Metric Spaces in Practice

TA: Nate Clause

Point Clouds

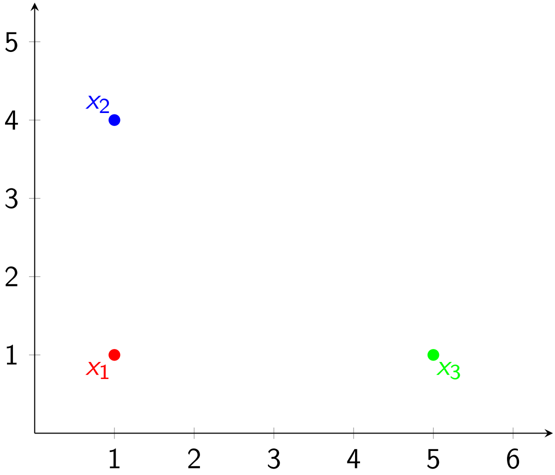
- In TDA, data is often provided in the form of a finite set of points X, called a *point cloud*.
- To use many tools in TDA, we need to convert X into a metric space, (X, d_X).

Point Clouds

- In TDA, data is often provided in the form of a finite set of points X, called a *point cloud*.
- To use many tools in TDA, we need to convert X into a metric space, (X, d_X).
- In many cases, X is viewed as a subset of an ambient metric space, and we restrict the ambient metric to X.
- Ex: $X \subset \mathbb{R}^n$, $d_p : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ the Minkowski metric with $p \in [0, \infty]$, then let $d_X := d_p|_{X \times X}$.

Examples

- Let $X = \{x_1, x_2, x_3\}$, with $x_1 = (1, 1)$, $x_2 = (1, 4)$, $x_3 = (5, 1)$.

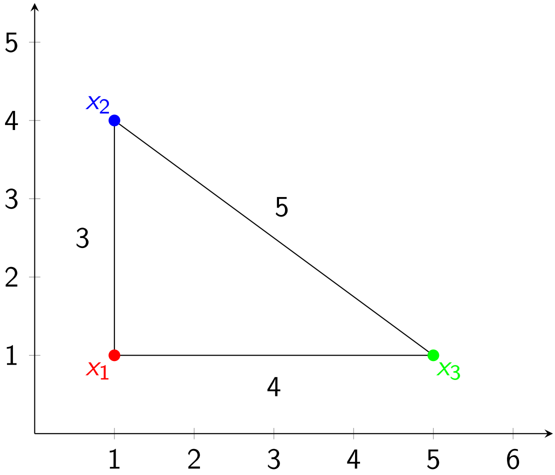


Examples

- If we define $d_X := d_2|_{X \times X}$, we get:

Examples

- If we define $d_X := d_2|_{X \times X}$, we get:



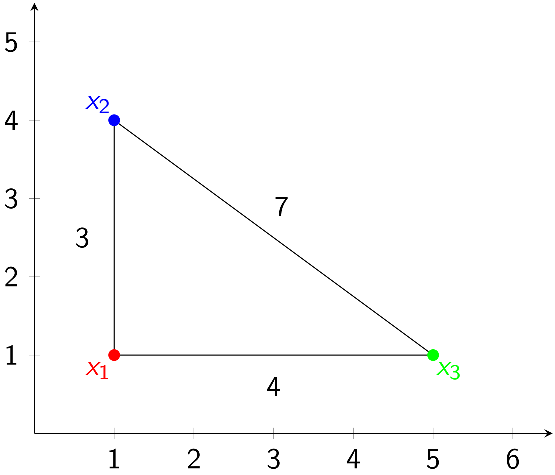
- So $d_X(x_1, x_2) = 3$, $d_X(x_1, x_3) = 4$, $d_X(x_2, x_3) = 5$.

Examples

- If we define $d_X := d_1|_{X \times X}$, we get:

Examples

- If we define $d_X := d_1|_{X \times X}$, we get:



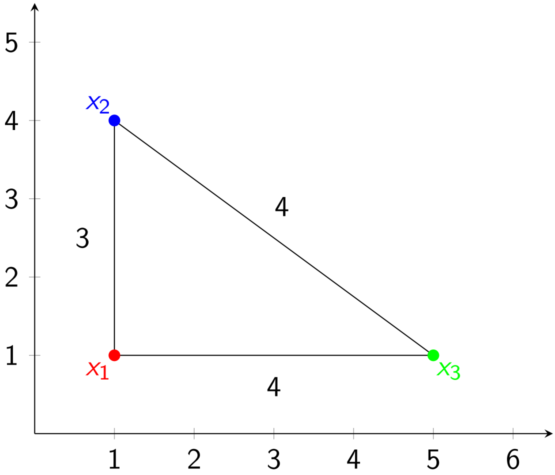
- So $d_X(x_1, x_2) = 3$, $d_X(x_1, x_3) = 4$, $d_X(x_2, x_3) = 7$.

Examples

- If we define $d_X := d_\infty|_{X \times X}$, we get:

Examples

- If we define $d_X := d_\infty|_{X \times X}$, we get:



- So $d_X(x_1, x_2) = 3$, $d_X(x_1, x_3) = 4$, $d_X(x_2, x_3) = 4$.

Distance Matrix

- If (X, d_X) is a metric space with $|X| = n$, we can enumerate the points $X = \{x_1, x_2, \ldots, x_n\}$. The information of d_X can be encoded in a *distance matrix*:

$$D = \left[d_{i,j}\right]_{1 \leq i,j \leq n},$$

where $d_{i,j} := d_X(x_i, x_j)$

Distance Matrix

- If (X, d_X) is a metric space with $|X| = n$, we can enumerate the points $X = \{x_1, x_2, \ldots, x_n\}$. The information of d_X can be encoded in a *distance matrix*:

$$D = \left[d_{i,j}\right]_{1 \leq i,j \leq n},$$

  where $d_{i,j} := d_X(x_i, x_j)$

- From previous examples, we have:

$$D_1 = \begin{bmatrix} 0 & 3 & 4 \\ 3 & 0 & 7 \\ 4 & 7 & 0 \end{bmatrix}, \quad D_2 = \begin{bmatrix} 0 & 3 & 4 \\ 3 & 0 & 5 \\ 4 & 5 & 0 \end{bmatrix}, \quad D_\infty = \begin{bmatrix} 0 & 3 & 4 \\ 3 & 0 & 4 \\ 4 & 4 & 0 \end{bmatrix}$$

Graphs

- Some real-world data is input as a graph or a weighted graph:

Definition

A *weighted graph* is $G = (V, E, w)$, with a set V of *vertices*, *edges* $E \subseteq V \times V$, and *weights* $w : E \to \mathbb{R}_{>0}$.

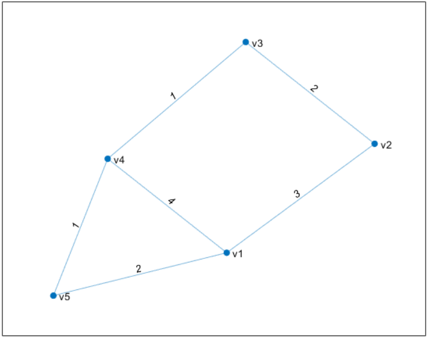
Graphs

- Some real-world data is input as a graph or a weighted graph:

Definition

A *weighted graph* is $G = (V, E, w)$, with a set V of *vertices*, *edges* $E \subseteq V \times V$, and *weights* $w : E \to \mathbb{R}_{>0}$.

- Example:



Weighted Graph to Metric Space

Definition

Let $G = (V, E, w)$ be a weighted graph. A *path* p, between vertices $u, v \in V$, denoted $p : u \to v$, is a sequence of vertices $\{u = v_0, v_1, \ldots, v_n = v\}$ of V such that for all $0 \leq i \leq n-1$, $(v_i, v_{i+1}) \in E$.

Weighted Graph to Metric Space

Definition

Let $G = (V, E, w)$ be a weighted graph. A *path* p, between vertices $u, v \in V$, denoted $p : u \to v$, is a sequence of vertices $\{u = v_0, v_1, \ldots, v_n = v\}$ of V such that for all $0 \leq i \leq n-1$, $(v_i, v_{i+1}) \in E$.

Definition

The *cost* of a path p is:

$$c(p) := \sum_{i=0}^{n-1} w((v_i, v_{i+1}))$$

If $G = (V, E)$ is an unweighted graph, we can view it as a weighted graph by setting $w((u, v)) := 1$ for all $(u, v) \in E$.

Weighted Graph to Metric Space

Definition

Define $d_V : V \times V \to \mathbb{R}$ as:

$$d_V(u, v) := \inf_{p: u \to v} c(p)$$

Weighted Graph to Metric Space

Definition

Define $d_V : V \times V \to \mathbb{R}$ as:

$$d_V(u, v) := \inf_{p:u \to v} c(p)$$

- For $v \in V$, (v, v) may not be in E. By convention, we assume there is a singleton path $p : v \to v$, with $p = \{v\}$ and $c(p) := 0$.

Weighted Graph to Metric Space

Definition

Define $d_V : V \times V \to \mathbb{R}$ as:

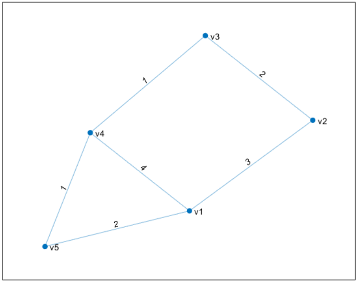
$$d_V(u, v) := \inf_{p:u \to v} c(p)$$

- For $v \in V$, (v, v) may not be in E. By convention, we assume there is a singleton path $p : v \to v$, with $p = \{v\}$ and $c(p) := 0$.
- Using this convention, we have the following:

Proposition

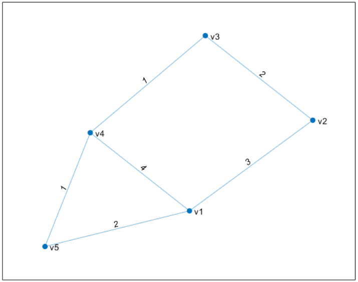
(V, d_V) is a metric space. d_V is often called the *shortest path distance*.

Example



- We compute d_V as the distance matrix D:

Example



- We compute d_V as the distance matrix D:

$$D = \begin{bmatrix} 0 & 3 & 4 & 3 & 2 \\ 3 & 0 & 2 & 3 & 4 \\ 4 & 2 & 0 & 1 & 2 \\ 3 & 3 & 1 & 0 & 1 \\ 2 & 4 & 2 & 1 & 0 \end{bmatrix}$$

MATLAB Examples