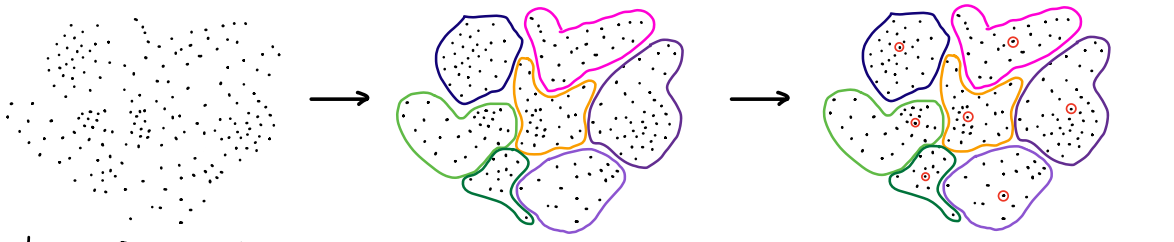
Algorithm

① sample data by k-means



Large Data set!

Partition into K-clusters

These selected points are used as a representative sample of the dataset.

How to maximize representativeness?

Choosing how many clusters: we iterated over different # of clusters and computed the silhouette score. This measures how well each point "fits" in its cluster.

Best Score → optimum # of clusters

② Run Ripser on sampled data

③ Retrieve Cocycles in H^1 → Give them all a persistance score.

φ : Most persistant feature in H^1 → $\max\{b-a : (a,b) \in dgm_1\}$

ψ : 2nd most persistant feature in H^1

④ Restrict $\varphi \to \varphi|_\alpha$

→ only want the edges it has in common psi.

⑤ Compute the cup product $\underbrace{\varphi|_\alpha \smile \psi}$ on all the triangles

$$\varphi|_\alpha \smile \psi(\{x_1, x_2, x_3\})$$

$$= \varphi|_\alpha(\{x_1, x_2\}) \cdot \psi(\{x_2, x_3\})$$

$$\forall \{x_1, x_2, x_3\} \in \text{dgm } 1$$

⑥ Find the boundary matrix δ

Take $\delta^\perp$ to get the coboundary matrix.

⑦ Reduce $\delta^\perp$

$$\left[\delta^\perp\right] \xrightarrow{\text{Reduction Algorithm}} \left[\delta^\perp_{red}\right], \left[V^\perp\right]$$

$$\left[\delta^\perp\right]_{N \times N}$$

(a) Initialize $R = \delta^\perp$, $V = I_{N \times N}$

(b) Find <u>pivot entries</u>: last nonzero entry in a column, if none exist, then inf

and <u>row pivots</u>: the row that the pivot entry is in, if inf, then 0.

(c) Iterate through the columns. Should the current column share a pivot row with any of the columns before it, kill the row entry of the current column. Repeat until it does not share

a row pivot with any previous columns.

* Make sure you do the same column operations to V!

Output: $\left[\delta^{\perp}_{red}\right]$

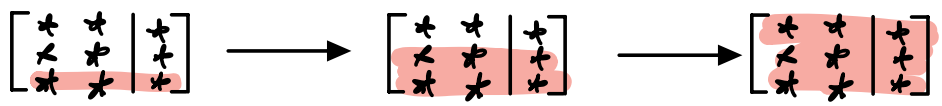
⑧ To solve

$$\underset{\text{Matrix}}{\left[\delta^{\perp}_{red}\right]}\underset{\text{vector}}{\left[X\right]} = \underset{\text{vector}}{\left[\varphi|_{\alpha} \smile \Psi\right]}$$

Take the augmented matrix

$$\left[\delta^{\perp}_{red} \;\middle|\; \varphi|_{\alpha} \smile \Psi\right]$$

Row by Row, we will check when this has no solution.



⑨ Get the triangle associated to the Row where this has no solution.

⑩ Find the cohomological death of $\varphi|_{\alpha} \smile \Psi$.

→ Run Ripser at different values of α until the triangle appears.

The cohomological birth is $\max(\text{birth}\ \varphi|_\alpha, \text{birth}\ \psi)$

(11) plot persistance diagram with $\varphi|_\alpha \cup \psi$.