{Y}^{-1}\right)$$

- Let $\boldsymbol{S}_1, \boldsymbol{S}_2 \in \mathcal{T}_{\boldsymbol{X}}\mathcal{P}$
- Let $\mathrm{P}_{\boldsymbol{X}\to\boldsymbol{Y}}\left(\boldsymbol{S}\right) = \boldsymbol{E}\boldsymbol{S}\boldsymbol{E}^T$ be a parallel transport where $\boldsymbol{E} = \boldsymbol{X}^{\frac{1}{2}}\left(\boldsymbol{X}^{-\frac{1}{2}}\boldsymbol{Y}\boldsymbol{X}^{-\frac{1}{2}}\right)^{\frac{1}{2}}\boldsymbol{X}^{-\frac{1}{2}}$.

2.5

Show that:

$$\langle \boldsymbol{S}_1, \boldsymbol{S}_2\rangle_{\boldsymbol{X}} = \langle \mathrm{P}_{\boldsymbol{X}\to\boldsymbol{Y}}\left(\boldsymbol{S}_1\right), \mathrm{P}_{\boldsymbol{X}\to\boldsymbol{Y}}\left(\boldsymbol{S}_2\right)\rangle_{\boldsymbol{Y}}$$

Hint: Show that $\boldsymbol{E}^T\boldsymbol{Y}^{-1}\boldsymbol{E} = \boldsymbol{X}^{-1}$

2.6 Implementation and applications

Solve this section in the attached notebook.