Non-injectivity via persistence topology

Marek Kaluba

Consider cloud of points in $\mathbb{R}^n$. We would like to know if the points has been sampled from a (local) patch of a manifold. Such sampling usually occurs with some addition of noise, so we may think of points as values of a stochastic function:

$$Y = F(X, N),$$

where under the assumptions:

- $Y|X$ is uni-modal and
- $N$ is (strongly) independent of $X$.

The basic problem approached can be summarised as follows. Given a set of points \( \{(X_i, Y_i)\}_{i=1}^n \subset \mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^{n-k} \) determine when (or how well) a function

$$f: \mathbb{R}^k \to \mathbb{R}^{n-k} \quad \text{or} \quad g: \mathbb{R}^{n-k} \to \mathbb{R}^k$$

could represent the data.

I will present easy 2-dimensional case where the persistence homology of appropriately thresholded Delaunay complex can supply computationally cheap hints to the true answer. Higher dimensional cases are also tractable with Rips complex. These considerations are motivated by applications to the causality problem, where the task to infer the underlying causal structure of two random variables $X, Y$ when only a finite sample from the joint distribution is given. Surprisingly though, the presented approach works for non-functional relations (e.g. in the presence of con-founder).