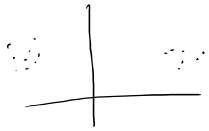
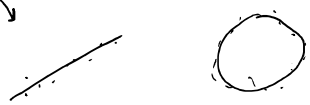
Monday, January 7, 2019 10:24

Goal: Structure in "big" data

eg patient $\longrightarrow \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{pmatrix} \in \mathbb{R}^N$

M patients $\longrightarrow$ M points in $\mathbb{R}^N$.

Similarity metric d

- How to find clusters of similar data pts $\longrightarrow$ 
- How to determine if data lies along a subspace of $\mathbb{R}^N$ 
- How to compare data sets:

M_1 pts in $\mathbb{R}^{N_1}$

M_2 pts in $\mathbb{R}^{M_2}$

eg social network graph

- Clusters
- Comparing social networks

Note: data is noisy, so methods must be robust.

Topology - "large-scale structure & connectivity"

"invariant under continuous deformations"

Algebraic Topology - provides methods to compute topological invariants

eg homology vector spaces

eg H_0 & H_1

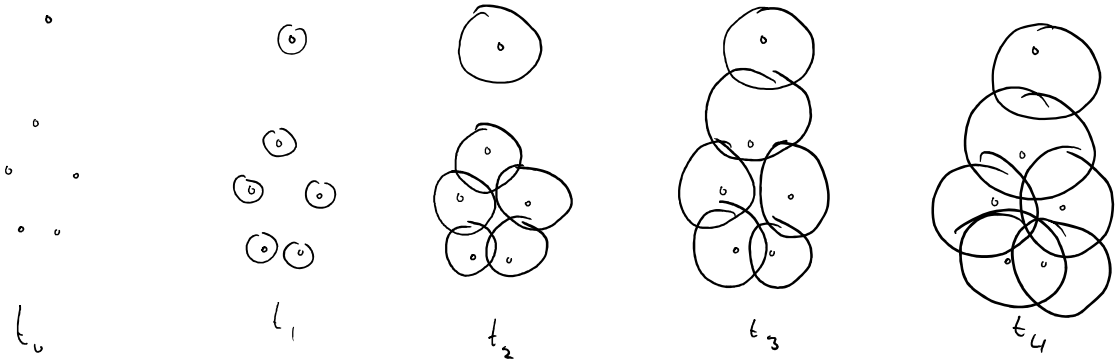
$\mathbb{R}^1$ $\{0\}$

$\mathbb{R}^1$ $\mathbb{R}^1$

$\mathbb{R}^2$ $\{0\}$

Classical topology treats cts shapes, but real data is discrete

Novel idea: ('00s, Edelsbrunner, Carlsson)



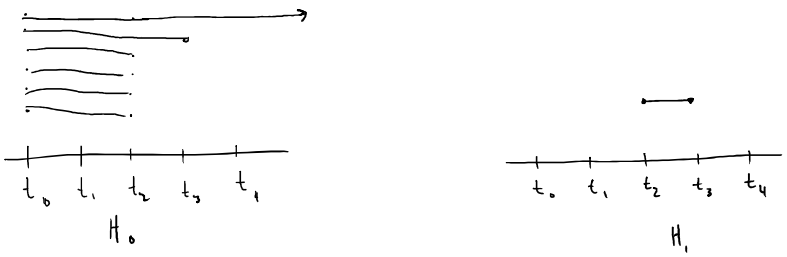
$H_0 = \mathbb{R}^6$, $H_1 = \{0\}$

$H_0 = \mathbb{R}^2$, $H_1 = \mathbb{R}^1$

$H_0 = \mathbb{R}^1$, $H_1 = \mathbb{R}^1$

$H_0 = \mathbb{R}^1$, $H_1 = \{0\}$

Topological Signature called a barcode:



TDA pipeline

Road Map: Linear Algebra

Metric Space topology

Simplicial homology

Persistent homology

Main theorems of TDA

Topics