Unsupervised Learning Methods 2022
Problem Set I –
Optimization

Due: 28.03.2022

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 PDF + ipynb files).
- You **may** (and should) use the forums if you have any questions.
- Good luck!

1 Convexity

Convex set

Let:

$$\mathbb{R}_{\geq 0}^{d}=\left\{\boldsymbol{x}\in\mathbb{R}^{d}\;\middle|\;\min_{i}x_{i}\geq 0\right\}$$

where $\boldsymbol{x}=\begin{bmatrix}x_{1}\\x_{2}\\\vdots\\x_{d}\end{bmatrix}$.

1.1

Prove or disprove: $\mathbb{R}_{\geq 0}^{d}$ is convex.

Solution:
Type your solution here...

Convex combination

Let $\mathcal{C} \subseteq \mathbb{R}^d$ be a convex set and consider $\{\boldsymbol{x}_i \in \mathcal{C}\}_{i=1}^N$.

1.2

Prove that for any $N \in \mathbb{N}$:

$$\sum_{i=1}^{N} \alpha_i \boldsymbol{x}_i \in \mathcal{C}$$

where α_i are such that:

- $\alpha_i \geq 0$ for all i.
- $\sum_{i=1}^{N} \alpha_i = 1$.

Solution:
Type your solution here...

Let $\mathcal{C} \subset \mathbb{R}^2$ be convex, and consider $\{\boldsymbol{x}_i \in \mathcal{C}\}_{i=1}^{10}$ such that $\boldsymbol{x}_i \neq \boldsymbol{x}_j$ for all $i \neq j$.

1.3

Prove or disprove: Necessarily, any point $\boldsymbol{y} \in \mathcal{C}$ can be represented as a convex combination of $\{\boldsymbol{x}_i\}_{i=1}^{10}$.

Solution:
Type your solution here...

---

2 The Gradient

Note: Assume all functions are differentiable.

Directional derivative

Let $f : \mathbb{R}^d \to \mathbb{R}$ and let $\boldsymbol{x}_0 \in \mathbb{R}^d$.

2.1

Prove that:

$$\forall \boldsymbol{h} \in \mathbb{R}^d : \nabla f(\boldsymbol{x}_0)[\boldsymbol{h}] = \langle \boldsymbol{g}_0, \boldsymbol{h} \rangle \implies \boldsymbol{g}_0 = \nabla f(\boldsymbol{x}_0)$$

Solution:
Type your solution here...

Definition
$f : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ is said to be <u>linear</u> if:

$$f(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = \alpha f(\boldsymbol{x}) + \beta f(\boldsymbol{y})$$

for all $\alpha, \beta \in \mathbb{R}$ and for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{d_1}$.

Let $f : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ be a linear function.

2.2

Prove that:

$$\nabla f(\boldsymbol{x})[\boldsymbol{h}] = f(\boldsymbol{h})$$

for all $\boldsymbol{x}, \boldsymbol{h} \in \mathbb{R}^{d_1}$.

Solution:
Type your solution here...

2.3 Some useful exercises

Compute the directional derivative $\nabla f(\boldsymbol{x})[\boldsymbol{h}]$ and the gradient $\nabla f(\boldsymbol{x})$ of:

1. $$f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}$$

   Solution:
   Type your solution here...

2.
$$f\left(\boldsymbol{X}\right)=\operatorname{Tr}\left\{\boldsymbol{X}^{T}\boldsymbol{A}\boldsymbol{X}\right\}$$

where $\boldsymbol{X}\in\mathbb{R}^{N\times d}$ and $\operatorname{Tr}\left\{\cdot\right\}$ is the trace operator.

Solution:
Type your solution here...

3.
$$f\left(\boldsymbol{x}\right)=\left\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\right\|_2^2$$

Solution:
Type your solution here...

4.
$$f\left(\boldsymbol{X}\right)=\left\|\boldsymbol{Y}-\boldsymbol{A}\boldsymbol{X}\right\|_F^2$$

where:

(a) $\boldsymbol{Y}\in\mathbb{R}^{D\times N}$, $\boldsymbol{A}\in\mathbb{R}^{D\times d}$ and $\boldsymbol{X}\in\mathbb{R}^{d\times N}$.

(b) $\left\|\cdot\right\|_F^2$ is the Frobenius norm, that is, $\left\|\boldsymbol{X}\right\|_F^2=\left\langle\boldsymbol{X},\boldsymbol{X}\right\rangle=\operatorname{Tr}\left\{\boldsymbol{X}^{T}\boldsymbol{X}\right\}$.

Solution:
Type your solution here...

5.
$$f\left(\boldsymbol{X}\right)=\left\langle\boldsymbol{X}^{T}\boldsymbol{A},\boldsymbol{Y}^{T}\right\rangle$$

where $\boldsymbol{Y}\in\mathbb{R}^{D\times N}$, $\boldsymbol{A}\in\mathbb{R}^{d\times D}$ and $\boldsymbol{X}\in\mathbb{R}^{d\times N}$.

Solution:
Type your solution here...

6.
$$f\left(\boldsymbol{x}\right)=\boldsymbol{a}^{T}g\left(\boldsymbol{x}\right)$$

where:

(a) $g\left(\boldsymbol{x}\right):=\begin{bmatrix}g\left(x_1\right)\\ \vdots\\ g\left(x_d\right)\end{bmatrix}\in\mathbb{R}^{d}$

Solution:
Type your solution here...

7.
$$f\left(\boldsymbol{X}\right)=\left\langle \boldsymbol{A},\log\left[\boldsymbol{X}\right]\right\rangle$$

where:

(a) $\boldsymbol{X}\in\mathbb{R}^{d\times d}$

(b) $\log\left[\boldsymbol{X}\right]$ is an element-wise log, that is:

$$\boldsymbol{M}=\log\left[\boldsymbol{X}\right]\implies\boldsymbol{M}\left[i,j\right]=\log\left(\boldsymbol{X}\left[i,j\right]\right)$$

Solution:
Type your solution here...

8.
$$f\left(\boldsymbol{X}\right)=\left\langle \boldsymbol{a},\mathrm{diag}\left(\boldsymbol{X}\right)\right\rangle$$

where:

(a) $\boldsymbol{X}\in\mathbb{R}^{d\times d}$

(b) $\mathrm{diag}:\mathbb{R}^{d\times d}\to\mathbb{R}^{d}$ returns the diagonal of a matrix, that is:

$$\boldsymbol{b}=\mathrm{diag}\left(\boldsymbol{X}\right)\implies\boldsymbol{b}\left[i\right]=\boldsymbol{X}\left[i,i\right]$$

Solution:
Type your solution here...

3 Descent Methods (Gradient Descent and Momentum)

Solve this section in the attached notebook.

4 Constraint optimization

Minimax

Let $G(x,y) = \sin(x+y)$.

4.1

Show that:

1. $\underset{x}{\min}\underset{y}{\max} G(x,y) = 1$
2. $\underset{y}{\max}\underset{x}{\min} G(x,y) = -1$

Solution:
Type your solution here...

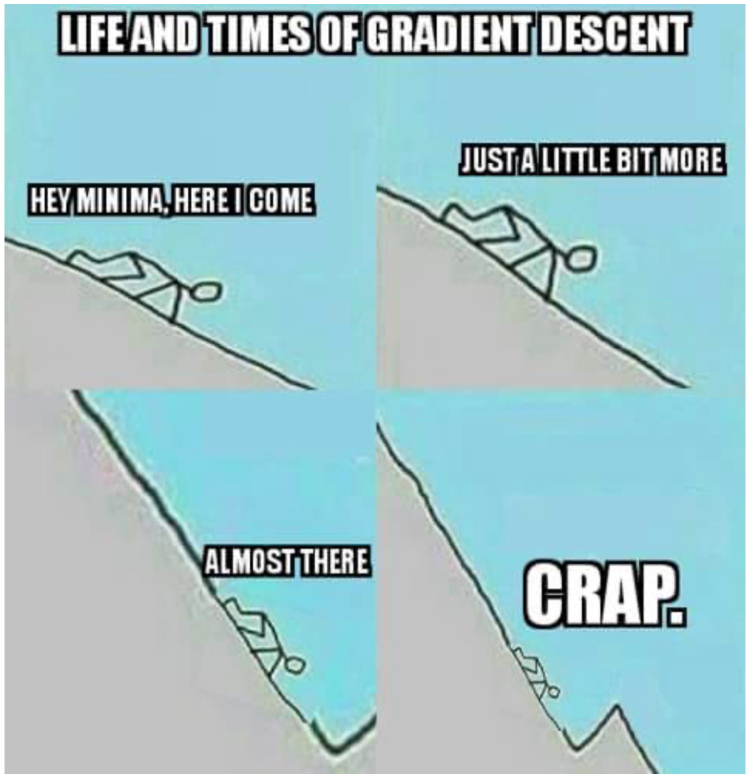
Rayleigh quotient

- The Rayleigh quotient is defined by:
$$f(\boldsymbol{x}) = \frac{\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}$$
for some symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{d \times d}$.

4.2

1. Show that
$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{cases} \min_{\boldsymbol{x}} \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \\ \text{s.t.} \quad \|\boldsymbol{x}\|_2^2 = 1 \end{cases}$$
2. Write the Lagrangian of the constraint objective $\mathcal{L}(\boldsymbol{x}, \lambda)$.
3. Show that:
$$\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \lambda) = 0 \iff \boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$$
in other words, the stationary points $(\boldsymbol{x}, \lambda)$ are the eigenpairs of $\boldsymbol{A}$ (eigenvectors and eigenvalues).

Solution:
Type your solution here...


LIFE AND TIMES OF GRADIENT DESCENT
HEY MINIMA, HERE I COME
JUST A LITTLE BIT MORE
ALMOST THERE
CRAP.