Approximation of functions
I. Metric spaces

Kamel ATTAR

Abstract

This worksheet explores the concepts of metric spaces through several key examples. In the first part, it addresses distance in finite spaces, such as graphs and sets of words. Next, it examines the distance between vectors, often calculated using the Euclidean norm. The worksheet also deals with distances in an abstract manner, without a specific context, to understand their fundamental properties. Finally, it looks at function spaces, using distances such as L_p to compare functions over a given domain.

Contents

1 Distances over finite spaces

Distance in a graph is a key measure for evaluating the proximity between two nodes. This concept is fundamental for many applications in computer science, such as finding optimal paths, detecting cycles, or analyzing networks, and plays a central role in modeling and solving problems on discrete structures.

Exercice 1 **Distance in a graph**

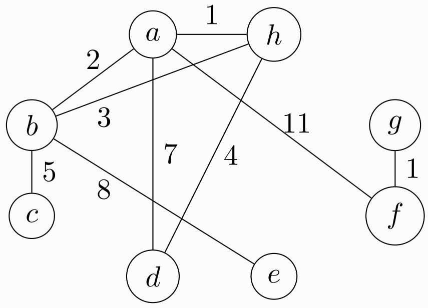
For mathematicians, an ***undirected simple graph*** is a couple $G = (S, E)$, where S is a finite set whose elements are called ***vertices*** of G, and E a set of pairs of elements of G called ***edges*** of G. A given graph G is said to be ***weighted*** if it is considered together with a map $w : E \to \mathbb{R}$ associating a weight to each edge.

❶ Represent graphically the weighted graph G given by:

- $S = \{0, \dots, 5\}$
- $E = \{\{0,1\}, \{1,2\}, \{2,5\}, \{1,3\}, \{3,4\}, \{3,5\}\}$
- w is the constant map over E equal to 2.

❷ Reminder: a path between two vertices s_1 and s_2 of a weighted graph (G, w) is a sequence of contiguous edges whose starting and arrival vertices are S_1 and s_2 respectively. Under which condition(s) will the shortest path between two vertices (that is to say, the sum of the weights of the edges of such path) define a distance over S?

❸ Given that these conditions are duly fulfilled, think about how to determine the balls of radius $r > 0$ centred at a given vertex. Then, find the closed ball centered at a with radius 6 for the following graph.



The data structures handled by a computer scientist are inherently discrete and finite, especially when implemented. Concepts of distance between these structures, such as for strings, or between the elements they contain, naturally arise. These notions can influence the evaluation of an algorithm's complexity when using these structures or be integrated into the modeling of the object itself. In this section, we examine two simple examples of distances applicable in this type of context: **Hamming distance**, which measures the difference between two strings of the same length, and **Levenshtein distance**, which quantifies the minimum number of operations required to transform one string into another.

Exercice 2 **Hamming and Levenshtein distance**

In this exercise, we get interested in two classic ditances over ***words***, that is to say finite sequences of characters from a given alphabet, or just character strings as you are used to call them.

❶ What is the Hamming distance between two words ? Explain why it is a distance mathematically speaking.

❷ What is the Levenshtein distance between two words ? Explain why it is a distance mathematically speaking.

2 Distances between vectors

In this section, we talk about distances as they appear in geometry. More specifically, we study examples allowing to measure a distance between two points of $\mathbb{R}^n$ for $n \in \mathbb{N}_{\geq 1}$. Some elements of this section will be dwelled upon more in detail in the `FPVA` lesson concerning the functions of several variables. Notably, you will see how the notion of distance enables us to extend the notion of smoothness[1] to the functions of several variables.

Exercice 3 — **p-norm**

Let $p \in [1, +\infty[$ and $n \in \mathbb{N}_{\geq 1}$. We call ***p-norm*** over $\mathbb{R}^n$ the norm $\|\cdot\|_p$ defined for every $x \in \mathbb{R}^n$ by

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

This definition is extended to the case $p = +\infty$ by:

$$\|x\|_\infty = \max_{1 \leq i \leq n} \{|x_i|\}$$

This norm induces the distance d_p defined for every x, $y \in \mathbb{R}^n$ by

$$d_p(x, y) = \|x - y\|_p.$$

We will look more closely at $p \in \{1, 2, +\infty\}$.

❶ Prove[a] that d_1, d_2 and d_∞ are distances over $\mathbb{R}^n$.

❷ Represent graphically the distance between two vectors of $\mathbb{R}^2$ for the distances d_1, d_2 and d_∞.

❸ Represent graphically the balls centred at $(0, 0)$ of radius 1 in $\mathbb{R}^2$ for each of the distances d_1, d_2 and d_∞.

❹ What is the shape of those balls if we now work in $\mathbb{R}^3$?

❺ Justify that the following inequalities are true for every $x \in \mathbb{R}^n$:

$$\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq n\|x\|_\infty.$$

❻ Characterise simply the fact that a sequence $(x^{(k)})_{k\in\mathbb{N}}$ of vectors of $\mathbb{R}^n$ converges towards a limit $\ell \in \mathbb{R}^n$.

[a] In order to prove the triangular inequality for d_2, we will use the Cauchy-Schwarz inequality:

$$\left|\sum_{i=1}^{n} \alpha_i \beta_i\right| \leq \left(\sum_{i=1}^{n} \alpha_i^2\right)^{\frac{1}{2}} \cdot \left(\sum_{i=1}^{n} \beta_i^2\right)^{\frac{1}{2}}$$

Exercice 4 — **Distance doesn't always originate from a norm**

Find examples of distances over $\mathbb{R}^n$ for $n \in \mathbb{N}_{\geq 1}$ that are not induced by a norm over $\mathbb{R}^n$.

3 Distances without context

It seems natural, when looking back at the already studied examples, to imagine that every set comes together with an ***obvious*** definition for a distance. This however is a completely false perception: the notion of distance is strongly related to the context we want to model. There are particularly on every set some distances that only add few if any usable knowledge.

[1] Continuity, differentiability etc.

Exercice 5 **Discrete distance**

Let E be whatever (nonempty) set. Let us denote by d the map defined for every x, $y \in E$ by

$$d(x,y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{else} \end{cases}$$

❶ Show that d defines a distance over E.

❷ What information does d bring?

Exercice 6 **String distance**

E is a finite set that has been ordered, each element being indexed by its rank. Thus E can be written $E = \{x_1, \ldots, x_n\}$. Let d be the map defined for every $(x_p, x_k) \in E^2$ by

$$d(x_k, x_p) = |p - k|.$$

❶ Prove that d defines a distance over E.

❷ What information does d bring?

4 Distances over spaces of functions

We start here working with the classic examples of distances over spaces of functions. The elements introduced here will be later used in the lessons about signal mathematics, continuous probabilities and numerical analysis. These are fields of mathematical engineering in which the notion of approximation of a function (it may represent an analogue signal, a probability density or a cost function) by sequences of functions easier to work with (polynomial or trigonometric sequences, easier probability densities, piecewise linear maps) are of the utmost importance. Yet, approximating something implies measuring the distance towards this thing.

Exercice 7 **p-norm over $\mathcal{C}(I, \mathbb{R})$**

Let $p \in [1, +\infty[$ and let I be a segment of $\mathbb{R}$. We denote by $\mathcal{C}(I, \mathbb{R})$ the vector space of continuous functions over I. We call p-norm over $\mathcal{C}(I, \mathbb{R})$ the norm defined for every $f \in \mathcal{C}(I, \mathbb{R})$ by

$$\|f\|_p = \left(\int_I |f(x)|^p \mathrm{d}x \right)^{1/p}.$$

This definition is extended to $p = +\infty$ by: $\|f\|_\infty = \sup_{x \in I}\{|f(x)|\}$
This norm induces the distance d_p defined for every f, $g \in \mathcal{C}(I, \mathbb{R})$ by

$$d_p(f, g) = \|f - g\|_p.$$

We insist particularly on $p \in \{1, 2, +\infty\}$.

❶ Prove[a] that d_1, d_2 and d_∞ are distances over $\mathcal{C}(I, \mathbb{R})$.

❷ Find sequences of functions that converge towards the 0 function when relating to d_1 yet not to d_∞. Explain why this situation is totally different from what can happen in $\mathbb{R}^n$.

[a] In order to prove the triangular inequality for d_2, once again we have to use the Cauchy-Schwarz inequality:

$$\left| \int_I fg \right| \leq \left(\int_I f^2 \right)^{\frac{1}{2}} \cdot \left(\int_I g^2 \right)^{\frac{1}{2}}$$

Exercice 8

❶ We use the same context than in exercise 4 except that we loosen the hypotheses concerning the regularity (smoothness) of the functions. We work now in the vector space $\mathcal{E}$ of piecewise continuous[a] functions over I.

Do the expressions of d_1, d_2 and d_∞ still define distances on $\mathcal{E}$?

❷ More generally, what is the minimal necessary condition on functions for d_∞ to define a distance on their set?

[a] These are the functions that are continuous over I except in a finite number of discontinuity points, where they still admit a finite limit on the left and on the right, except obviously at the boundaries of I where they can admit only one.

A Reminders: Supremum of a subset of $\mathbb{R}$

Here are reminded some elements pertaining to the notion of supremum of a subset of $\mathbb{R}$, and more particularly what concerns the supremum of a function.

Definition A.1. Given a subset A of $\mathbb{R}$, we call ***upper bound*** of A any real number M such that

$$\forall x \in A, \quad x \leq M.$$

It means that every element of A is less than M.

Example A.1. For the segment $[0, 1]$ of $\mathbb{R}$, elements 2, e and 1 are upper bounds of $[0, 1]$. Among all the possible upper bounds of $[0, 1]$, 1 is the least; this least upper bound here belongs to $[0, 1]$.

Example A.2. The interval $[0, 1[$ still admits 2, e and 1 as upper bounds. The real number 1 is still the least upper bound, yet it does not belong to $[0, 1[$.

The interest of determining the least upper bound in the previous cases comes from a standard approach in analysis. Analysis consist to a large extent in estimating, finding upper and lower bounds. And when one gets an estimation, lower or upper bound of an expression, the next question is whether this result is the best that can be reached; so, when finding upper bounds it is natural to aim for the best one, that is to say the least one.

Except for the obvious cases when A is empty or not bounded above, in which case we say that the least upper bound is respectively $-\infty$ and $+\infty$, a nonempty bounded-above subset of $\mathbb{R}$ always admits a least upper bound. This property is deeply anchored in the construction of $\mathbb{R}$, through what we call the ***completeness*** of real numbers, and the contrary assumption (that a bounded-above subset may have no least upper bound) would be in contradiction with the way the set of real numbers has been mathematically built.

Definition A.2. Let A be a ***nonempty*** bounded-above subset of $\mathbb{R}$. We call ***supremum*** of A the least upper bound of A.

Hypothesis A.1. The supremum of $\emptyset$ is $-\infty$, that of an unbounded-above subset is $+\infty$.

Notation. We denote by $\sup(A)$ the supremum of a subset A of $\mathbb{R}$.

As we already noticed, the supremum of a subset A of $\mathbb{R}$ does not always belong to A. Particularly, if A is a right-open interval, its supermum is not an element of A.

Definition A.3. Let A be a subset of $\mathbb{R}$. When $\sup(A)$ belongs to A, it is called ***maximum*** and denoted by $\max(A)$.

Remark 1. An upper bound m of a subset A that belongs to A is necessarily the maximum of A. Indeed, m has to be less than any upper bound of A since it is an element of A. As it is also supposed to be an upper bound of A, it is the least upper bound of A.

We call ***supremum of a function*** $f : I \to \mathbb{R}$ the supremum of the subset of $\mathbb{R}$

$$f(I) = \{f(x) \mid x \in I\}.$$

More accurately, it is the least upper bound of the set of values reached by f on the interval I. We shortcut this denomination by ***sup(remum) of*** f or ***max(imum) of*** f when this supremum is actually reached by f for an starting value in I.

We try M to be an upper bound of $f(I)$ through the following estimation bias

$$\forall x \in I, \quad f(x) \leq M. \tag{M upper bound}$$

In that case, M is an upper bound of f. The supremum of f being its least upper bound, the equation (M upper bound) systematically implies a relation of the type

$$\sup_{x \in I} f(x) \leq M.$$

Given how a maximum is defined, the function f admits a maximum if there exists a value $m \in I$ such that $f(m)$ is equal to the supremum de f.

Example A.3. The real numbers 1, 2, π, 10^{10} are all upper bounds of the sin function. 1 is the supremum of sin, it is even a max since it is reached in $\frac{\pi}{2}$ (among others).

Example A.4. The numbers 1, 10 and 32 are upper bounds of the function defined over $\mathbb{R}$ by $f : x \mapsto 1 - e^{-x}$. It seems natural, given that f is increasing and tends towards 1 in $+\infty$, to suppose that 1 is the supremum of f. This is true, saying that the limit of $f(x)$ when $x \to +\infty$ is the upper bound 1 justifies this. Indeed, let us assume that the upper bound 1 is not the supremum of f; if we denote by M said supremum, we thus get the relation $M < 1$. However, as f tends towards 1 in $+\infty$,there exists values of f in the interval $]M, 1]$, which is a contradiction to M being an upper bound; thus, 1 is the supremum of f.

The previous example can be generalised. Given a function $f : I \to \mathbb{R}$ and M an upper bound of f, if there exists a sequence (x_n) of elements of I such that $\big(f(x_n)\big)$ converges towards M, then M is the supremum of f. The proof would be similar to the one given for the example.