AMAT 342 Lec 23, 11/19/19

Exam review

Edit distance on DNA sequences

Let S denote the set of finite strings of the letters $\{A, T, C, G\}$.

e.g. $ATTGG \in S$, $CTAGG \in S$

An elementary operations on S:

- Replace a letter, e.g. $TTT \to TAG$
- Insert a letter, e.g. $TTT \to TCTT$
- Remove a letter, e.g. $ATC \to AC$.

The edit distance is the metric d_{edit} on S given by $d_{edit}(X,Y)$ = minimum # of elementary ops needed to transform X into Y.

Exercise: $X = ATAT$ What is $d_{edit}(X,Y)$?
$Y = TATA$

$ATAT \to TAT \to TATA$ $\boxed{\text{Answer is 2.}}$

Metric Spaces + Open Sets

A **metric** on a set S is a function $d: S \times S \to [0,\infty)$ satisfying:

1. $d(x,y)=0$ iff $x=y$,
2. $d(x,y)=d(y,x)$
3. $d(x,z) \leq d(x,y)+d(y,z)$

A **metric space** is a pair (S,d), where S is a set and d is a metric on S.

Example: 3 metrics on $\mathbb{R}^n$:

d_2 [Euclidean distance], $d_2(x,y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}$

d_1 [Manhattan distance] $d_1(x,y)=\sum_{i=1}^{n}|x_i-y_i|$

d_{max} $\qquad d_{max}(x,y)=\max_{1\leq i\leq n}|x_i-y_i|$.

Remark If $M=(S,d)$ is a metric space, we often abuse notation/terminology slightly and conflate M with its underlying set S

Examples: $x \in M$ means $x \in S$,
$U \subset M$ means $U \subset S$,
If $M' = (S', d')$ is a second metric space, a function $f: M \to M'$ is understood to be a function $f: S \to S'$.

Subspaces of metric spaces

For $M = (S, d)$ a metric space and $S' \subset M$, we can define a metric d' on S' by $d'(x,y) = d(x,y) \; \forall \; x,y \in S'$.

In this way, we can regard any $S \subset M$ as a metric space

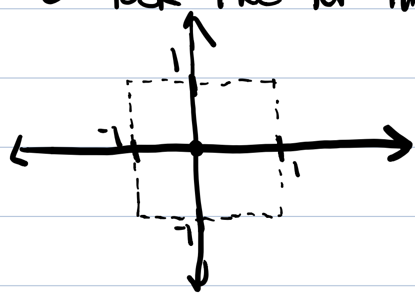
Open Sets in a metric Space

Let M be a metric space w/ metric d_M.
For $x \in M$ and $r > 0$, define

$$B(x,r) = \{ y \in M \mid d_M(x,y) < r \}$$

$B(x,r)$ is called the open ball of radius r centered at x, or simply an open ball.

Exercise: What does the open ball of radius 1 centered at 0 look like for the metric d_{max} on $\mathbb{R}^2$?



Unions (Possibly infinite) Let A be any set, and suppose that for each $x \in A$, we have a set S_x.

The union of the sets $\{S_x\}_{x\in A}$, denoted $\bigcup_{x\in A} S_x$ is the set containing an element y iff $y \in S_x$ for some $x \in A$.

Example: $A = \{1,2,3\}$. $S_1 = \{a,b\}$, $S_2 = \{b,c\}$, $S_3 = \{c,d\}$.

Then $\bigcup_{x\in A} S_x = \bigcup_{x\in\{1,2,3\}} S_x = S_1 \cup S_2 \cup S_3 =$

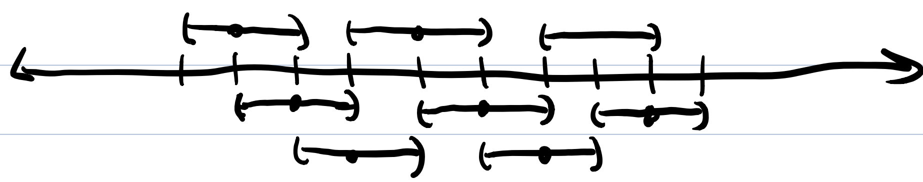
$\{a,b\} \cup \{b,c\} \cup \{c,d\} = \{a,b,c,d\}$.

Definition: For M a metric space, an open set in M is a union of open balls.

More formally, $U \subset M$ is open iff there exists a set $A \subset M$ and a number $r_x > 0$ for each $x \in A$ such that $U = \bigcup_{x\in A} B(x, r_x)$.

Example: Let's write $\mathbb{R}$ as a union of balls in a couple of ways ways:

$$\mathbb{R} = \bigcup_{x \in \mathbb{Z}} B(x,1) \qquad (\text{here } \Lambda = \mathbb{Z})$$



$$\mathbb{R} = \bigcup_{x \in \mathbb{R}} B(x,1)$$

Proposition: If U is an open set of M, then for each $x \in U$, $B(x,r) \subset U$ for some $r > 0$.

Boundaries

Let U be a subset of a metric space M. $x \in M$ is a <u>boundary point</u> if every open ball centered at x contains at least one point in U and one point not in U.

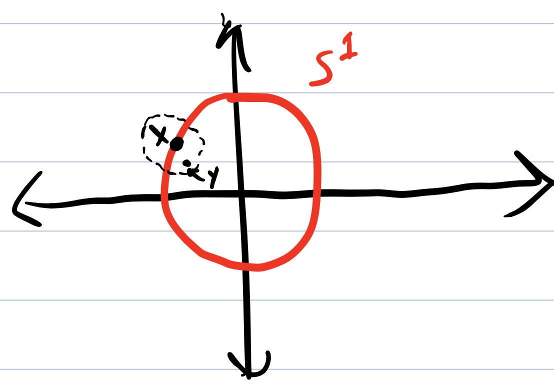
Proposition: For any metric space M, $U \subset M$ is open iff U contains none of its boundary points.

Def: The boundary of U is the set of all of its boundary points.

HW 6 #1

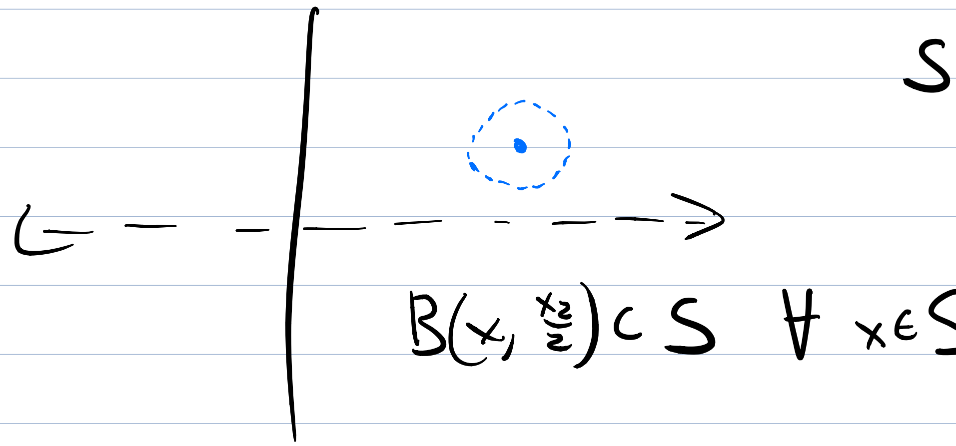
For each of the following sets $S \subset \mathbb{R}^2$, give the boundary of S and state whether S is open.

a. $S = S^1$.



If $x \in S^1$ and $r > 0$, $B(x, r)$ contains $x \in S^1$ and a point $y \notin S^1$. $\Rightarrow$ Boundary$(S^1) = S^1 \Rightarrow$ S^1 is not open.

b. $S = \{(x,y) \mid x > 0\}$



$B(x, \frac{x_2}{2}) \subset S \quad \forall\, x \in S.$