AMAT 584 5/4/20 Last lecture!

Today: Stability of Persistent Homology

The constructions of persistent homology we've considered (Vietoris-Rips, Čech/Delaunay) are stable with respect to perturbations of the data.

The stability of persistent homology is a non-trivial and mathematically interesting theorem. (Or several closely related theorems, to be precise.)

It is arguably the most important result in the persistence theory. It is very useful.

Recall: Persistent homology takes as input a data set and outputs a barcode.

To formulate the stability theorem, one needs to specify

1. a metric (notion of distance) on data sets
2. a metric on barcodes.

Stability then says that data sets which are close have close barcodes (with respect to these metrics).

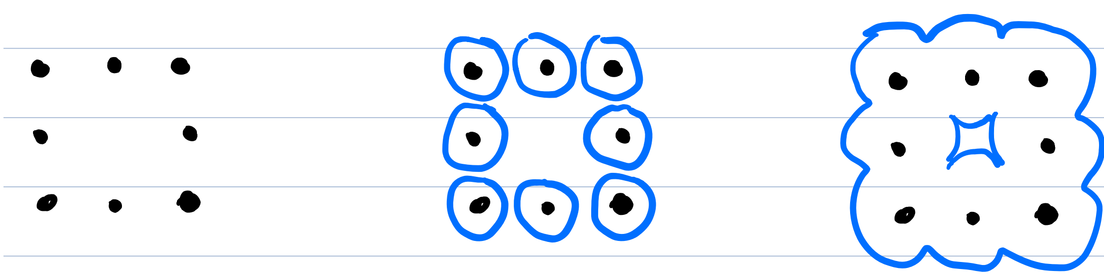
We will now specify these metrics, restricting attention to data in $\mathbb{R}^n$, for simplicity. (But there is also a similar stability theory for finite metric spaces).

Hausdorff Distance

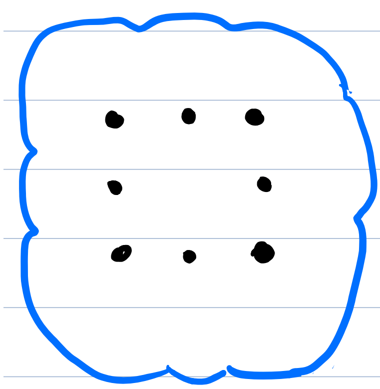
Recall that for $X \subset \mathbb{R}^n$, $U(X,r)$ denotes the union of (closed) balls of radius r centered at the points of X.

We saw this definition lecture 9, where we explained that $U(X,r)$ is a "thickening" of X.

Here is an illustration from that lecture:



$X = U(X,0)$ $\quad$ $U(X,r_1)$ $\quad$ $U(X,r_2)$



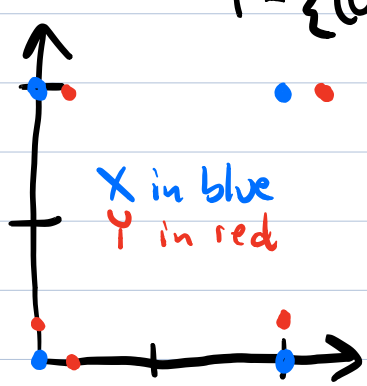
$U(X, r_3)$

For $X, Y \subset \mathbb{R}^n$ finite, we define $d_H(X,Y)$, the Hausdorff distance between X and Y, by

$$d_H(X,Y) = \min\{r \geq 0 \mid X \subset U(Y,r) \text{ and } Y \subset U(X,r)\}$$

(The definition also extends to infite sets, using "inf" instead of "min.")

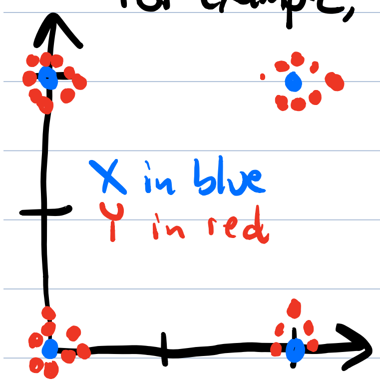
<u>Example:</u> $X = \{(0,0), (2,0), (0,2), (2,2)\}$
$Y = \{(0,\tfrac{1}{4}), (\tfrac{1}{4},0), (2, \tfrac{1}{4}), (\tfrac{1}{4},0), (2\tfrac{1}{4}, 2)\}.$



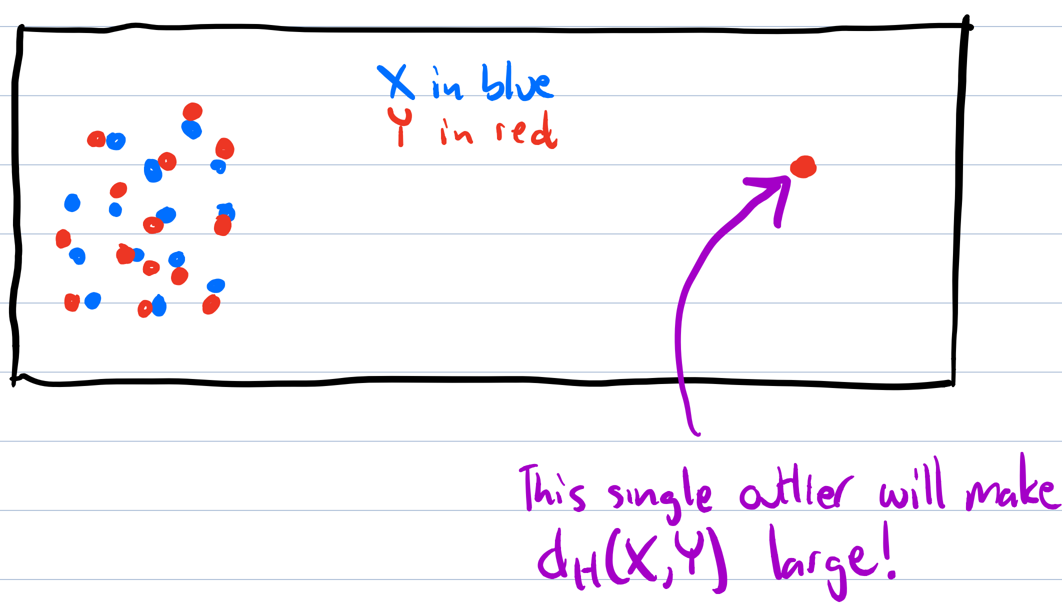
Then $d_H(X,Y) = \frac{1}{4}$, as is easily checked.

Note: $d_H(X,Y)$ can be small even though number of points in X and Y is very different.

For example, these data sets X and Y will have $d_H(X,Y) = \frac{1}{4}$.



However, even one outlying point can make d_H large:



Note: d_H is a metric on finite subsets of $\mathbb{R}^n$

Bottleneck Distance

There are a number of metrics on barcodes one considers in TDA, but for theoretical purposes, the Bottleneck distance is the most popular.

Recall: A barcode is a multiset of intervals in $\mathbb{R}$.

Assume for simplicity that:
1) Each barcode is finite
2) Each interval in each barcode is of the form $[a,b)$, $a,b \in \mathbb{R}$.

For C and D two barcodes, a (partial) matching between C and D is a bijection $\sigma: C' \to D'$, where $C' \subset C$, and $D' \subset D$.

Example: $C = \{[1,5), [1,2)\}$.

$D = \{[1,4)\}$

$C' = \{[1,5)\}$ $\quad\sigma: C' \to D'$, $\sigma([1,5)) = [1,4)$.

$D' = D = \{[1,4)\}$.

For $\sigma: C \to D$ a matching, we define $\mathrm{cost}(\sigma)$ by

$$\mathrm{cost}(\sigma) = \max\left(\begin{array}{l} \left\{ \frac{b-a}{2} \,\middle|\, [a,b) \subset C \cup D \text{ unmatched} \right\} \\ \cup \left\{ \max(|a-a'|, |b-b'|) \,\middle|\, \sigma \text{ matches } [a,b) \text{ to } [a',b') \right\} \end{array} \right).$$

<u>Intuition</u>:
$\mathrm{cost}(\sigma)$ is the maximum amount σ moves any endpoint of an interval in C to transform C into D.

<u>Example</u>: For σ as in the previous example,

$$\mathrm{cost}(\sigma) = \max\left(\frac{2-1}{2}, \max(|1-1|, |5-4|)\right) = \max\left(\tfrac{1}{2}, 1\right) = 1.$$

<u>Definition</u>: For barcodes C and D

$$d_B(C,D) = \min\{\mathrm{cost}(\sigma) \mid \sigma : C \to D \text{ a matching}\}).$$

<u>Example</u>: For C, D as in the previous example, it is easily checked that $d_B(C,D) = 1$.

<u>Theorem [stability]</u>: For $X, Y \subset \mathbb{R}^n$ and $i \geq 1$

$$d_H(X,Y) \geq d_B(\mathrm{Barc}(H_i(VR(X))), \mathrm{Barc}(H_i(VR(Y))))$$

The same is also true using the Čech/Delauney filtrations instead of Rips.

For Čech, this is due originally to Cohen-Steiner, Edelsbrunner, and Harer (2007).

The VR case is due to Chazal et al. (2009).

Looks like that's all we have time for. We'll pick up with many of these ideas in TDA III. Thank you for an enjoyable semester/year! And hope to see you all in person again after this pandemic resolves itself.

-Mike.