**Metric Spaces**

- **Metric space**
  \((X, d)\), \(d \colon X \times X \to \mathbb{R}^+ \cup \{0\}\)
  
  - Symmetry: \(d(p, q) = 0 \iff p = q\)
  
  - Triangle inequality: \(d(p, q) + d(q, r) \geq d(p, r)\)

- **Metric ball**
  
  \(B_o(\text{center}, \text{radius}) := \{ x \in X \mid d(x, c) < r \}\)

- **Metric space topology**
  
  \((X, d) \Rightarrow (X, T_d)\) where \(T_d = \{ B_o(c, r) \mid c \in X, 0 < r \}\) \(U\) (union of balls)

---

**Ex:**

1. \(d \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+ \cup \{0\}\)
   
   \((x, y) \mapsto |x - y|\)

   Euclidean metric on \(\mathbb{R}\)

   \(B_o(c, r) = \{ x \in \mathbb{R} \mid |x - c| < r \}\) = \((c - r, c + r)\)

   ![Diagram of Euclidean metric on R](image)

   \((\mathbb{R}, T_d)\) is \(\mathbb{R}\) with the usual topology
2. \( d : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^+ \cup \{0\} \)
\[
\begin{bmatrix}
x_1 \\
\vdots \\
x_n \\
y_1 \\
\vdots \\
y_n
\end{bmatrix} 
\rightarrow \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}
\]

Euclidean metric on \( \mathbb{R}^n \)

(\( \mathbb{R}^n, \tau_d \)) is known as \( \mathbb{R}^n \) with the usual topology.

3. \( X \neq \emptyset \). \( d : X \times X \rightarrow \mathbb{R}^+ \cup \{0\} \)
\[
(\chi, \psi) \rightarrow \begin{cases} 
0, & \text{if } \chi = \psi \\
1, & \text{if } \chi \neq \psi
\end{cases}
\]

Discrete metric

(\( X, \tau_d \)) is \( X \) with the discrete topology.
4. \( d : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^+ \cup \{0\} \)
\[
\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \mapsto \max\left\{ |x_1 - y_1|, |x_2 - y_2| \right\}
\]

5. \( d : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^+ \cup \{0\} \)
\[
\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \mapsto |x_1 - y_1| + |x_2 - y_2|
\]

Exercise: Find \( B_0(c, r) \) symbolically and graphically.

Taxicab Metric

A graph is specified by a set of vertices and a set of edges connecting the vertices. A weighted graph has weights (nonnegative numbers) attached to the edges. More precisely, a weighted graph is a tuple \( G = (V, E, W) \) with vertex set \( V \), edge set \( E \subset V \times V \), and weights \( W : E \rightarrow \mathbb{R}^+ \).

Regarding this graph as undirected and stipulating that there are no edges with non-zero weight from any vertex \( v \) to itself, the graph metric on a weighted graph is a metric on the set of vertices of the graph. The metric is defined so that the distance between vertices \( v \) and \( w \) is the minimal length of a path connecting \( v \) and \( w \):

\[
\partial_G(v, w) = \min_{v, z_0, z_1, \ldots, z_{k-1}, z_k, w | z_i \in V} \left( W(v, z_0) + \sum_{i=0}^{k-1} W(z_i, z_{i+1}) + W(z_k, w) \right).
\]

\[
\partial_G(A, B) = \min \left\{ 5 + 2 + 9, 5 + 5 + 7 \right\} = 16
\]
Hamming metric

Fix an alphabet \( \Sigma \), i.e., a set of symbols we will call letters. Let \( x \) and \( y \) be words of length \( n \) with letters in \( \Sigma \). Then the Hamming distance between \( x \) and \( y \) is defined to be the number of positions at which the letters of \( x \) and \( y \) differ:

\[
\partial_H(x, y) = \# \{ i \mid x_i \neq y_i \}.
\]

Consider the collection of possible bases in a DNA strand: \( \Sigma = \{A, C, G, T\} \)

\[
\in \Sigma^4, \quad \partial_H(\text{ACGT}, \text{ACAA}) = \#(\{3, 4\}) = 2
\]

At a very high level, the genome can be understood as a long word whose letters are the four nucleotide bases, denoted (A, C, G, and T or U, in the case of RNA). The length of this word varies dramatically across different organisms. The shortest, called viroids, are a few hundred bases. Humans have roughly three billion bases. And plant genomes can be two orders of magnitude larger (e.g., the genome of \textit{Paris japonica}, a rare and beautiful plant from alpine regions in Japan, has a genome of 150 billion bases). The situation is further complicated by the fact that in multicellular organisms, such as humans, different cells will have similar but not necessarily identical genomes. Mathematically, different organisms can be regarded as producing distinct points in the genotype space, and so can different cells from a single organism.
Minimum distance between a point and a set: $d(x, A) = \inf \left\{ d(x, a) : a \in A \right\}$, $A \neq \emptyset$.

Minimum distance between two sets: $d(A, B) = \inf \left\{ d(a, b) : a \in A, b \in B \right\}$, $A, B \neq \emptyset$.

Limit point: $(X, d)$, $A \subseteq X$, $p \in X$.
- $p$ is a limit point of $A$ if $\forall \varepsilon > 0$, $\exists \alpha \in A$, $a \neq p$ and $d(p, a) < \varepsilon$.

- $(X, \tau)$, $A \subseteq X$, $p \in X$.
- $p$ is a limit point of $A$ if $\forall U \in \tau$, $p \in U$, $U \cap A \neq \emptyset$.

Diameter of a set: $\text{diam}(A) = \sup \left\{ d(a, b) : a, b \in A \right\}$, least upper bound.
- **Closure, Interior, Boundary, Compactness** (in terms of limit points)

  closed & bounded

  closer to the intuition of finiteness

- **(Euclidean) (n-1)-sphere**

\[
S(c, r) = \left\{ x \in \mathbb{R}^n : (x_1 - c_1)^2 + \cdots + (x_n - c_n)^2 = r^2 \right\}
\]

indicates the dimension

\[
B^n
\]
\[
S^{n-1}
\]
\[
B^*_n
\]
\[
H^n
\]