MATHEMATICS OF DATA SCIENCE

September 12, 2023

Today:

→ Logistics
→ Introduction
→ Basics (notations, etc)

A) Logistics

→ Everything on the course webpage!

https://www.lamsade.dauphine.fr/~croyer/teachMDS.html

→ For any questions, email me!

clement.royer@lamsade.dauphine.fr

→ 12 lectures + 12 tutorials (in groups)

⚠ Schedule will change in the upcoming weeks!

→ End: mid-December (break in October)

→ Evaluation:
- 70% Exam (January, French & English)
- 30% Homework

→ Expectations
- Be on time
- Play along with the English

B) Motivation

About: Maths tools for understanding and modeling data science tasks

→ Data-driven perspective

* Want to use data in the best way possible ⇒ Optimization
* Helpful to consider data originating from a probability distribution

⇒ Statistics

Outline

① Optimization

→ Convexity and why it matters

→ Convex optimization problems

② Statistics

→ Exploit the statistical nature of the data (typically defining an optimization problem)

→ Leverage randomness to obtain approximation results

Most of the time:
Inequalities!

C) Background on linear algebra

$\mathbb{R}^n$: vectors with n real components

$\forall x \in \mathbb{R}^n$, $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ $x_i \in \mathbb{R}$ $\forall i = 1..n$

$x^T := [x_1 \; \ldots \; x_n]$

$\forall (x,y) \in (\mathbb{R}^n)^2$, $x + y = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}$

$\forall x \in \mathbb{R}^n$, $\forall \lambda \in \mathbb{R}$, $\lambda x = \begin{bmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{bmatrix}$

Euclidean norm

$\forall x \in \mathbb{R}^n$, $\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}$

Scalar / inner product

$\forall (x,y) \in (\mathbb{R}^n)^2$,

$$x^T y := \sum_{i=1}^{n} x_i y_i$$

Cauchy-Schwarz inequality

$\forall (x,y) \in (\mathbb{R}^n)^2$,

$$x^T y \leq \|x\| \, \|y\|$$

NB: $\|x\|^2 = x^T x$

Matrix linear algebra

$\mathbb{R}^{m \times n}$: matrices with m rows and n columns

$A \in \mathbb{R}^{m \times n}$

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ A_{m1} & & A_{mn} \end{bmatrix} \qquad A_{ij} \in \mathbb{R}$$

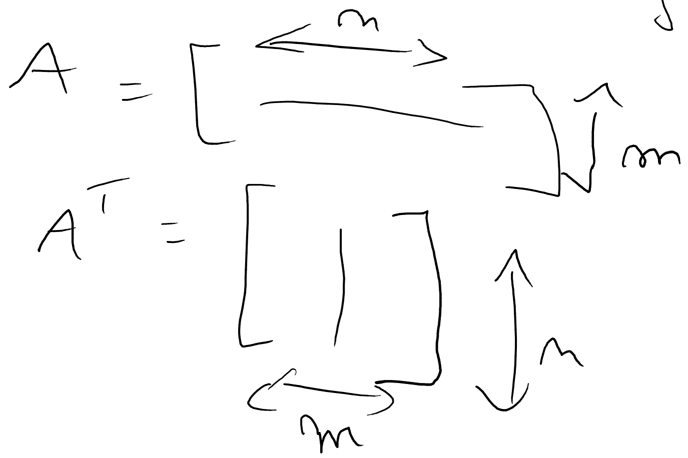
$$A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} \qquad a_i \in \mathbb{R}^n$$

$$A = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \qquad a_j \in \mathbb{R}^m$$

$\forall A \in \mathbb{R}^{m \times n}$, $A = [A_{ij}]_{\substack{1 \le i \le m \\ 1 \le j \le n}}$

A^T "A transpose" $:= [B_{ji}]_{\substack{1 \le j \le n \\ 1 \le i \le m}}$

$B_{ji} = A_{ij} \quad \forall (i,j)$



$A \in \mathbb{R}^{n \times n}$: Square matrix

$A \in \mathbb{R}^{n \times n}$ is symmetric if

$$A^T = A$$

Ex) $I_n = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$ "Identity matrix"

Th) Spectral / Eigen decomposition

Let $A \in \mathbb{R}^{n \times n}$ be symmetric.

Then $$A = U \Lambda U^T$$

where $\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$

$\lambda_i \in \mathbb{R} \; \forall_i$

and U is orthogonal

Orthogonal matrix: $U \in \mathbb{R}^{n \times n}$

is orthogonal if $U U^T = U^T U = I_n$

Product of two matrices

$\forall A \in \mathbb{R}^{m \times n}, \forall B \in \mathbb{R}^{n \times l}$

$$C = AB := [C_{ij}]_{\substack{1 \le i \le m \\ 1 \le j \le l}} \in \mathbb{R}^{m \times l}$$

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

Singular value decomposition

Let $A \in \mathbb{R}^{m \times n}$

Then $A = U \Sigma V^T$

where $U \in \mathbb{R}^{m \times m}$ is orthogonal

$V \in \mathbb{R}^{n \times n}$ is orthogonal

$$\Sigma \in \mathbb{R}^{m \times n}$$

$$r \leq \min(m,n) \qquad \Sigma = \begin{array}{c} \begin{array}{cc} \overleftrightarrow{\;\; r \;\;} & \overleftrightarrow{\;\; n-r \;\;} \end{array} \\ \left[\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right] \end{array} \begin{array}{l} \updownarrow r \\ \updownarrow m-r \end{array}$$

$$\underbrace{\sigma_1 \geq \cdots \geq \sigma_r}_{\text{singular values of } A} > 0$$

r : rank of matrix A

<u>Pseudo-inverse</u>

$\forall\, A \in \mathbb{R}^{m \times n}$, the pseudo-inverse of A is defined as

$$A^+ = V \Sigma^+ U^T \in \mathbb{R}^{n \times m}$$

where $A = U \Sigma V^T$ SVD of A

and

$$\mathbb{R}^{n \times m} \ni \Sigma^+ = \begin{array}{l} \updownarrow r \\ \\ \end{array} \begin{array}{c} \overleftrightarrow{\;\; r \;\;} \\ \left[\begin{array}{ccc|c} 1/\sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & 1/\sigma_r & \\ \hline & 0 & & 0 \end{array}\right] \end{array}$$

D) Motivating example

Solving a linear system of equations

Given $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, find $x \in \mathbb{R}^n$

such that $Ax = b$

⊕ The system may not have a solution

⊖ Maybe it is not suitable to solve it exactly

One approach: Solve a convex optimization problem instead of the linear system exactly

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \|Ax - b\|^2$$

$\Rightarrow$ $x_{opt} = A^+ b$ is a solution of the pb

$\alpha_{opt} := \|A x_{opt} - b\|^2$ optimal value of the problem (lowest error possible)

<u>Pb</u>: Computing A^+ is expensive as $m, n \gg 1$

<u>Statistical approach</u>

$m \gg 1$

Define $\tilde{A} \in \mathbb{R}^{\tilde{m} \times n}$ by sampling $\tilde{m}$ rows of A with $\tilde{m} \ll m$

$$A = \begin{bmatrix} \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \end{bmatrix} \qquad \tilde{A} = \begin{bmatrix} \text{---} \\ \text{---} \end{bmatrix}$$

Define $\tilde{b}$ accordingly and consider

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \|\tilde{A}x - \tilde{b}\|^2$$

$\tilde{A}, \tilde{b}$ are approximations of A, b with smaller sizes

$\Rightarrow$ $(\tilde{x}_{opt}, \tilde{\alpha}_{opt})$ solution

$\tilde{\alpha}_{opt} = \|\tilde{A}\tilde{x}_{opt} - \tilde{b}\|^2$ and optimal value

If rows are sampled appropriately and $\tilde{m} = O(n/\varepsilon)$, then with probability 0.8,

$$\tilde{\alpha}_{opt} \leq \alpha_{opt} \leq (1+\varepsilon)\tilde{\alpha}_{opt}$$

$\Rightarrow$ We'll explain why this is!