Warm-up sessions

Copernicus Master on Digital Earth

Linear Algebra and probabilities for data science

Prof. Nicolas Courty
ncourty@irisa.fr

Linear Algebra

Set of mathematical tools operating in the continuous domain (by opposition to discrete mathematics) essential to understanding machine learning tools.

We will discuss among other things :

- scalar values, vector matrices, tensors
- basic operations between these quantities (addition, products)
- Vector spaces generated by a base, independence
- diagonalization, factorization

Quantities

Scalar variables: denoted by a lowercase letter

- e.g. $x \in \mathbb{R}$ is the slope of a straight line
- e.g. $n \in \mathbb{N}$ is the number of elements in a set

Vector variables: array of ordered values, denoted in lowercase bold

- $\mathbf{v} \in \mathbb{R}^{256}$: point in a real space at 256 dimensions
- $\mathbf{v}^T = [v_1 v_2 \dots v_i \dots v_n]$, every value of the table is indexed by an integer $i \in \{1, n\}$.
- the values v_i are the coordinates along each axis of the space

Matrix variables: two-dimensional (2D) array of ordinate values, written uppercase BOLD

- $\mathbf{A} \in \mathbb{R}^{3\times3}$: matrix expressing a application of $\mathbb{R}^3 \to \mathbb{R}^3$
- these values are indexed by i (line number) and j (line number): $A_{i,j}$

Tensors

Sometimes extra dimensions are needed to translate complex relationships between several elements

- e.g. multispectral image

tensors: n-dimensional (n-d) array of ordered values (based on an regular grid), upper-case denoted $\mathbf{BOLD}$

- $\mathbf{T} \in \mathbb{R}^{28\times 28}$: tensor of $\mathbb{R}^{28\times 28\times 28}$
- these values are indexed by $\mathrm{i}, \mathrm{j}, \mathrm{k}, \ldots$: $\mathrm{T}_{\mathrm{i},\mathrm{j},\mathrm{k}}$ in the previous case

Operations

Addition/multiplication of matrices

If $\mathbf{A}$ and $\mathbf{B}$ are the same size (e.g. $m \times n$)

- $\mathbf{C} = \mathbf{A} + \mathbf{B} \equiv \mathbf{C}_{ij} = \mathbf{A}_{ij} + \mathbf{B}_{ij}, \forall i, j$
- addition/multiplication by a scalar : $\mathbf{C} = e\mathbf{A} + f \equiv \mathbf{C}_{ij} = e\mathbf{A}_{ij} + f$
- addition of a vector (non-standard notation): $\mathbf{C} = \mathbf{A} + \mathbf{v} \equiv \mathbf{C}_{ij} = \mathbf{A}_{ij} + \mathbf{v}_j$
- also called broadcasting in English

Operations

If $\mathbf{v}$ and $\mathbf{u}$ are the same size m then

- we note $\mathbf{v}^{\mathbf{T}}\mathbf{u}$ the scalar product between these two vectors
- $\mathbf{v}^{\mathbf{T}}\mathbf{u} = \sum_{\mathrm{k}=1}^{\mathrm{m}} \mathbf{v}_{\mathrm{k}}\mathbf{u}_{\mathrm{k}}$

If $\mathbf{A}$ and $\mathbf{B}$ are $\mathrm{m} \times \mathrm{r}$ and $\mathrm{r} \times \mathrm{n}$ then

- $\mathbf{C} = \mathbf{A}\mathbf{B}$ is of size $\mathrm{m} \times \mathrm{n}$
- $\mathbf{C}_{\mathrm{ij}} = \sum_{\mathrm{k}=1}^{\mathrm{r}} \mathbf{A}_{\mathrm{ik}} + \mathbf{B}_{\mathrm{kj}}$
- $\mathbf{C}_{\mathrm{ij}}$ is the scalar product between line i of $\mathbf{A}$ and column j of $\mathbf{B}$

Operations

Matrix Product Properties

- Distributivity versus addition : $\mathbf{C}(\mathbf{A}+\mathbf{B}) = \mathbf{CA}+\mathbf{CB}$.
- Associativity: $\mathbf{C}(\mathbf{AB}) = (\mathbf{CA})\mathbf{B}$
- Non-commutating: generally $\mathbf{AB} \neq \mathbf{BA}$
- but the scalar product is : $\mathbf{v}^{\mathbf{T}}\mathbf{u} = \mathbf{u}^{\mathbf{T}}\mathbf{v}$
- transposed from a product: $(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$

Linear systems

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

.

- If $\mathbf{A}$ is of size $\mathrm{m} \times \mathrm{n}$, $\mathbf{x}$ and $\mathbf{b}$ from sizes n, we have a system with m equations and n unknown
- case where $\mathrm{m} = \mathrm{n}$. Then the system solution, if it exists, is given by:

  $$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

  .
- $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$, i.e. such that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_{\mathrm{n}}$
- $\mathbf{I}_{\mathrm{n}}$ size identity matrix $\mathrm{n} \times \mathrm{n}$

resolution of linear systems

- Gauss Pivot

$$\begin{aligned} x + 3y - 2z &= 5 \\ 3x + 5y + 6z &= 7 \\ 2x + 4y + 3z &= 8 \end{aligned}$$

$$\left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 3 & 5 & 6 & 7 \\ 2 & 4 & 3 & 8 \end{array}\right] \overset{L_2-3L_1 \rightarrow L_2}{\sim} \left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 0 & -4 & 12 & -8 \\ 2 & 4 & 3 & 8 \end{array}\right] \overset{L_3-2L_1 \rightarrow L_3}{\sim} \left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 0 & -4 & 12 & -8 \\ 0 & -2 & 7 & -2 \end{array}\right] \overset{-L_2/4 \rightarrow L_2}{\sim} \left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 0 & 1 & -3 & 2 \\ 0 & -2 & 7 & -2 \end{array}\right]$$

$$\sim \left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 0 & 1 & -3 & 2 \\ 0 & 0 & 1 & 2 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & 2 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 3 & 0 & 9 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & 2 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & -15 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

- many other methods exist (LU, decomposition of Cholesky, with additional constraints on the shape of $\mathbf{A}$)
- $\mathbf{A}^{-1}$ does not depend on $\mathbf{b}$, and can be used to solve several problems
- however in practice $\mathbf{A}^{-1}$ is never really calculated as is ($\mathrm{o}(\mathrm{n}^3)$ operations) and we use the value of $\mathbf{b}$ in the resolution (looking for example at the difference between $\mathbf{b}$ and $\mathbf{Ax}$ at a given iteration

$$\arg\min_{\mathrm{x}} ||\mathbf{Ax} - \mathbf{b}||^2$$

Vector space

Space generated by a set of vectors

The space generated by a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is the set of points formed by all linear combinations of these vectors, i.e. $\mathbf{p} = \sum_i \alpha_i \mathbf{v}_i$

- to know if $\mathbf{A}\mathbf{x} = \mathbf{b}$ admits a solution amounts to to know if $\mathbf{b}$ is in the space generated by $\mathbf{A}$
- if $\mathbf{A}$ is of size $n \times n$, then $\mathbf{A}$ must be formed of linearly independent vectors
- $\mathrm{rang}(\mathbf{A}) = n$.

For matrices of size $m \times n$, the rank of $\mathbf{A}$ is at better m. Other methods of inversion exist for this type of problem.

- Example: Moore Penrose's inverse nickname $\mathbf{A}^+$
- if $\mathrm{rank}(\mathbf{A}) = m$, $\mathbf{A}^+ = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$
- if $\mathrm{rank}(\mathbf{A}) = n$, $\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$

Norms

A norm is a function f used to measure the length of a vector. It respects the 3 following conditions:

- $f(\mathbf{v}) = 0 \implies \mathbf{v} = 0$.
- $f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$ (triangular inequality)
- $f(\alpha\mathbf{x}) = |\alpha| f(\mathbf{x})$ (homogeneity)

Examples: L^p norm : $||\mathbf{x}|||_p = \sqrt[p]{\sum_i |\mathbf{x}_i|^p}$.

- $p = 2$. Euclidean norm, also note that $||\mathbf{x}||_2^2 = \mathbf{x}^T\mathbf{x}$.
- $p = 1$ useful to differentiate values close to 0
- $p = \infty$ max norm $||\mathbf{x}||_\infty = \max_i |\mathbf{x}_i|$

Special cases of matrices

- Diagonal Matrix: non-zero entries only on the diagonal. One notes $\mathbf{V} = \mathrm{diag}(\mathbf{v})$
- Symmetrical Matrix: $\mathbf{A} = \mathbf{A}^{\mathrm{T}}$
- Orthogonal matrix: $\mathbf{A}^{-1} = \mathbf{A}^{\mathrm{T}}$. All columns $\mathbf{a}_{\bullet\mathrm{i}}$ are orthogonal to each other, i.e. $\forall \mathrm{i}, \mathrm{j} | \mathrm{i} \neq \mathrm{j}, {\mathbf{a}_{\bullet\mathrm{i}}}^{\mathrm{T}} \mathbf{a}_{\bullet\mathrm{j}} = 0$

Matrix decomposition

Matrices can be decomposed into factors (product of matrices) to gain understanding of their structures.

- Spectral decomposition, also called decomposition into eigen values/vectors
- an eigenvector $\mathbf{v}$ of a matrix $\mathbf{A}$ is such that it exists a scalar λ such as

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

.

  - λ is the eigenvalue associated with $\mathbf{v}$.
  - all multiples of $\mathbf{v}$ (i.e. for $s \in \mathbb{R}$, we form $s\mathbf{v}$) are considered to be eigenvectors of $\mathbf{A}$

- finding the eigenvalues is equivalent to solving $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$.
- the roots of this polynomial are the eigenvalues of $\mathbf{A}$.
- example with

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Eigenvector decomposition (Eigendecomposition)

If $\mathbf{A}$ has n independent eigenvectors then it can be written as follows

$$\mathbf{A} = \mathbf{V}\mathrm{diag}(\lambda)\mathbf{V}^{-1}$$

- $\mathbf{V}$: eigenvector matrix
- λ : eigenvalue vector

Case of a symmetrical matrix $\mathbf{S}$:

$$\mathbf{S} = \mathbf{V}\mathrm{diag}(\lambda)\mathbf{Q}^{\mathrm{T}}$$

- $\mathbf{Q}$ is an orthogonal matrix
- λ : eigenvalue vector

Useful properties of eigenvalue decomposition

- A matrix is singular if at least one of its eigenvalue is zero
- the rank of a matrix is equal to the number of non null eigenvalues
- the eigenvalue decomposition is useful to solve some optimization problems
- ex: $\mathrm{f}(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ with $|||\mathbf{x}||_2 = 1$
  - quadratic form
  - if $\mathbf{x}$ is an eigenvector, then $\mathrm{f}(\mathbf{x})$ is the corresponding eigenvalue
  - $\min \mathrm{f} = \min \lambda$, $\max \mathrm{f} = \max \lambda$.
- all strictly positive eigenvalues: positive matrix
  - guarantees that $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} \geq 0$
- all eigenvalues positive or null: matrix Positive Semi-Definite (PSD)

SVD decomposition

The SVD decomposition (Singular Values Decomposition) is a more general decomposition that fits to the rectangular matrices:

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}$$

.

- If $\mathbf{A}$ is size $\mathrm{m} \times \mathrm{n}$, then $\dim(\mathbf{U}) = \mathrm{m} \times \mathrm{m}$, $\dim(\mathbf{D}) = \mathrm{m} \times \mathrm{n}$ and $\dim(\mathbf{V}) = \mathrm{n} \times \mathrm{n}$
- the elements of $\mathbf{D}$ are called singular values
- the vectors of $\mathbf{U}$ and $\mathbf{V}$ are respectively vectors right and left singulars (resp. eigenvectors of $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$)
- practical aspect: pseudo-inverse calculation with SVD : $\mathbf{A}^{+} = \mathbf{V}\mathbf{D}^{+}\mathbf{U}^{\mathrm{T}}$, where $\mathbf{D}^{+}$ is formed with the reciprocal of the non-zero elements of $\mathbf{D}$.

Practical session

Let's practice now:

- Numpy, scipy and matplotlib for manipulating data matrices
- Notebooks

The end.