Topological Machine Learning: Homework 1
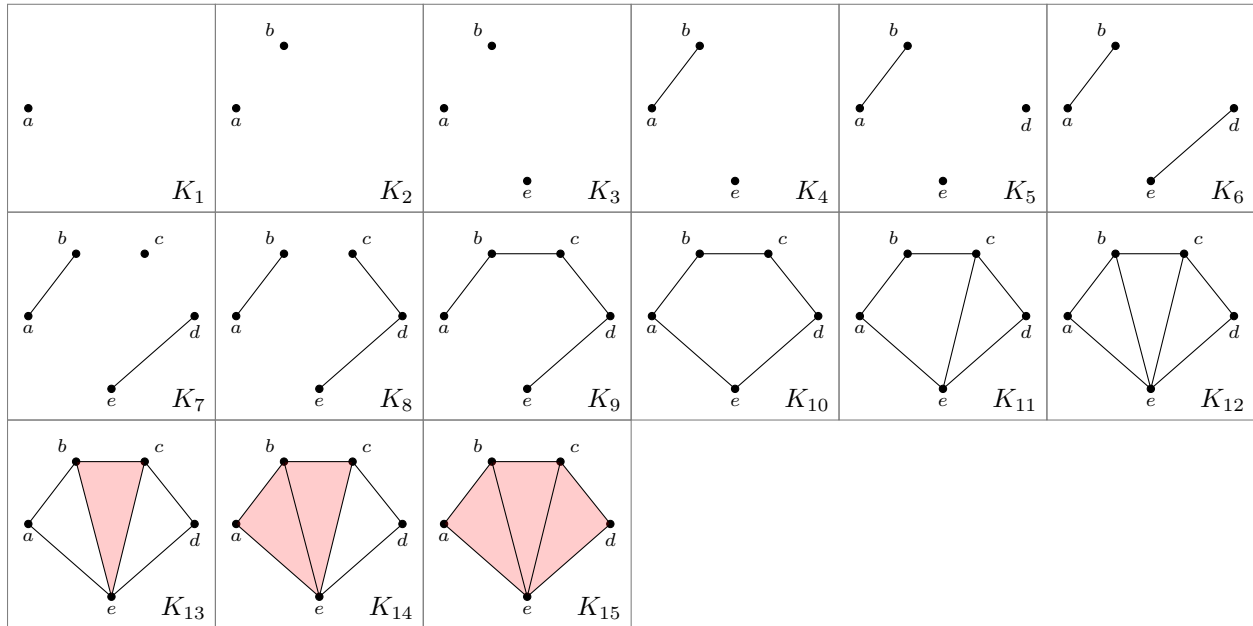
Due on May 16, 2025 at 23:59

Tao Hou

Hashem A. Damrah
UO ID: 952102243

Problem 1. Using the algorithm described in class, compute the PD of the following simplexwise filtration:

Specifically, provide intervals in the PD of each dimension and also for each finite interval (not ending with $+\infty$), provide its representative (aka. its entry in the "ζ" table as demonstrated in class). Notice that a representative (chain) should be denoted as a sum of σ_i's as done in class, where i is the index of the simplex added in the filtration. For example, the chain a $a + b$ should be denoted as $\sigma_1 + \sigma_2$ where $\sigma_1 = a$ and $\sigma_2 = b$.

Remark: There are several ways to quickly check whether your results make sense:

(i) Each index should occur exactly once in an interval, either as the start or the end.

(ii) The representative for an interval $[b, d)$ should be a cycle created in K_b and becoming a boundary in K_d.

Solution. The PD is given by

$$\mathrm{PD}_0(\mathcal{F}) = \{(2, 4), (5, 6), (7, 8), (3, 9)\}$$
$$\mathrm{PD}_1(\mathcal{F}) = \{(12, 13), (11, 14), (10, 15)\}.$$

The $\zeta(i)$ table is given by

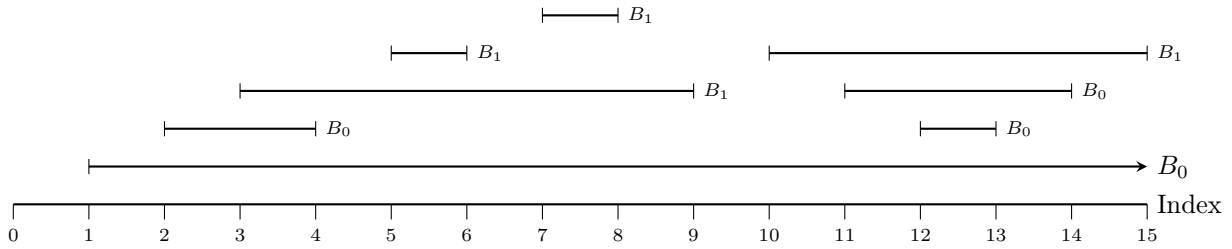| i	$\zeta(i)$
2	$\sigma_1 + \sigma_2$
3	$\sigma_2 + \sigma_3$
5	$\sigma_3 + \sigma_5$
7	$\sigma_5 + \sigma_7$
10	$\sigma_6 + \sigma_8 + \sigma_{10}$
11	$\sigma_4 + \sigma_9 + \sigma_{10} + \sigma_{11}$
12	$\sigma_9 + \sigma_{11} + \sigma_{12}$

Problem 2. Given a filtration $\mathcal{F}$, its *p-th Betti curve* $B_p : I \to \mathbb{N}$ is a map from the indices I of the filtration $\mathcal{F}$ to the natural numbers $\mathbb{N}$ such that $B_p(i)$ is the p-th Betti number of the complex K_i in $\mathcal{F}$. Compute the 0th and 1st Betti curve for the filtration in Question 1. Specifically, fill out a table of the following form:

Index	1	2	$\cdots$	15
B_0				
B_1				

Hint: Recall that there is a way mentioned in class about reading off the Betti numbers for a complex in a filtration using the persistence diagram (barcode).

Solution. Graphing the Betti curves is a good way to visualize the Betti numbers. The 0th Betti number counts the number of connected components, and the 1st Betti number counts the number of holes in the complex. The Betti curves are shown below.

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
B_0	1	2	3	2	3	2	3	2	1	1	1	1	1	1	1
B_1	0	0	0	0	0	0	0	0	0	1	2	3	2	1	0

$\square$

Problem 3. Consider the following non-simplex-wise filtration $\mathcal{F}$ and its corresponding simplexwise filtration $\mathcal{F}'$ where the dashed lines indicate equality of complexes:

$$\mathcal{F} : K_0 \dashrightarrow K_1 \dashrightarrow K_2 \dashrightarrow K_3 \dashrightarrow K_4 \dashrightarrow K_5 \dashrightarrow K_6$$

$$\mathcal{F}' : K_0' \hookrightarrow K_1' \hookrightarrow K_2' \hookrightarrow K_3' \hookrightarrow K_4' \hookrightarrow K_5' \hookrightarrow K_6' \hookrightarrow K_7' \hookrightarrow K_8' \hookrightarrow K_9' \hookrightarrow K_{10}' \hookrightarrow K_{11}' \hookrightarrow K_{12}' \hookrightarrow K_{13}'$$

Convert the following intervals in the PD of $\mathcal{F}'$ to intervals in the PD of $\mathcal{F}$ (if an interval does not have correspondence in $\mathcal{F}$, just say "none"):

(i) $[2, 5)$

(ii) $[7, 13)$

(iii) $[2, 4)$

Solution.

(i) From the diagram, we notice that $K_2' = K_1 \subset \mathcal{F}$ and that K_5' lies strictly before $K_4' = K_2 \subset \mathcal{F}$. Therefore, the feature born at index 2 in $\mathcal{F}'$ is born in K_1 of $\mathcal{F}$, and it dies before reaching K_2. Thus, the corresponding interval in $\mathcal{F}$ is $[1, 2)$.
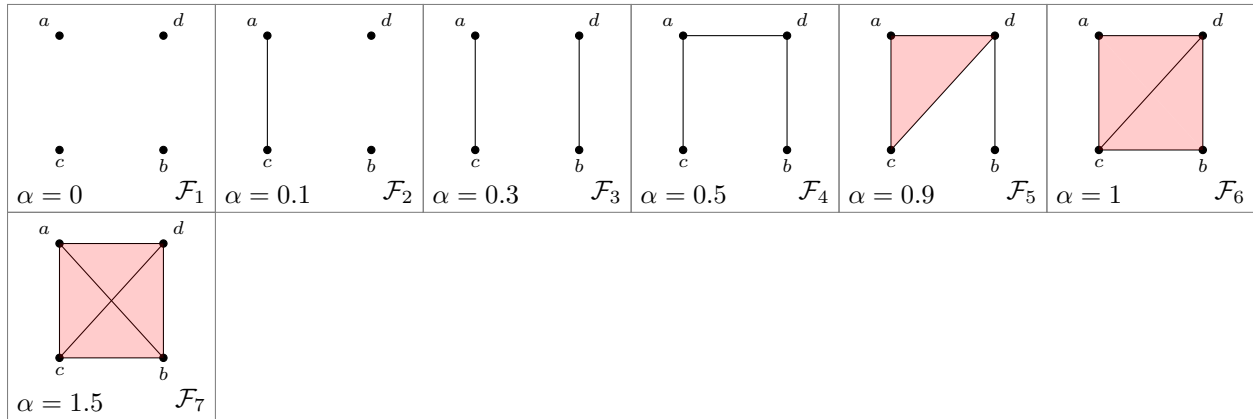

(ii) We notice that $K_6' = K_3$ in $\mathcal{F}$ and K_7' immediately follows it, so K_7' is contained within K_3. Also, we notice that $K_{13}' = K_6$ in $\mathcal{F}$, so, its death occurs just before the inclusion into K_6. Therefore, the feature is born in K_3 and dies before entering K_6, giving the interval $[3, 6)$.

(iii) Here, both K_2' and K_4' lie within K_1 of $\mathcal{F}$. Since the entire birth-death interval lies strictly inside a single complex in $\mathcal{F}$, this feature does not appear in the persistence diagram of $\mathcal{F}$. Therefore, the corresponding interval in $\mathcal{F}$ is "none". $\qquad\square$

Problem 4. Consider four points a, b, c, and d with the following pair-wise distance:

	a	b	c	d
a		1.5	0.1	0.5
b			1.0	0.3
c				0.9
d				

Construct the Vietoris-Rips filtration for the four points. You only need to list the different complexes in the filtration, by denoting each simplex as, say, $abcd$. (Recall that to build VR filtration, you only need to first figure out the order of the edges being introduced, and then you just take the clique complexes once you figure out the edges.)

Solution. The VR filtration is given by the following complexes:

All the complexes from $\alpha = 0$ to $\alpha = 0.5$ are given by

$$\begin{aligned}
\alpha = 0 : & K_1 = \{ab, c, d\} \\
\alpha = 0.1 : & K_2 = \{ab, c, d, ac\} \\
\alpha = 0.3 : & K_3 = \{ab, c, d, ac, bd\} \\
\alpha = 0.5 : & K_4 = \{ab, c, d, ac, bd, ad\}.
\end{aligned}$$

Then, at $\alpha = 0.9$, the triangle acd is complete since we have the edges ac, ad, cd. Therefore,

$$K_5 = \{ab, c, d, ac, bd, ad, acd\}.$$

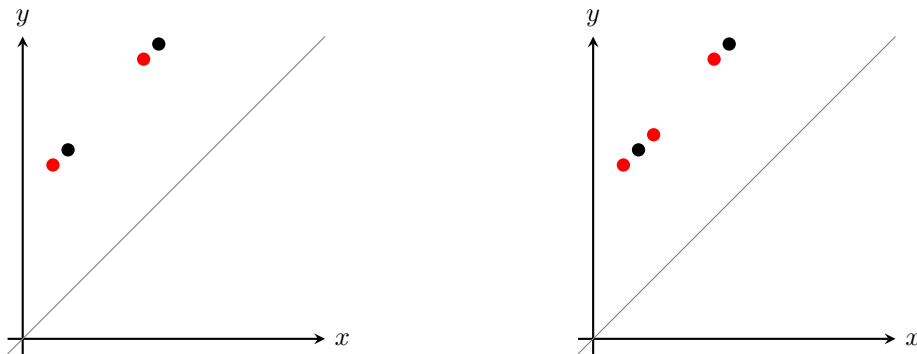At $\alpha = 1$, the triangle bcd is complete since we have the edges bd, cd, bc. Therefore,

$$K_6 = \{ab, c, d, ac, bd, ad, acd, bcd\}.$$

At $\alpha = 1.5$, the triangle abd is complete since we have the edges ab, ad, bd. Also, another triangle is complete, abc, since we have the edges ab, ac, bc. Now the 3-simplex $abcd$ is present, since all 6 edges are in the complex. Therefore,

$$K_7 = \{a, b, c, d, ac, bd, ad, acd, bcd, abd, abc, abcd\}. \qquad \square$$

Problem 5.

Consider the two PDs on the left $\mathbb{R}^2$ plane where one PD contains red points and another PD contains black points. Assume the following:

(i) Each red or black point (interval) has the same length (aka. $d - b$) of 1.0.

(ii) The first red point and the first black point are 0.2 distance away, and the second red point and the second black point are 0.2 distance away.

(iii) The first black point and the second red points are 1.5 distance away.

We then have that, the two PDs (on the left $\mathbb{R}^2$ plane) have a bottleneck distance of 0.2, with the two closest pairs of red and black points perfectly matched.

Now, add another red point whose length is also 1.0 and whose distance to the first black point is 0.2, as indicated in right figure. What is the bottleneck distance of the two PDs in this case? You should also provide brief justifications for your answer.

Solution. In the original persistence diagrams (PDs), each red point can be perfectly matched to a black point with distance 0.2, and the bottleneck distance is 0.2. After adding a third red point in the right figure, there are now three red points and only two black points. Since the bottleneck distance involves computing a matching between the points in the two PDs (allowing unmatched points to be matched to the diagonal), we now must match one red point to the diagonal because there are more red points than black ones.

There are two possible matchings. The two black points are still matched to the two closest red points at distance 0.2 or the extra red point must be matched to the diagonal. Its birth-death pair has length 1.0. In bottleneck distance, we take the maximum over the distances in the optimal matching. In this case, the distances are $\{0.2, 0.2, 1\}$. Thus, the bottleneck distance is the maximum of these, which is 1.

Since we are allowed to match extra points to the diagonal, and this red point is unmatched (there are not enough black points), its best possible match is to the diagonal, which incurs a cost of 1. The optimal matching minimizes the maximum of the matching distances, so the bottleneck distance increases from 0.2 to 1. $\qquad \square$