Data Science / Analyse et Traitement de l'Information

TP1: Linear Algebra (100 pts).

---

1 Matrix

1. Find a quadratic polynomial, say $f(x) = ax^2 + bx + c$, such that

$$f(1) = -5, \quad f(2) = 1, \quad f(3) = 11.$$

2. Let a, b be some fixed parameters. Solve the system of linear equations

$$\begin{cases} x + ay = 2 \\ bx + 2y = 3 \end{cases}$$

3. Find the inverse of the following matrix, if it exists:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 0 \end{pmatrix}$$

4. Solve the following matrix equation for X:

$$AX = B, \quad \text{where} \quad A = \begin{pmatrix} 1 & 3 \\ 2 & 5 \end{pmatrix}, \quad B = \begin{pmatrix} 7 & 8 \\ 9 & 10 \end{pmatrix}.$$

2 The importance of the mathematical concept behind a code

1. Download, open and run several times the PYTHON file "TP1/some script.py". Explain the function `def project_on_first(u, v)` in geometrical terms and then explain the results of this code with clear mathematical concepts (you have to talk about projections and scalar product). *It is very important to understand the mathematical concepts behind a program!!*

2. Consider $u = (u_i)_{1 \leq i \leq 7}$ and $v = (v_i)_{1 \leq i \leq 7}$, two vectors of length 7. How would you clearly write the following part of the script using some mathematical notation?

```python
# import numpy library as np in order to work on tensors
import numpy as np
# create 2 random vectors of length 7
u = np.random.randn(7)
v = np.random.randn(7)
# perform some computations
r = 0
for ui, vi in zip(u, v) :
    r += ui * vi
```

Name the mathematics behind and rewrite the last 3 lines of code in one.

3. Complete the code to construct a vector v orthogonal to the vector u and of the same norm. Comment each line of your code.

4. Given two vectors u and v in $\mathbb{R}^n$, write Python code to compute the cosine of the angle between them. Explain the mathematical concept behind this computation, and show how you would write it in one equation.

3 Computing Eigenvalues, Eigenvectors, and Determinants

1. Find the determinant, eigenvectors, and eigenvalues of the matrix

$$A = \begin{bmatrix} 5 & 6 & 3 \\ -1 & 0 & 1 \\ 1 & 2 & -1 \end{bmatrix}.$$

Compare your results with those of the computer.

2. The *covariance matrix* for the n samples $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, each represented by a $d \times 1$ column vector, is given by

$$C = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T,$$

where C is a $d \times d$ matrix and $\boldsymbol{\mu} = \sum_{i=1}^{n} \boldsymbol{x}_i / n$ is the *sample mean*. Prove that C is always positive semidefinite. (Note: A symmetric matrix C of size $d \times d$ is *positive semidefinite* if $v^T C v \geq 0$ for every $d \times 1$ vector v.)

3. In this portion of the exercise, we will calculate the eigenvalues of the covariance matrices of six data sets listed as follows:

filename	n	d	description
`tp1_artificialdata[1-3]`	1024	100	Artificial data generated from various auto-regression (AR-1) models
`tp1_artificialdata4`	1024	100	Random Gaussian data
`tp1_freyfaces`	1965	560	Facial images of a man named Brendan
`tp1_digit2`	5958	784	Hand-written images of "2"

To access each data set, go to `''TP1/data''` and download `tp1_*`. Each file contains a $n \times d$ data matrix with rows representing n different samples in $\Re^d$. For example, `tp1_artificialdata1` contains a data matrix of size 1024×100 (1024 samples, 100 features). Once each dataset has been imported into `Python`, complete the following tasks:

(a) Compute the covariance matrix of each dataset;

(b) Compute the eigenvalues for each covariance matrix;

(c) Compute the determinant of each covariance matrix;

(d) Compute the product of the eigenvalues for each covariance matrix;

(e) Display the eigenspectrum for each covariance matrix in a 2D plot, where the x-axis shows the rank of the eigenvalues, ranging from 1 (the largest eigenvalue) to 100 (the 100-th largest eigenvalue), and the y-axis shows the corresponding eigenvalue.

Describe the relationship between the product of the eigenvalues and the determinant of each covariance matrix. In addition, describe the observations you made regarding the plot of the spectrum of eigenvalues of each covariance matrix. How can you explain the disparities between datasets?

4. Let $A \in \mathbb{R}^{n \times n}$ be a square matrix. Show that if A has an eigenvalue $\lambda = 0$, then the matrix A is singular. Furthermore, prove that the determinant of A must be zero if any eigenvalue is zero.

4 Computing Projection Onto a Line

There are given a line $\alpha : 3x + 4y = -6$ and a point A with the coordinates $(-1, 3)$.

1. Find the distance from the point A to the line α using linear algebra (not the coordinate method).

2. Explain your code using a sketch and some mathematics (there should be a scalar product somewhere).

Submission

Please archive your report and code (*do **not** include dataset*) in "Prénom Nom.zip" (replace "Prénom" and "Nom" with your real name), and upload to "`Upload TP1: Linear Algebra.`" on `https://moodle.unige.ch` before **Monday, October 7 2024, 21:59 PM**. Note, that the assessment is mainly based on your report, which should include your answers to all questions and the experimental results. *Importance is given on the mathematical explanations of your works and your codes should be commented* **Please use this overleaf template.**

Supplements

1. Define and explain the mean and the variance on some examples.

2. Define and explain what is a vector space, a projection and a scalar product.

3. Define and explain what is a vector space, a basis and a change of basis transformation matrix.

4. Present the eigen-value decomposition and the singular-value decomposition.

5. Be ready to answer questions from slide 37 *DS.02.highDimension*.