## Goal

Topological methods in machine learning aim to quantitatively encode shape information from multi-dimensional data points. Validation relies on defining a validation measure to compare topological models. What could be a validation measure relating topological properties of the model and statistical properties of the data for the Mapper [1] and the Generative Simplicial Complex [2,3,4] models?

## Research directions for validation

### Different samples (blue/red) from the same distribution.
But very different Mapper nerves...

### 1) Input data
### 2) Filter function $f$
### 3) Cover $Im(f)$
### 4) Cluster preimages
### 5) Compute nerve

### 1) Input (labeled) data $x \in X$
### 2) Gaussian Mixture Model - Parameters $\theta$ estimated with Expectation Maximization (EM)
### 3) Delaunay complex of GMM centers $\mathcal{W}$
### 4) Gaussian kernel $g(x, w, \theta)$ convoluted to each simplex $\mathcal{W}$ with its own prior weight $\pi_{\sigma}$
### 5) EM: prior weights of generative simplices which do not explain data tend towards 0
### 6) BIC: Simplices with 0 prior get pruned
### 7) Max A Posteriori gives class label for each simplex
### 8) Summary graph/simplex based on connected components in initial and pruned Delaunay complex

Looking for a Post-doc or PhD on these topics, please contact us!