TDA Breakout: Day 3 Worksheet

1 Bottleneck Distance Between Persistence Diagrams

Let D and D' be persistence diagrams. Recall that the *bottleneck distance* between them is given by

$$d_b(D, D') = \min_{\phi: A \to A'} \max \left\{ \max_{p \in A} c_m(p, \phi(p)), \max_{p \in D \setminus A} c_u(p), \max_{p' \in D' \setminus A'} c_u(p') \right\},$$

where the minimum is over *partial matchings*; i.e., bijections $\phi : A \to A'$ where $A \subset D$ and $A' \subset D'$. We use the *matching cost* between points $p = (b, d)$ and $p' = (b', d')$ given by

$$c_m(p, p') = \max\{|b - b'|, |d - d'|\}.$$

Another way to write this is

$$c_m(p, p') = \|p - p'\|_\infty,$$

where $\| \cdot \|_\infty$ is the ℓ_∞-*norm*, defined on an aribtrary vector in $\mathbb{R}^2$ by

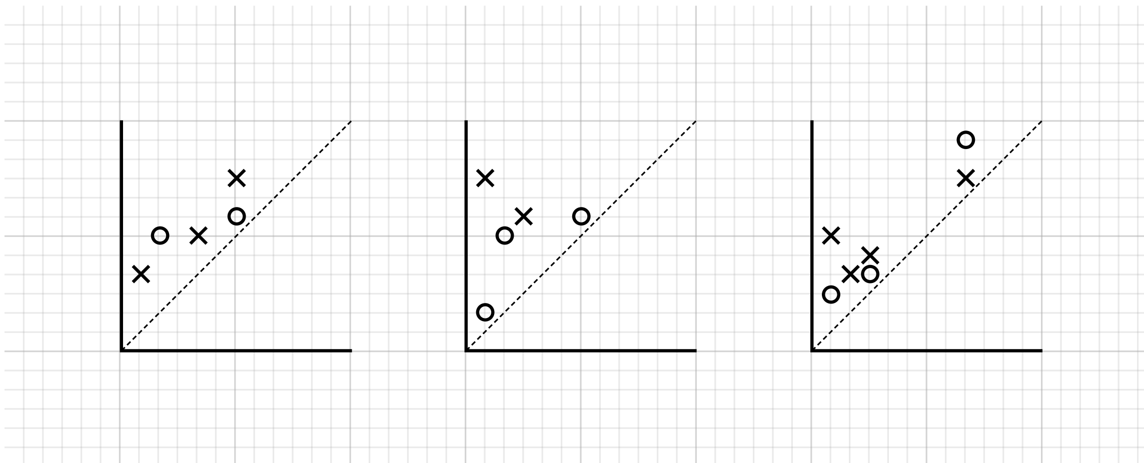
$$\|(x, y)\|_\infty = \max\{|x|, |y|\}.$$

We also use a cost for each unmatched point $p = (b, d)$ given by

$$c_u(p) = \frac{d - b}{2}.$$

Exercise. Show that unmatched cost is ℓ_∞ distance from the point to the diagonal line $y = x$. Can you think of an interpretation of this cost in terms of "matching topological features"?

Exercise. Compute bottleneck distances between the following pairs of persistence diagrams. In each example, D consists of O's and D' consists of X's. Assume the gridlines are 1 unit apart (and that everything is actually centered correctly on the grid).



2 Gromov-Hausdorff Distance

2.1 Metric Spaces

Let (X, d) be a metric space. That is, X is a set, $d : X \times X \to \mathbb{R}_{\geqslant 0}$ is a function with $d(x_1, x_2)$ representing the "distance" between points $x_1, x_2 \in X$. It must satisfy:

- Positivity: $d(x, y) = 0 \Leftrightarrow x = y$ (the distance from x to any other point is positive)
- Symmetry: $d(x, y) = d(y, x)$ (the distance from x to y is the same as the distance from y to x)
- Triangle Inequality: $d(x, z) \leqslant d(x, y) + d(y, z)$ (taking a detour from x to z always increases distance).

If the set $X = \{x_1, \ldots, x_n\}$ is finite, then any metric d can be represented by a *distance matrix*

$$\begin{pmatrix} d(x_1,x_1) & d(x_1,x_2) & \cdots & d(x_1,x_n) \\ d(x_2,x_1) & d(x_2,x_2) & \cdots & d(x_2,x_n) \\ \vdots & \vdots & & \vdots \\ d(x_n,x_1) & d(x_n,x_2) & \cdots & d(x_n,x_n) \end{pmatrix}.$$

Exercise. Explain why any distance matrix will be a symmetric matrix with zeros on the diagonal.

2.2 Hausdorff Distance

Let (Z, d_Z) be a metric space and let $A, B \subset Z$. The *Hausdorff distance* between A and B is

$$d_H^Z(A, B) = \max\{\sup_{a \in A} \inf_{b \in B} d_Z(a, b), \sup_{b \in B} \inf_{a \in A} d_Z(a, b)\}.$$

Exercise. Compute the Hausdorff distance between a (filled in) disk of radius 1 and a disk of radius 2 with the same center in $\mathbb{R}^2$, using Euclidean distance.

2.3 Comparing Finite Metric Spaces

The big question to answer in many applied fields (data science, shape analysis, etc.): how do we quantitatively compare finite metric spaces? That is, *we want a metric on the set of metric spaces*!

A famous answer to the question is given by *Gromov-Hausdorff distance*, d_{GH}. Let (X, d_X) and (Y, d_Y) be finite metric spaces. We compute the Gromov-Hausdorff distance between them as follows:

Let $Z = X \sqcup Y$ (the disjoint union of X and Y). We *extend* the metrics of $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_m\}$ to Z by defining a metric d_Z on Z with distance matrix

$$\left(\begin{array}{cccc|cccc}
0 & d_X(x_1,x_2) & \cdots & d_X(x_1,x_n) & d_{XY}(x_1,y_1) & d_{XY}(x_1,y_2) & \cdots & d_{XY}(x_1,y_m) \\
d_X(x_2,x_1) & 0 & \cdots & d_X(x_2,x_n) & d_{XY}(x_2,y_1) & d_{XY}(x_2,y_2) & \cdots & d_{XY}(x_2,y_m) \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
d_X(x_n,x_1) & d_X(x_n,x_2) & \cdots & 0 & d_{XY}(x_n,y_1) & d_{XY}(x_n,y_2) & \cdots & d_{XY}(x_n,y_m) \\
\hline
d_{XY}(y_1,x_1) & d_{XY}(y_1,x_2) & \cdots & d_{XY}(y_1,x_n) & 0 & d_Y(y_1,y_2) & \cdots & d_Y(y_1,y_n) \\
d_{XY}(y_2,x_1) & d_{XY}(y_2,x_2) & \cdots & d_{XY}(y_2,x_n) & d_Y(y_2,y_1) & 0 & \cdots & d_Y(y_2,y_n) \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
d_{XY}(y_n,x_1) & d_{XY}(y_n,x_2) & \cdots & d_{XY}(y_n,x_n) & d_Y(y_n,y_1) & d_Y(y_n,y_2) & \cdots & 0
\end{array}\right)$$

$$= \left(\begin{array}{c|c}
d_X & d_{XY} \\
\hline
d_{XY}^T & d_Y
\end{array}\right),$$

where we must choose the entries of the matrix d_{XY}. Remember that this matrix must overall define a metric on Z, so there are considerations (namely, the triangle inequality) to make when defining d_{XY}.

Then Gromov-Hausdorff distance is defined by

$$d_{GH}(X,Y) = \min_{d_{XY}} d_H^Z(X,Y)$$

where we take the minimum over all admissible choices of d_{XY}.

Exercise. Show that for any metric space (X, d_X), $d_{GH}(X, X) = 0$.

Exercise. Let (X, d_X) be a one point space $X = \{x\}$ (so $d_X(x, x) = 0$ and the distance matrix for X is just (0)). Let (Y, d_Y) be the metric space with $Y = \{y_1, y_2\}$ and distance matrix

$$d_Y = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Compute $d_{GH}(X, Y)$.

Exercise. The *diameter* of a metric space (Y, d_Y) is

$$\mathrm{diam}(Y) = \max\{d_Y(y, y') \mid y, y' \in Y\}.$$

Show that if X is the one-point metric space and Y is an *arbitrary* finite metric space, then

$$d_{GH}(X, Y) = \frac{1}{2}\mathrm{diam}(Y).$$

Exercise. Let $X = \{x_1, x_2, x_3\}$ with metric given by the matrix

$$d_X = \begin{pmatrix} 0 & \sqrt{2} & 1 \\ \sqrt{2} & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

and let $Y = \{y_1, y_2, y_3\}$ have metric given by the matrix

$$d_Y = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Think about how you might compute $d_{GH}(X, Y)$ (but don't do it unless you really want to). The complexity of the problem, even in this simple case, should start to become apparent.

2.4 Comparing General Metric Spaces

Gromov-Hausdorff distance can be defined without the finiteness requirement. For (compact) metric spaces (X, d_X) and (Y, d_Y), we define

$$d_{GH}(X, Y) = \inf_{Z,\phi,\psi} d_H^Z(\phi(X), \psi(Y)),$$

where the infimum is over *all* metric spaces (Z, d_Z) and all *distance-preserving* maps $\phi : X \to Z$ and $\psi : Y \to Z$ (i.e. $d_Z(\phi(x), \phi(x')) = d_Z(x, x')$).

Exercise. Convince yourself that if X and Y are finite, then this is *the same definition* as what was given above.

Open Problem (!). Compute the Gromov-Hausdorff distance between the interval $[0, 1]$ and the unit circle S^1 (with their natural metrics).

2.5 Gromov-Hausdorff and Bottleneck Distance

You should now be ready to fully appreciate the theorem we stated previously:

Theorem 2.1 (Chazal, Cohen-Steiner, Guibas, Mémoli and Oudot, 2009)**.** *Let (X, d_X) and (Y, d_Y) be finite metric spaces. Let $D_k(X)$ and $D_k(Y)$ be the persistence diagrams for the k-dimensional persistent homology of their Vietoris-Rips complexes. Then*

$$d_b(D_k(X), D_k(Y)) \leqslant d_{GH}(X, Y).$$