Tutorial on Topological Data Analysis

Kunlin

July 15, 2020

Contents

1 Tutorial on LaTeX **1**

2 Metric Space **2**

Abstract

This is for the first tutorial of topological data analysis.

1 Tutorial on LaTeX

This is a **bold text**.
This is a underlined text.
This is a *italic text*.

This could also be ***combined***.
If you don't want to choose manually, you could use *emphsize*.
emphsize *under italic text*. ***emphasize* under bold text**.

- List one
- List two

1. Number one
2. Number two

Metric: (i) $d(x, y) \geq 0$ for all x, $y \in X$.

$$d(x,y) = \begin{cases} 0000000 & \text{if } x = y, \\ 1 & \text{if } x \neq y, \end{cases} \tag{1}$$

2 Metric Space

Definition: Let X be a set and $d : X^2 \to \mathbb{R}$ a function with the follow properties:

(i) $d(x, y) \geq 0$ for all x, $y \in X$.

(ii) $d(x, y) = 0$ if only if $x = y$.

(iii) $d(x, y) = d(y, x)$ for all x, $y \in X$.

(iv) $d(x, y) + d(y, z) \geq d(x, z)$ for all x, y, $z \in X$. (This is called the *triangle inequality*).

Then we say that d is a *metric* on X and that (X, d) is a *metric space*.

Take away:

(i) Distances are always positive.

(ii) Two points are zero distance part if and only if they are the same point.

(iii) The distance form A to B is the same as the distance from B to A.

(iv) The distance form A to B via C is at least as great as the distance from A to B directly.

Exercise 2.1. *If $d : X^2 \to \mathbb{R}$ is a function with the following properties:*
(ii) $d(x, y) = 0$ if and only if $x = y$
(iii) $d(x, y) = d(y, x)$ for all $x, y \in X$
(iv)$d(x, y) + d(y, z) \geq d(x, z)$ for all $x, y, z \in X$
prove that d is a metric on X
[Thus condition (i) of the definition is redundant.]

Solution. Setting $z = x$ in condition (iv):

$$\begin{aligned} d(x, y) + d(y, x) &\geq d(x, x) \\ 2d(x, y) = d(x, y) + d(y, x) &\geq d(x, x) = 0 \\ 2d(x, y) &\geq 0 \\ d(x, y) &\geq 0 \end{aligned} \tag{2}$$

Definition 2.2. *A* metric *on the set X is a function d: $X \times X \to [0, \inf)$ such that the following conditions are satisified for all $x, y, z \in X$:*
(M1) Positive property: $d(x, y) = 0$ if and only if $x = y$
(M2) Symmetry property: $d(x, y) = d(y, x)$
(M3) Triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$

Example 2.3. *The set $X = \mathbb{R}$ with $d(x, y) = |x - y|$, the absolute value of the difference of $x - y$, prove that $d(x, y)$ is a metric on X.*

Solution. M1 and M2 are obvisous.
For M3:

$$d(x, y) = |x - y| = |x - z + z - y| \leq |x - z| + |z - y| = d(x, z) + d(z, y) \quad (3)$$

Example 2.4. *Let $X = \mathbb{R}^n$ and let $d : \mathbb{R}^n \times \mathbb{R}^n \to [0, \inf)$ be defined by:*

$$d((x_1, x_2 \cdots x_n), (y_1, y_2 \cdots y_n)) = \sum_{i=1}^{n} |x_i - y_i| \quad (4)$$

prove that $d(x, y)$ is a metric on X.

Solution.

$$\begin{aligned} &d((x_1, x_2 \cdots x_n), (y_1, y_2 \cdots y_n)) \\ &= \sum_{i=1}^{n} |x_i - y_i| \\ &= \sum_{i=1}^{n} |x_i - z_i + z_i - y_i| \\ &\leq \sum_{i=1}^{n} |x_i - z_i| + |z_i - y_i| \\ &= d((x_1, x_2 \cdots x_n), (z_1, z_2 \cdots z_n)) + d((z_1, z_2 \cdots z_n), (y_1, y_2 \cdots y_n)) \end{aligned} \quad (5)$$

Example 2.5. *The* Euclidean metric *on $\mathbb{R}^n$ is defined by formula:*

$$d((x_1, x_2 \cdots x_n), (y_1, y_2 \cdots y_n)) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (6)$$

for each $(x_1, x_2 \cdots x_n), (y_1, y_2 \cdots y_n) \in \mathbb{R}^n$, prove that $d(x, y)$ is a metric on X.

Solution. Let $x = (x_1, x_2 \cdots x_n)$, $y = (y_1, y_2 \cdots y_n)$, $z = (z_1, z_2 \cdots z_n)$.
Let $r_i = x_i - z_i$ and $s_i = z_i - y_i$, we need to prove that

$$\begin{aligned} d(x, y) &= \sqrt{\sum_{i=1}^{n} (r_i + s_i)^2} \\ &\leq \sqrt{\sum_{i=1}^{n} (r_i)^2} + \sqrt{\sum_{i=1}^{n} (s_i)^2} \\ &= d(x, z) + d(z, y) \end{aligned} \quad (7)$$

Since both sides of the inequality are positive. By squaring the above, it is equivalent to prove that:

$$\begin{aligned}
\sum_{i=1}^{n}(r_i+s_i)^2 &\leq \sum_{i=1}^{n} r_i^2 + \sum_{i=1}^{n} s_i^2 + 2\sqrt{\sum_{i=1}^{n}(r_i)^2}\sqrt{\sum_{i=1}^{n}(s_i)^2} \\
(\sum_{i=1}^{n} r_i s_i)^2 &\leq (\sum_{i=1}^{n}(s_i)^2)(\sum_{i=1}^{n}(r_i)^2)
\end{aligned} \tag{8}$$

The above could be derived from Cauchy-Schwartz inequality.

Example 2.6. *There is also a Manhattan distance on* $\mathbb{R}^n$*:*

$$d(x,y) = |y_1 - x_1| + \cdots |y_n - x_n| \tag{9}$$

Prove that it is also a metric.

Solution. Same as example 2.4

Example 2.7. *Consider box metric on* $\mathbb{R}^n$*:*

$$d(x,y) = max\{|x_i - y_i|\} \tag{10}$$

Prove that it is also a metric.

Solution. Let $x = (x_1, \cdots, x_n)$, $y = (y_1, \cdots, y_n)$ and $z = (z_1, \cdots, z_n) \in \mathbb{R}^n$. Then, for each $i = 1, \cdots, n$.

$$|x_i - y_i| = |(x_i - z_i) + (z_i - y_i)| \leq |x_i - z_i| + |z_i = y_i| \leq d(x,z) + d(z,y) \tag{11}$$

Example 2.8. *Let* X *be any set. The discrete metric on* X *is defined by*

$$d(x,y) = \begin{cases} 0 & \text{if } x = y, \\ 1 & \text{if } x \neq y, \end{cases} \tag{12}$$

Prove that it is also a metric.

Solution. Let $x, y, z \in X$. If $x = y$, then $d(x,y) = 0$, and there is nothing to check.
Suppose then taht $x \neq y$. Then, either $z = x$ or $z = y$ or $z \neq x, y$. Regardless the situation, we have then

$$1 = d(x,y) \text{ and } 1 \leq d(x,z) + d(z,y) \leq 2 \tag{13}$$

and the triangle inequality hodls.

Example 2.9. *Let d_1, d_2 and d_∞ be the following metrics $\mathbb{R}^2$:*

- $d_1((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|$
- $d_2((x_1, y_1), (x_2, y_2)) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$
- $d_\infty((x_1, y_1), (x_2, y_2)) = max\{|x_1 - x_2|, |y_1 - y_2|\}$

Then, for each $x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$, prove
$\frac{1}{2}d_1((x_1, y_1), (x_2, y_2)) \leq \frac{1}{\sqrt{2}}d_2((x_1, y_1), (x_2, y_2)) \leq d_\infty((x_1, y_1), (x_2, y_2))$

Solution. By definition of the metric d_2,

$$\begin{aligned} d_2((x_1, y_1), (x_2, y_2)) &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \\ &\leq \sqrt{2max\{(x_1 - x_2)^2, (y_1 - y_2)^2\}} \\ &= d_\infty((x_1, y_1), (x_2, y_2)) \end{aligned} \tag{14}$$

For d_1 and d_2, let $a = |x_1 - x_2|$ and $b = |y_1 -_2|$. Then

$$(a + b)^2 \leq 2(a^2 + b^2) \tag{15}$$

which is obviously true.