DDA3020: Machine Learning - Tutorial 2

- Linear Algebra in Python & Least Square/Linear Regression

Xudong Wang

SDS, CUHK(SZ)

September 20, 2022

Tutorial Information

Tutorial Venue: TB102, Tuesday 18:00 - 21:00pm from Week 2 -Week 13.
S1:18:00-18:50pm, S2: 19:00-19:50pm, S3: 20:00-20:50pm.
All the 3 sessions will give the same tutorial contents.
TAs:

- **Xudong Wang:** xudongwang@link.cuhk.edu.cn
  OH: Tue 4:30-5:30pm, Seat 70, SDS Lab, 4F ZX
- **Dan Qiao:** danqiao@link.cuhk.edu.cn
  OH: Fri 4:00-5:00pm, Seat 1, SDS Lab, 4F, ZX
- **Fei Yu:** feiyu1@link.cuhk.edu.cn
  OH: Thu 3:30-4:30pm, Seat 14, SDS Lab, 4F, ZX
- **Sho Inoue:** shoinoue@link.cuhk.edu.cn
  OH: Tue 5:30-6:30pm, ZOOM: 759 394 8941 (pw123456)

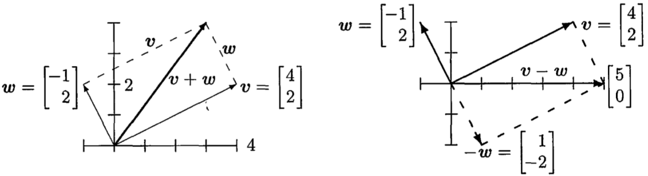
UStaff TAs(Appoint via email or Wechat group in advanced):

- **Yifan Wang:** 119010317@link.cuhk.edu.cn
  OH: Thu 10:30 - 11:30am, start-up zone library
- **Rongxiao Qu:** 120020144@link.cuhk.edu.cn
  OH: Tue 10:30 - 11:30am, start-up zone library
- **Jinrui Lin:** 120090527@link.cuhk.edu.cn
  OH: Fri 10:30 - 11:30am, start-up zone library

Vectors and their Operations

$$\boldsymbol{v} := \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \ \boldsymbol{w} := \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}, \ \boldsymbol{v}^\mathsf{T} = \begin{pmatrix} v_1 & v_2 \end{pmatrix}$$

Vector addition (head to tail) ***At the end of*** $\boldsymbol{v}$***, place the start of*** $\boldsymbol{w}$.



$$\boldsymbol{v} + \boldsymbol{w} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \qquad \boldsymbol{v} - \boldsymbol{w} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} - \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$

Vectors and their Operations

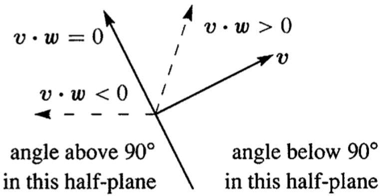
$$\boldsymbol{v} := \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \ \boldsymbol{w} := \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

$$\boldsymbol{v} \cdot \boldsymbol{w} := v_1 w_1 + v_2 w_2 = \boldsymbol{v}^\intercal \boldsymbol{w} = \langle \boldsymbol{v}, \boldsymbol{w} \rangle$$

$$\|\boldsymbol{v}\| := \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}} = \sqrt{v_1^2 + v_2^2}$$

Triangle inequality $\|\boldsymbol{w} + \boldsymbol{v}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\|$

Cauchy-Schwarz inequality $|\boldsymbol{v} \cdot \boldsymbol{w}| \leq \|\boldsymbol{v}\| \, \|\boldsymbol{w}\|$



Matrices and their Operations

A matrix $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ is represented as

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \ddots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

- $A^\intercal \in \mathbb{R}^{n \times m}$ is defined by swapping columns with rows of A: the $i-$th column of $A^\intercal$ is set to be the $i-$th row of A.
- A is called *square* if $m = n$
- A square matrix A is called *symmetric* if $A^\intercal = A$
- A square matrix $A \in \mathbb{R}^{n \times n}$ is called *invertible* if $\exists B \in \mathbb{R}^{n \times n}$ such that $AB = BA = I$, $B := A^{-1}$.
- A is called *orthogonal* if $A^\intercal A = I$
- In the space of $\mathbb{R}^{m \times n}$ matrices, $\langle A, B \rangle := \mathrm{tr}(A^\intercal B) = \mathrm{tr}(AB^\intercal) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}$

An Example

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $F : \mathbb{R}^n \to \mathbb{R}^m$, where $F(x) = (F_1(x), F_2(x), \cdots, F_m(x))^\intercal$ and $F_i : \mathbb{R}^n \to \mathbb{R}$

- The gradient of f at $x \in \mathbb{R}^n$ is

$$\nabla f = \begin{pmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{pmatrix} \in \mathbb{R}^n,$$

where $\frac{\partial}{\partial x_i} f(x)$ denotes the *partial derivative* of f at x w.r.t. x_i.

- The *Jacobian-Matrix* of F at x is

$$DF(x) = J_f(x) = \begin{pmatrix} \nabla F_1(x)^\intercal \\ \nabla F_2(x)^\intercal \\ \vdots \\ \nabla F_m(x)^\intercal \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x_1} F_1(x) & \cdots & \frac{\partial}{\partial x_n} F_1(x) \\ \vdots & \ddots & \\ \frac{\partial}{\partial x_1} F_m(x) & \cdots & \frac{\partial}{\partial x_n} F_m(x) \end{pmatrix}$$

Matrix norms

The matrix norm $\|\cdot\| : \mathbb{K}^{m\times n} \to \mathbb{R}$ satisfies for any $A, B \in \mathbb{K}^{m\times n}$ and $\alpha\mathbb{R}$

- $\|A\| \geq 0$
- $\|A\| = 0$ i.f.f. $A = \mathbf{0}$
- $\|\alpha A\| = |\alpha| \, \|A\|$
- $\|A + B\| \leq \|A\| + \|B\|$

Examples:

- (Frobenius norm) $\|A\|_F = \sqrt{\sum_{i,j=1}^{n} |a_{ij}|^2} = \mathrm{tr}(\sqrt{A^\intercal A}) = \sqrt{\sum_{i=1}^{\min\{i,j\}} \sigma_i^2(A)}$
  Where $\sigma_i(A)$ is the i-th eigenvalue of matrix A.

Subspace

- *Subspace* $S \subseteq \mathbb{R}^n$:
  - $\boldsymbol{v} + \boldsymbol{w} \in S, \forall \boldsymbol{v}, \boldsymbol{w} \in S$
  - $c\boldsymbol{v} \in S, \forall c \in \mathbb{R}, \boldsymbol{v} \in S$
- The *column space* of the matrix $A \in \mathbb{R}^{m\times n}$ consists of all linear combinations of the columns, denoted as $\mathcal{C}(A)$:

$$\mathcal{C}(A) := \{y \in \mathbb{R}^m \mid y = Ax, \forall x \in \mathbb{R}^n\} \subset \mathbb{R}^m$$

- The *nullspace* of the matrix $A \in \mathbb{R}^{m\times n}$ consists of all solutions to $Ax = 0, x \in \mathbb{R}^n$, denoted by $\mathcal{N}(A)$:

$$\mathcal{N}(A) := \{x \in \mathbb{R}^n \mid Ax = 0\} \subset \mathbb{R}^n$$

Rank

Let $A = (\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n) \in \mathbb{R}^{m \times n}$ with $\boldsymbol{a}_i \in \mathbb{R}^m$,

- A sequence of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_k$ is said to be *linear independent* if

$$\sum_{i=1}^{k} x_i \boldsymbol{v}_i = 0 \Rightarrow x_1 = x_2 = \cdots = x_k = 0.$$

- The *rank* of A is the maximum number of linearly independent columns of A:

$$\mathrm{rank}(A) = \dim(\mathcal{C}(A)).$$

- If there exists $x \in \mathbb{R}^n$ for $Ax = b$, then the vector $b \in \mathbb{R}^m$ can be represented as the linear combination of the columns of A, i.e., $b = \sum_{i=1}^{n} x_i \boldsymbol{a}_i \in \mathcal{C}(A)$

Systems of Linear Equations/Least Square

$A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, When b is not in the column space of A. Find the solution $x \in \mathbb{R}^n$ of

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|_2^2 \tag{1}$$

Let the derivative $A^\intercal(Ax - b) = 0$ yields

$$A^\intercal Ax = A^\intercal b$$

if $A^\intercal A$ is **invertible**,

$$x = (A^\intercal A)^{-1} A^\intercal b$$

Programming Part

Move to Jupyter Notebook

Summary

- Linear Algebra Review
- Linear Algebra in Python via Numpy
- Least Square/Linear Regression and Optimization

Thank you for listening!

Xudong Wang
xudongwang@link.cuhk.edu.cn

September 20, 2022