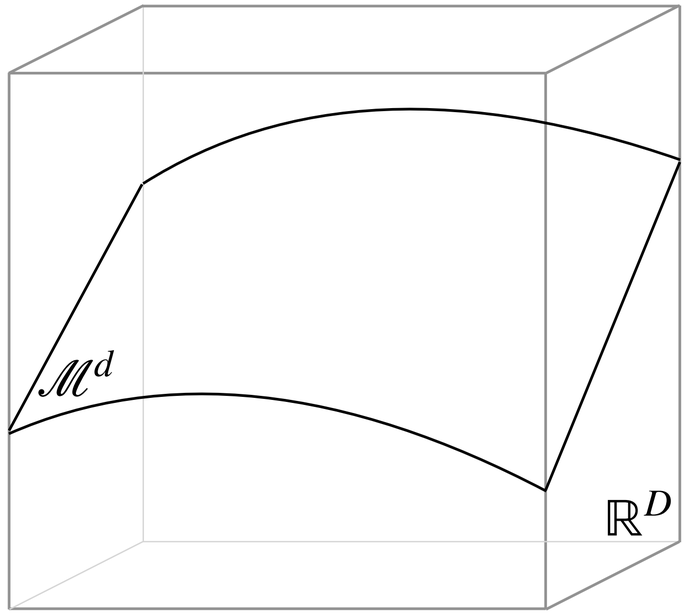
Topological Data Analysis

Lecture 12

Topological Data Analysis for Representation Learning

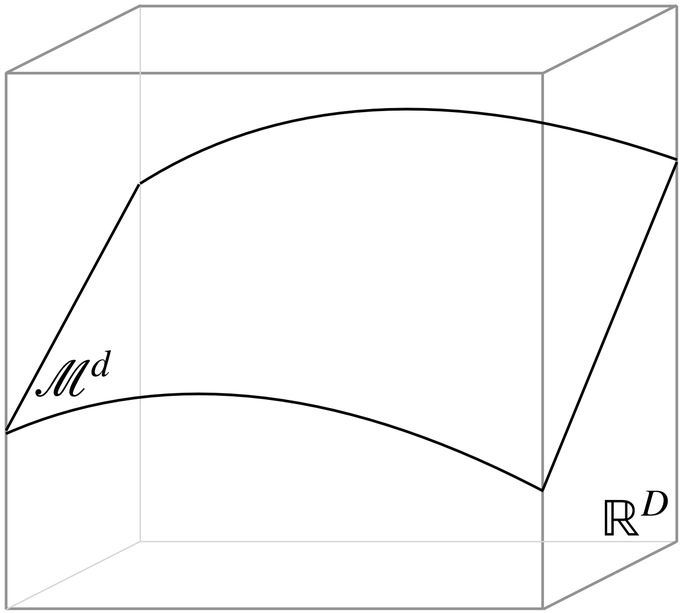
Oleg Kachan

Manifold hypothesis

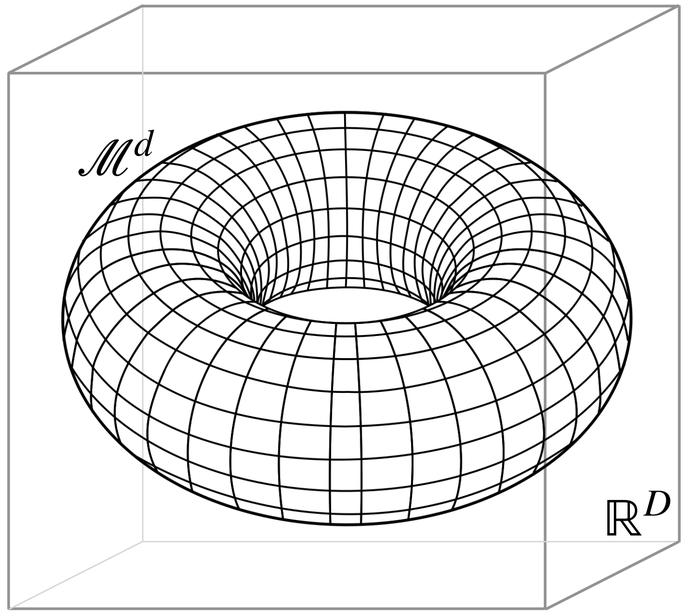


$d \ll D$

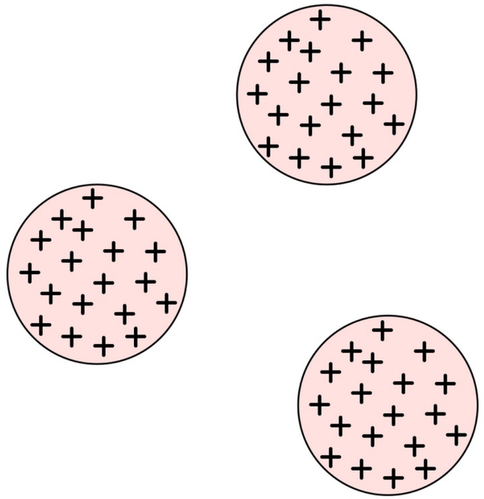
Manifold hypothesis



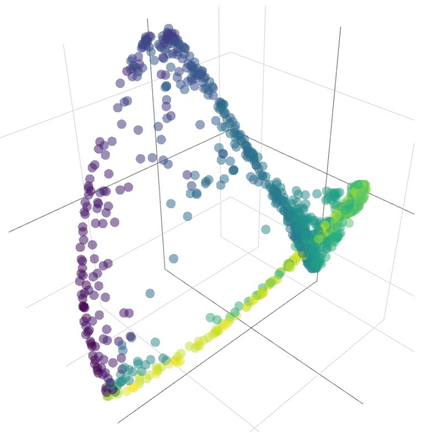
$d \ll D$



Manifold hypothesis

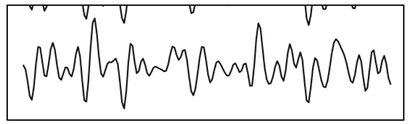


Connected components =
clusters

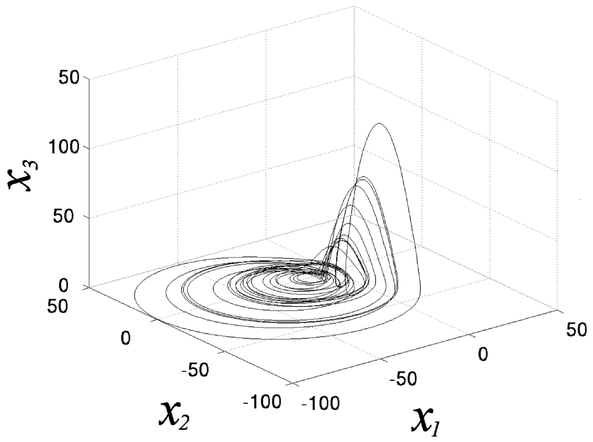


Cycles =
periodicity, regions of lower density

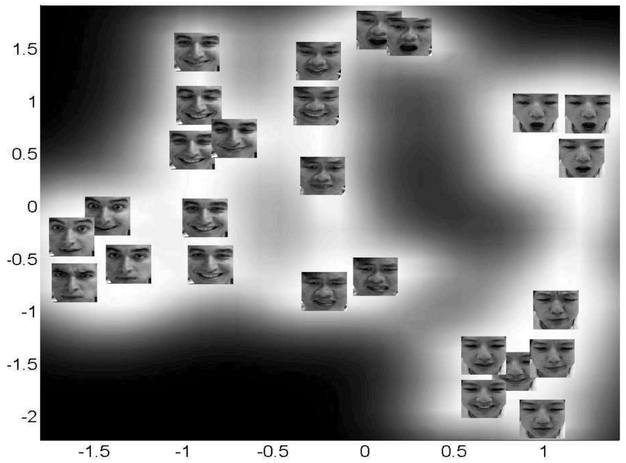
Manifold hypothesis



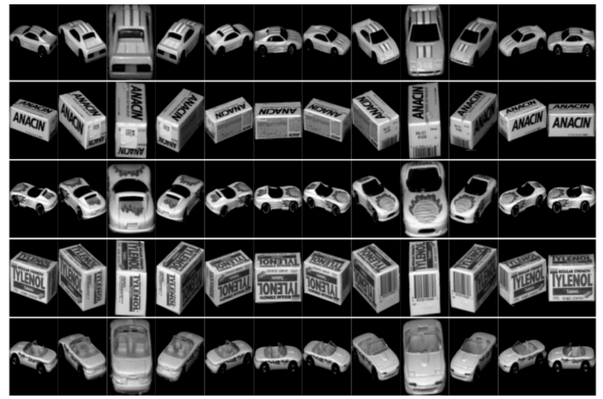
Time series

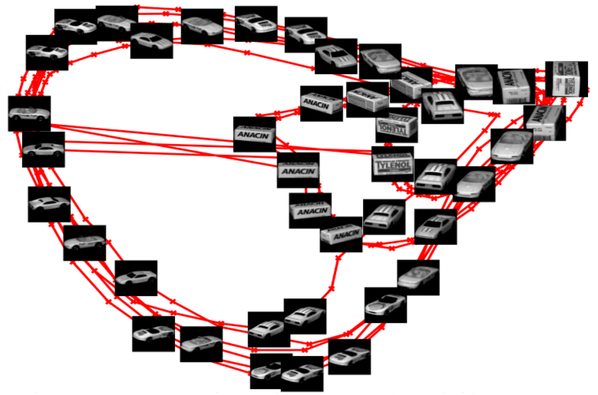


Delay embedding



COIL dataset





Dimensionality reduction

Given a dataset $X \in \mathbb{R}^m$ find a mapping $f : \mathcal{X} \in \mathbb{R}^D \rightarrow \mathcal{Z} \in \mathbb{R}^d$
where d << D, while optimizing/preserving some relevant properties of the data.

$$X \xrightarrow{f} Z$$

Properties

- variance/distances
- statistical independence

TDA for dimensionality reduction

Given a dataset $X \in \mathbb{R}^m$ find a mapping $f : \mathcal{X} \in \mathbb{R}^D \rightarrow \mathcal{Z} \in \mathbb{R}^d$ where d << D, while preserving topology of the input data.

$$X \xrightarrow{f} Z$$

$$Z \cong X$$

Properties

- variance/distances
- statistical independence
- *persistent homology groups*

TDA for representation learning

Given a dataset $X \in \mathbb{R}^m$ and mappings $f : \mathcal{X} \to \mathcal{Z}$ and $g : \mathcal{Z} \to \mathcal{Y}$
find a representation Z, close in topology to the input or output data.

$$X \xrightarrow{f} Z \xrightarrow{g} Y$$

$$Z \cong X \qquad Z \cong Y$$

Comparing data point clouds

Persistence diagram

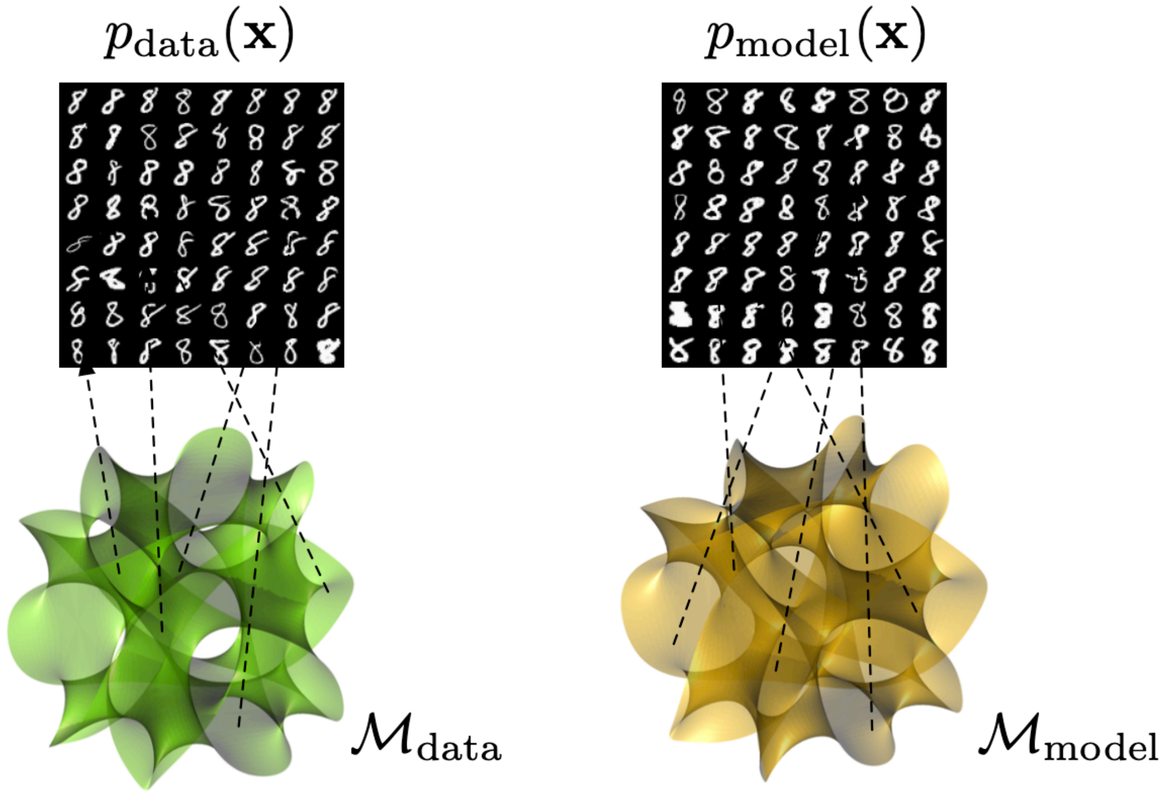
- Wasserstein distance

Persistence diagram statistics

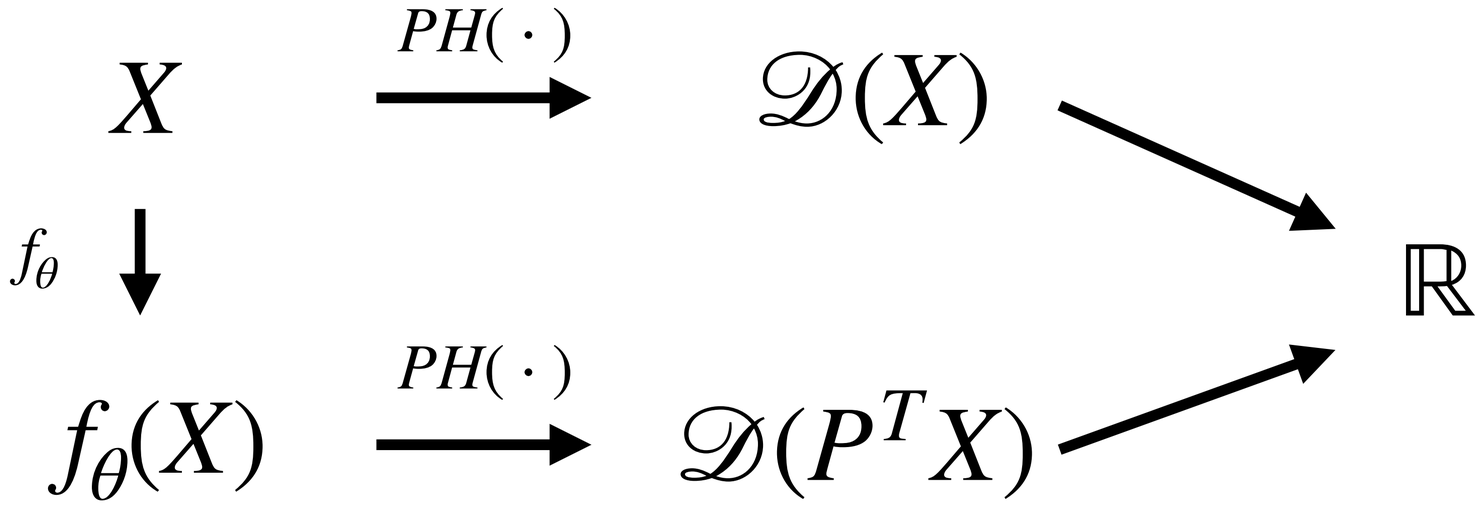
- Geometry Score (GS)

Cross-diagram

- Manifold topology divergence
- Representation topology divergence



Comparing diagrams



Comparing data point clouds

Wasserstein distance

Given a ground metric $d : X \times X \to \mathbb{R}$ optimal transport equips the space of measures $\mathscr{P}(X)$ with a metric referred to as the *p*-th Wasserstein distance

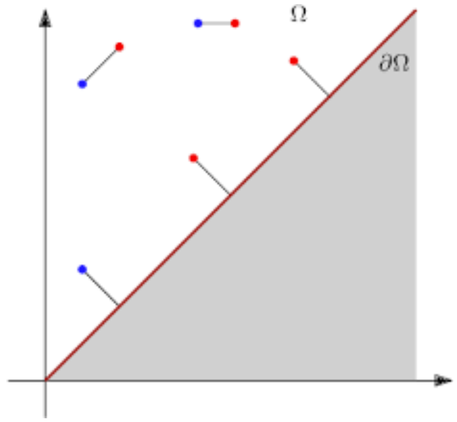
$$W_p^p(\mu, \nu) = \min_{\mathbf{T} \in \Pi(\mathbf{a}, \mathbf{b})} \langle \mathbf{T}, \tilde{\mathbf{M}} \rangle_F$$

$$\Pi(\mathbf{a}, \mathbf{b}) = \{\mathbf{T} \in \mathbb{R}_+^{n \times m} : \mathbf{T}\mathbf{1}_m = \mathbf{a}, \mathbf{T}^T\mathbf{1}_n = \mathbf{b}\}$$

Augmented cost matrix

$$\tilde{\mathbf{M}} = \begin{pmatrix} \mathbf{M} & \Delta_{D(X)} \\ \Delta_{D(Y)} & 0 \end{pmatrix}$$

$$(\mathbf{M})_{ij} = d^p(\mathbf{x}_i, \mathbf{y}_j) \qquad (\Delta_{D(X)})_i = d^p(\mathbf{x}_i, \partial\Omega)$$
$$(\Delta_{D(Y)})_j = d^p(\mathbf{y}_j, \partial\Omega)$$



Approximations

- Sinkhorn divergence, O(n^2)
- Sliced Wasserstein distance, O(*n*log*n*)

Comparing data point clouds

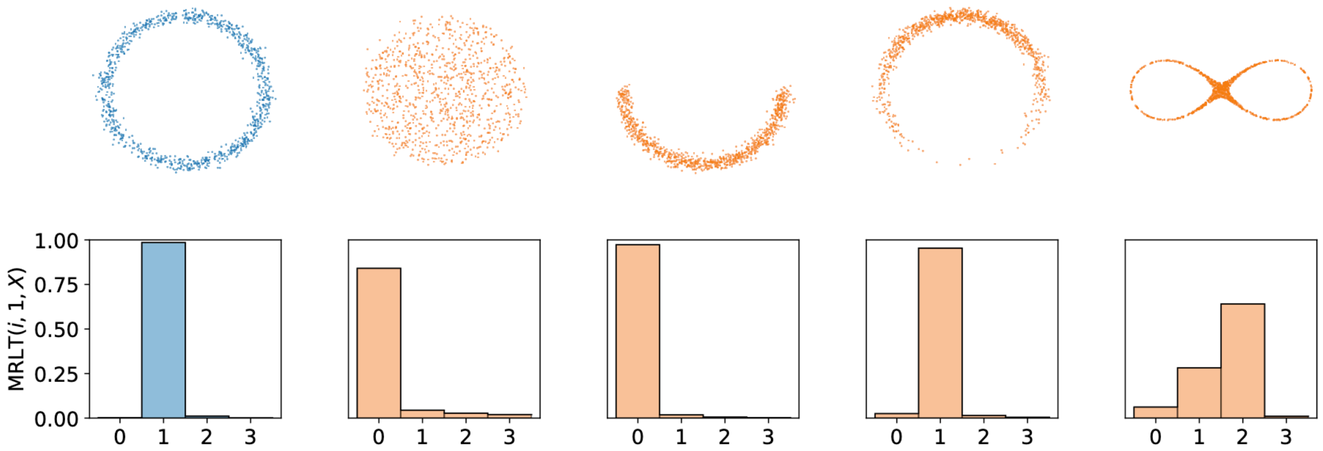
Geometry score

Given a *k*-th persistence diagram $D^k(X)$ the relative living time (RLT) is defined

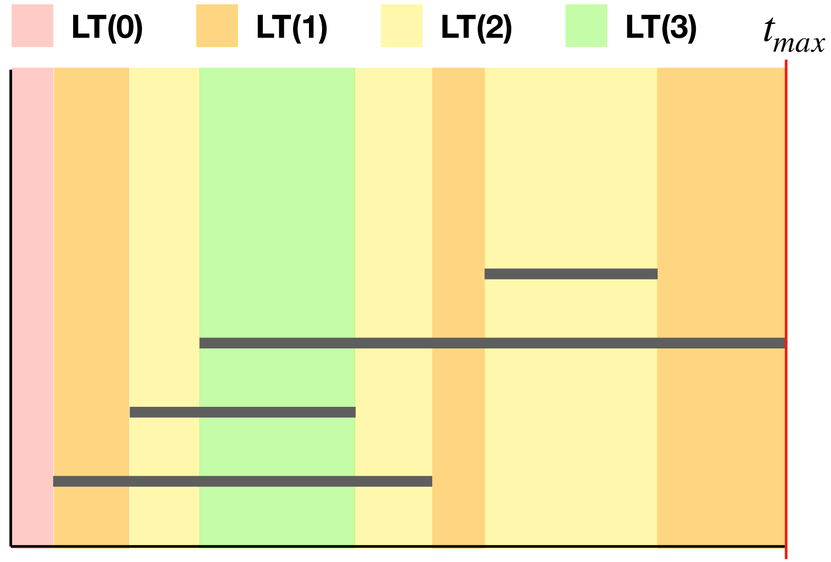
$$\mathrm{RLT}_X^k(n) = \frac{\mu\{t \in [0,\, t_{max}] \mid \beta_k(t) = n\}}{t_{max}}, \quad n \in \mathbb{Z}_{\geq 0}$$

Minimum enclosing radius

$$t_{max} = \min_{x \in X} \max_{x' \in X} d(x, x')$$



RLT[1] distributions over *n* for various point clouds



Relative living times for a given persistence barcode $D^k(X)$

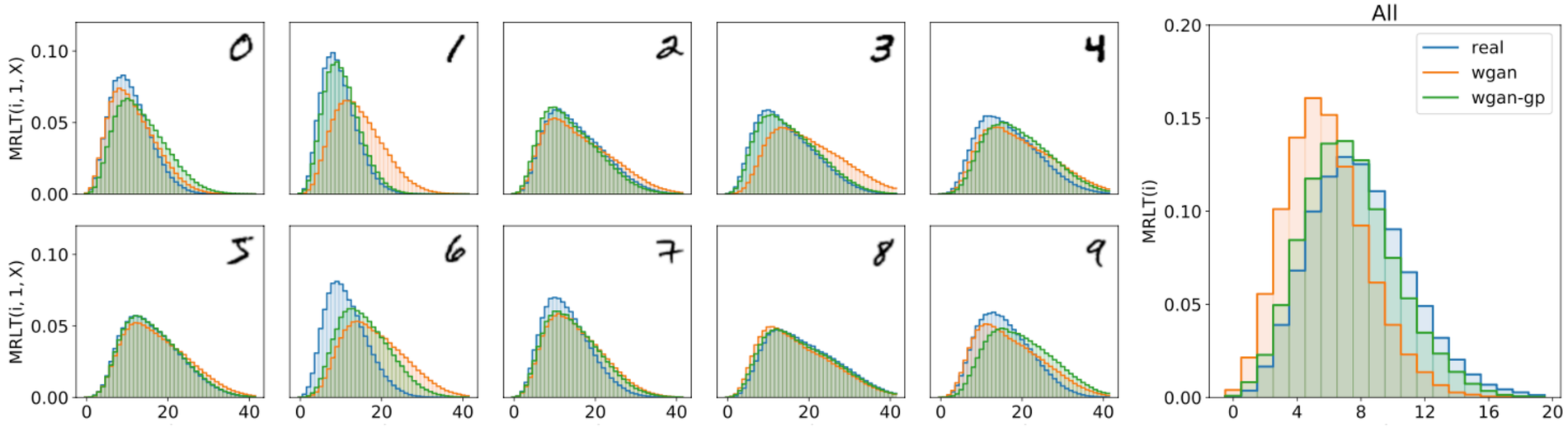
Geometry Score: A Method For Comparing Generative Adversarial Networks
Khrulkov V., Oseledets I. *International Conference on Machine Learning (2019)*

Comparing data point clouds

Geometry score

Given two point clouds X and Y the geometry score $\mathrm{GS}(\mathscr{X}, \mathscr{Y})$ is given by

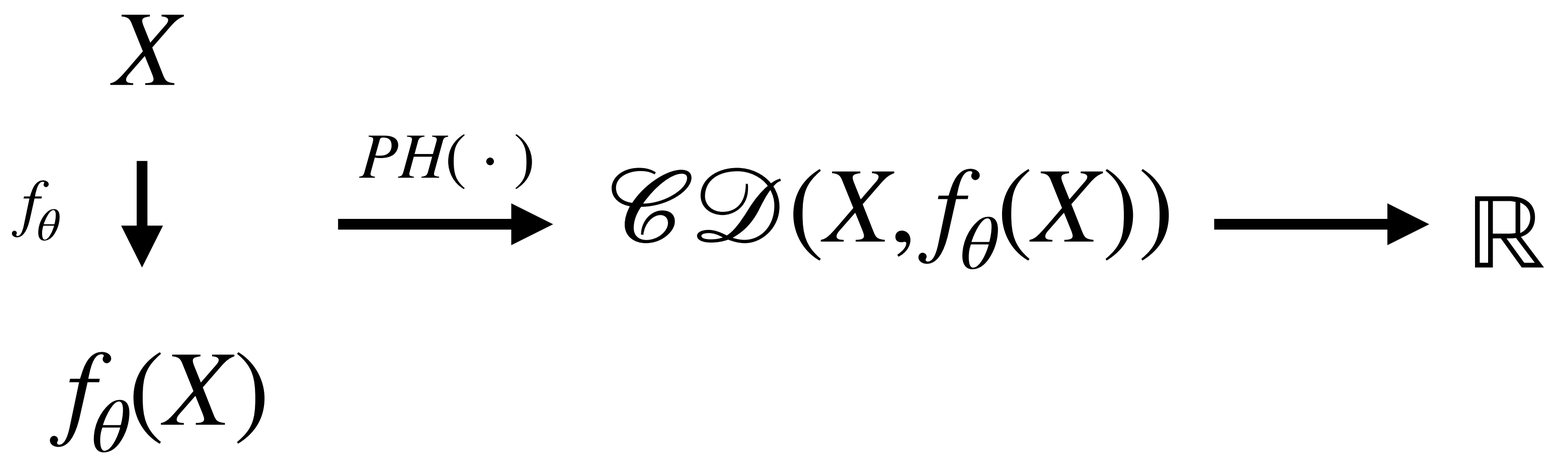
$$\mathrm{GS}(\mathscr{X}, \mathscr{Y}) = \sum_{n}^{n_{max}-1} \left(\mathrm{RLT}^1_X(n) - \mathrm{RLT}^1_Y(n)\right)^2$$



According to the Geometry Score WGAP-GP better fits MNIST dataset than vanilla WGAN

Geometry Score: A Method For Comparing Generative Adversarial Networks
Khrulkov V., Oseledets I. *International Conference on Machine Learning (2019)*

Comparing cross-diagrams



Manifold topology divergence

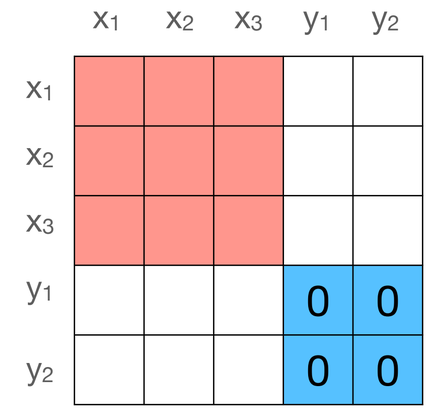
Cross-barcode

Given two point clouds $X, Y \in \mathcal{X}$ the cross-barcode $CB(X, Y)$ is a map to the persistent diagram of the Vietoris-Rips filtration of the union of X and Y quotiented by Y.

$$CB : (X, Y) \to D_{VR}((X \cup Y)/Y)$$

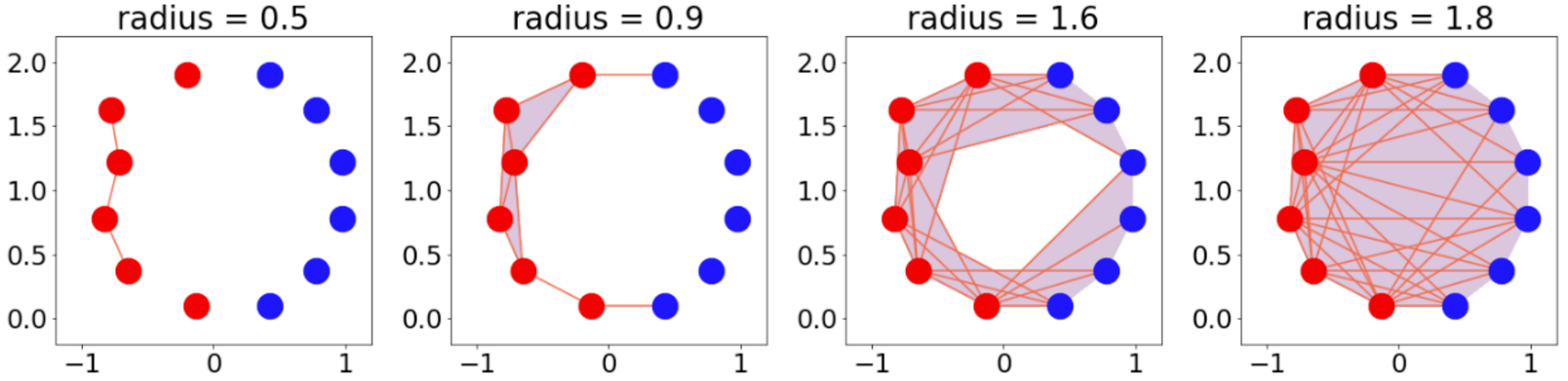
If point clouds X and Y coincide then $CB(X, Y) = \varnothing$.

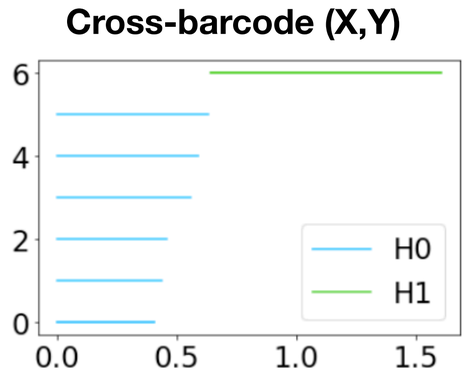
$$\mathbf{A}_{XY} =$$



$a_{ij} = d(x_i, x_j), d(x_i, y_j), d(x_j, y_i)$

$a_{ij} = 0, \quad \forall (y_i, y_j)$





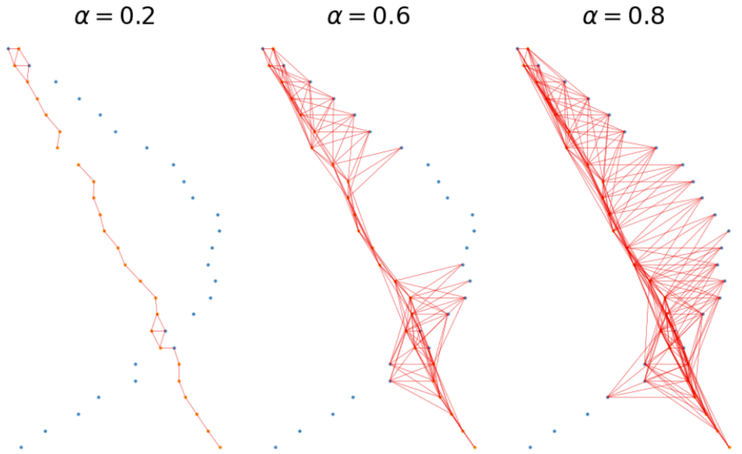
All simplices spanned by **Y** are born at 0 and not shown for perception ease.

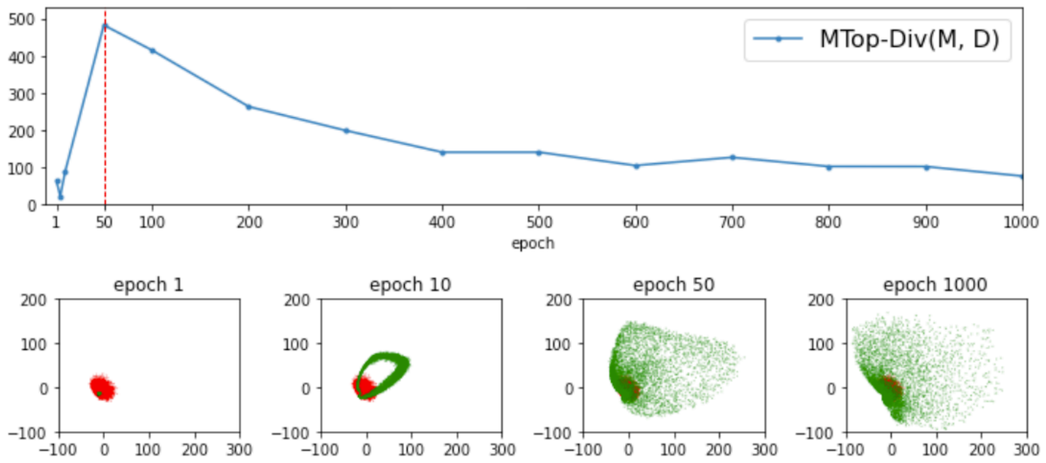
Manifold Topology Divergence: A Framework for Comparing Data Manifolds
Barannikov et al. *Neural Information Processing Systems (2021)*

Manifold topology divergence

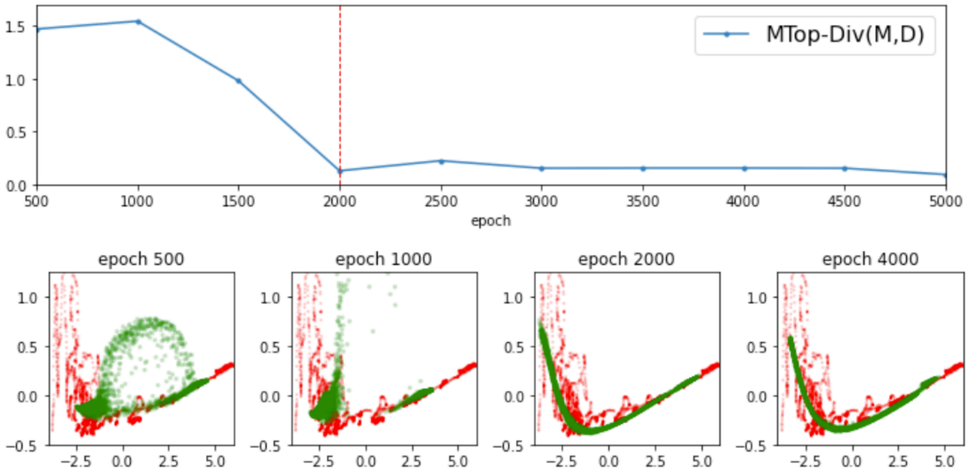
Given two point clouds $X, Y \in \mathcal{X}$ the manifold topology divergence $\mathrm{MTop} - \mathrm{Div}(X, Y)$ is given by the sum of persistences in $CB(X, Y)$

$$\mathrm{MTop} - \mathrm{Div}(X, Y) = \sum_{i}^{|CB(X,Y)|} (d_i - b_i)$$





Training dynamics of GAN applied to 3D shapes



Training dynamics of TimeGAN applied to market stock data

Manifold Topology Divergence: A Framework for Comparing Data Manifolds
Barannikov et al. *Neural Information Processing Systems (2021)*

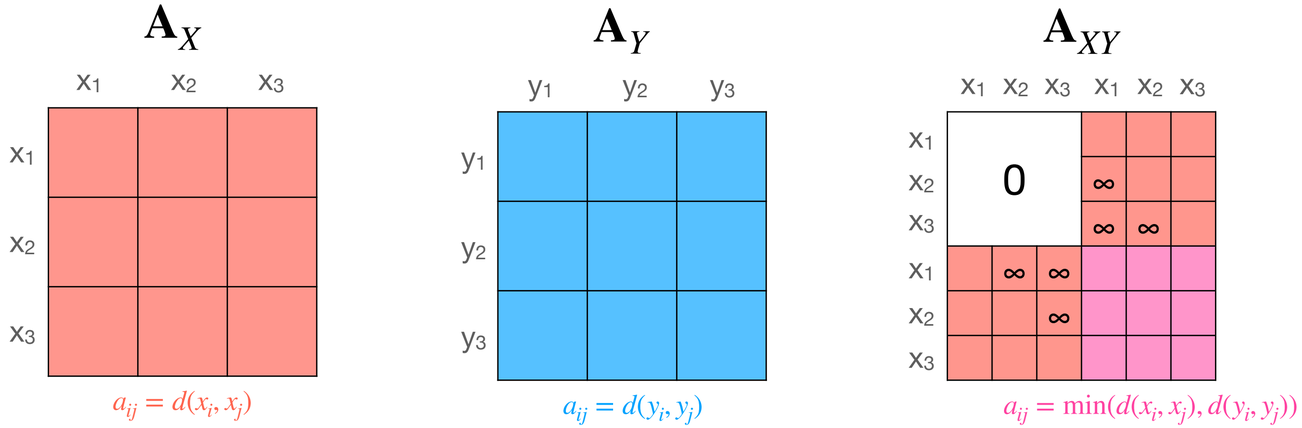
Representation topology divergence

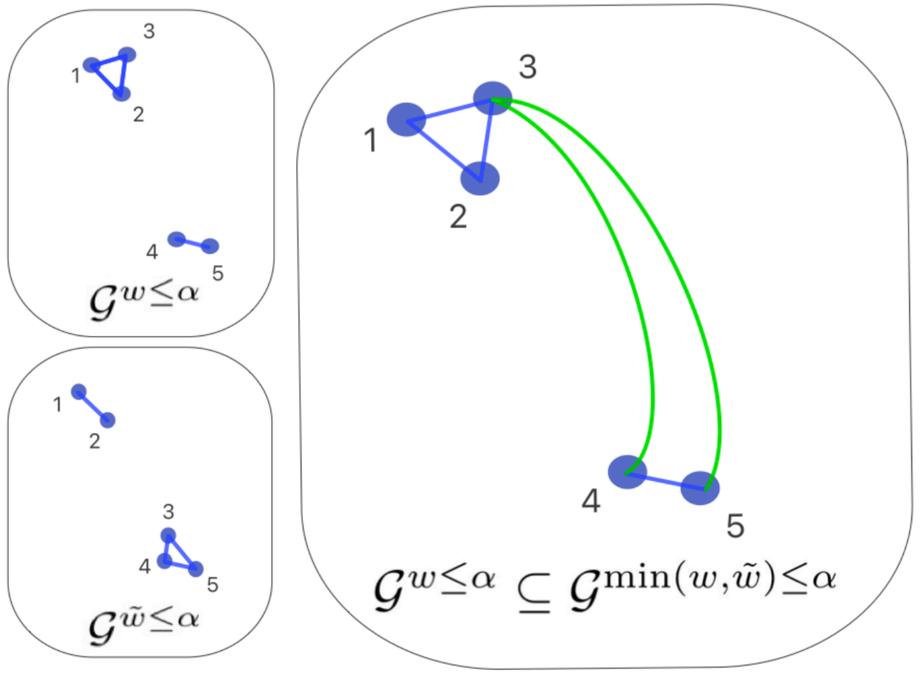
R-cross-barcode

Given two point clouds $X \in \mathcal{X}, Y \in \mathcal{Y}$ with point-to-point correspondence, i.e. two embeddings of the same data, possibly in different ambient spaces, an r-cross-barcode $RCB(X, Y)$

$$RCB : (X, Y) \to D_{VR}((X, \min(X, Y))$$

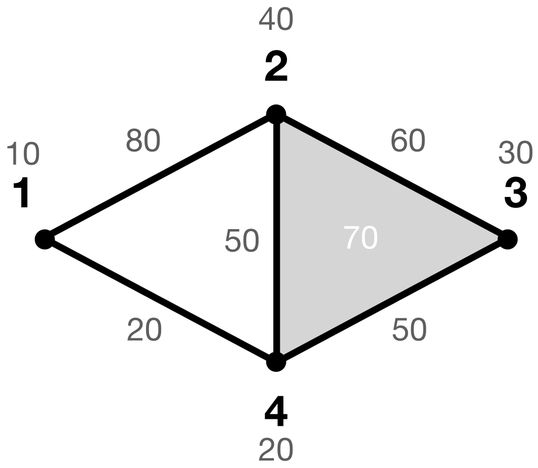
$$\text{RTop} - \text{Div}(X, Y) = \sum_{i}^{|RCB(X,Y)|} (d_i - b_i)$$





Representation Topology Divergence: A Framework for Comparing Data Manifolds
Barannikov et al. *Neural Information Processing Systems (2021)*

Differentiating persistent homology



$\mathbf{R} =$

	10	20	20	30	40	50	50	60	70	80
	1	**4**	**14**	**3**	**2**	**24**	**34**	**23**	**234**	**12**
0										
1			1							
4			**1**			1	1			
14										
3							**1**			
2						**1**				
24									1	
34									1	
23									**1**	
12										

Persistence pairing π_X

(4, 14) 0
(2, 24) 0
(3, 34) 0
(23, 234) 1
(1, $\underline{\emptyset}$) 0
(12, $\underline{\emptyset}$) 1

Persistence diagram D_X

(20, 20) 0
(40, 50) 0
(30, 50) 0
(60, 70) 1
(10, $\underline{\emptyset}$) 0
(80, $\underline{\emptyset}$) 1

Optimizing Persistent Homology-based Functions
Carriere et al. *International Conference on Machine Learning (2021)*

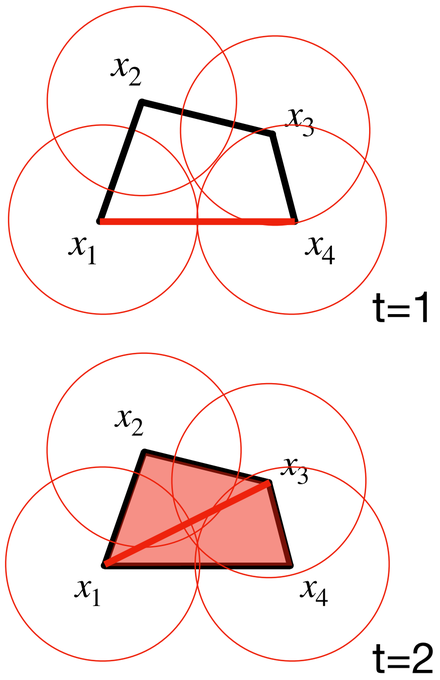
Continuation of Point Clouds via Persistence Diagrams
Gameiro M., Hiraoka Y., Obayashi I. *Physica D: Nonlinear Phenomena (2016)*

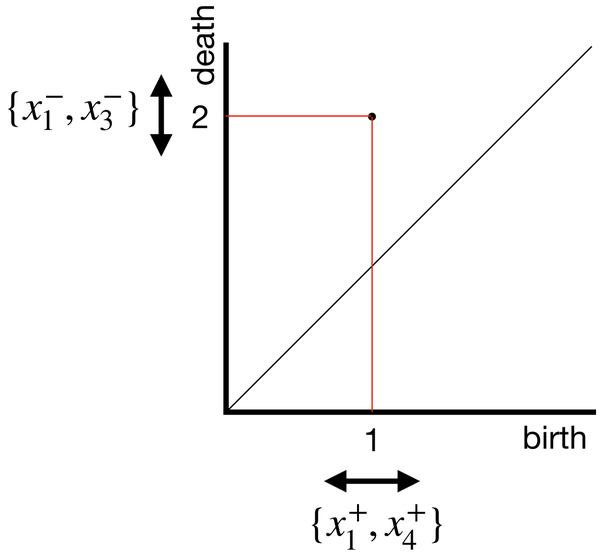
Differentiating persistent homology

Vietoris-Rips filtration

Persistent homology as a map: $PH : (X, f) \rightarrow D_f(X)$

Given a dataset X, the Vietoris-Rips simplicial complex $K_f(X)$ with
the filtration function $f(\sigma) = \max_{(x_i,x_j)\in\sigma} \|x_i - x_j\|_2$ the gradient of point in the persistent diagram





Inverse function

$$\pi(\sigma) = \arg \max_{(x_i,x_j)\in\sigma} \|x_i - x_j\|_2$$

Gradient $\dfrac{\partial D(X)}{\partial X}$

$$\frac{\partial b_k}{\partial x_i^+} = \partial_{x_i^+} f(\sigma) = +\frac{x_i^+ - x_j^+}{\|x_i^+ - x_j^+\|_2}, \quad \frac{\partial b_k}{\partial x_j^+} = \partial_{x_j^+} f(\sigma) = -\frac{x_i^+ - x_j^+}{\|x_i^+ - x_j^+\|_2}$$

$$\frac{\partial d_k}{\partial x_i^-} = \partial_{x_i^-} f(\sigma) = +\frac{x_i^- - x_j^-}{\|x_i^- - x_j^-\|_2}, \quad \frac{\partial d_k}{\partial x_j^-} = \partial_{x_j^-} f(\sigma) = -\frac{x_i^- - x_j^-}{\|x_i^- - x_j^-\|_2}$$

Optimizing Persistent Homology-based Functions
Carriere et al. *International Conference on Machine Learning (2021)*

Continuation of Point Clouds via Persistence Diagrams
Gameiro M., Hiraoka Y., Obayashi I. *Physica D: Nonlinear Phenomena (2016)*

Persistent homology-based projection pursuit

Given a dataset $X \in \mathbb{R}^m$ find a mapping $f : \mathcal{X} \in \mathbb{R}^m \rightarrow \mathcal{Z} \in \mathbb{R}^n$
where n << m, while optimizing/preserving some relevant properties of the data.

Properties

- variance/distances
- statistical independence
- persistent homology groups

Projection pursuit framework

- mapping in linear
- find best *index*, the value of loss function assigned for each projection Z

Persistent Homology-based Projection Pursuit
Kachan O. *ICML Differential Geometry in Machine Learning Workshop (2020)*

Persistent homology-based projection pursuit

Given a dataset $X \in \mathbb{R}^m$ find a mapping $f : \mathcal{X} \in \mathbb{R}^m \to \mathcal{Z} \in \mathbb{R}^n$, where $n << m$, while preserving topological properties of the data in terms of persistent homology groups

$$\min_{P \in St(n,m)} W_2(\mathscr{D}(X), \mathscr{D}(P^T X))$$

Bottleneck distance bounds Gromov-Hausdorff distance

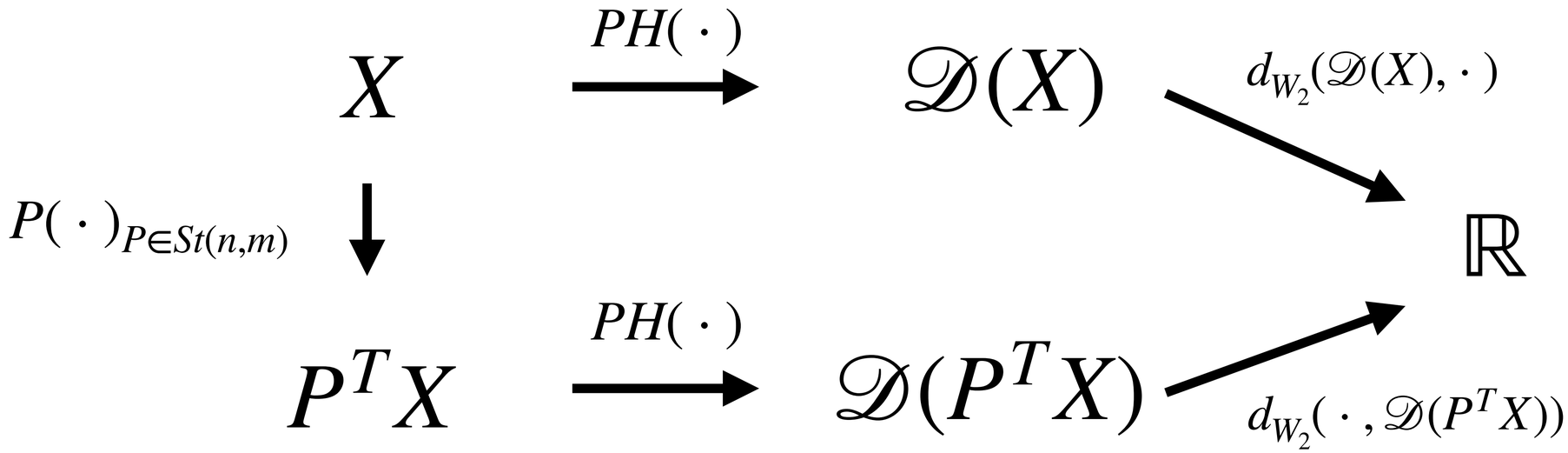
$$d_{W_\infty}(\mathscr{D}(X), \mathscr{D}(Y)) \leq d_{GH}(X, Y)$$

Persistent Homology-based Projection Pursuit
Kachan O. *ICML Differential Geometry in Machine Learning Workshop (2020)*

Persistent homology-based projection pursuit

$$X \to Z$$

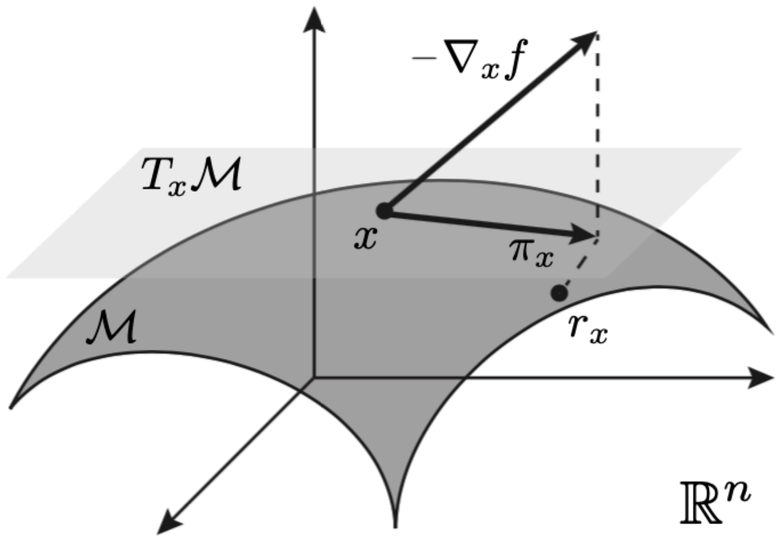
$$Z \cong X$$



Persistent Homology-based Projection Pursuit
Kachan O. *ICML Differential Geometry in Machine Learning Workshop (2020)*

Persistent homology-based projection pursuit

Riemannian optimization generalizes optimization algorithms to the Riemannian manifolds other than $\mathbb{R}^n$.



$$\min_{x \in St(n,m)} f(x)$$

Data: A manifold $\mathcal{M}$, a scalar field f on $\mathcal{M}$, a projection $\pi_x : \mathbb{R}^n \to T_x\mathcal{M}$, a retraction $r_x : T_x\mathcal{M} \to \mathcal{M}$, an initial iterate $x_0 \in \mathcal{M}$, step size $\alpha \in \mathbb{R}_+$

Result: Sequence of iterates $\{x_k\}$

for $k = 1, 2, \ldots$ **do**

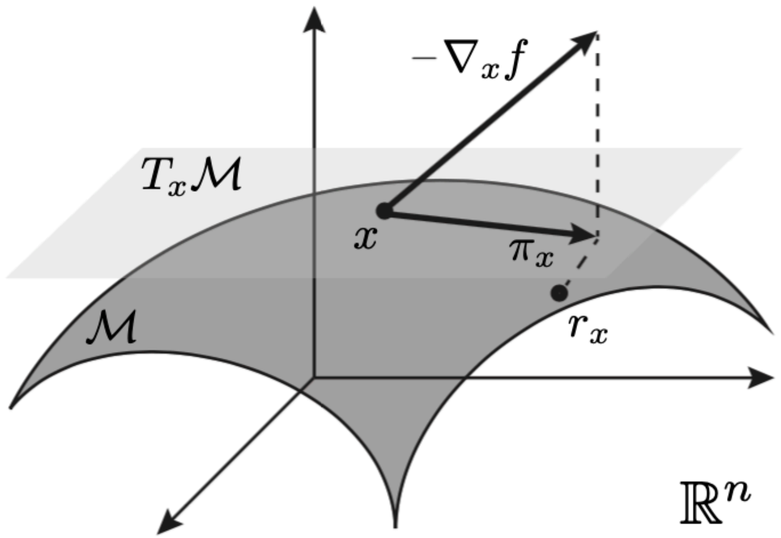
$x_{k+1} = (r_{x_k} \circ \alpha\pi_{x_k})(-\nabla_x f(x_k))$;

return $\{x_k\}$

Persistent Homology-based Projection Pursuit
Kachan O. *ICML Differential Geometry in Machine Learning Workshop (2020)*

Persistent homology-based projection pursuit

Riemannian optimization generalizes optimization algorithms to the Riemannian manifolds other than $\mathbb{R}^n$.



$$\min_{x \in St(n,m)} f(x)$$

Stiefel manifold

$$St(n,m) = \{\mathbf{X} \in \mathbb{R}^{m \times n} \mid \mathbf{X}^T\mathbf{X} = \mathbf{I}_n\}$$

Projection map $\pi_{\mathbf{X}} : \mathscr{R}^{m \times n} \mapsto T_{\mathbf{X}}St(n,m)$

Let $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, then $\pi_{\mathbf{X}}(\mathbf{A}) = \mathbf{U}\mathbf{V}^T$.

Retraction map $r_{\mathbf{X}} : T_{\mathbf{X}}St(n,m) \mapsto St(n,m)$

Via polar decomp. $r_{\mathbf{X}}(\mathbf{V}) = (\mathbf{X} + \mathbf{V})(\mathbf{I}_p + \mathbf{V}^T\mathbf{V})^{-1/2}$

Via QR decomp. $r_{\mathbf{X}}(\mathbf{V}) = qf(\mathbf{X} + \mathbf{V})$

Persistent Homology-based Projection Pursuit
Kachan O. *ICML Differential Geometry in Machine Learning Workshop (2020)*

Persistent homology-based projection pursuit

Given a ground metric $d : X \times X \to \mathbb{R}$ optimal transport equips the space of measures $\mathscr{P}(X)$ with a metric referred to as the Wasserstein distance

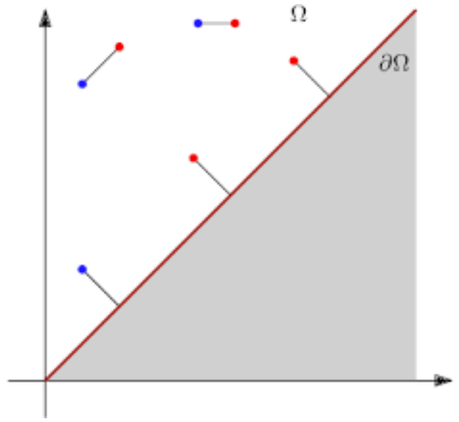
$$W_p^p(\mu, \nu) = \min_{\mathbf{T} \in \Pi(\mathbf{a}, \mathbf{b})} \langle \mathbf{T}, \tilde{\mathbf{M}} \rangle_F$$

$$\Pi(\mathbf{a}, \mathbf{b}) = \{\mathbf{T} \in \mathbb{R}_+^{n \times m} : \mathbf{T}\mathbf{1}_m = \mathbf{a}, \mathbf{T}^T\mathbf{1}_n = \mathbf{b}\}$$

Augmented cost matrix

$$\tilde{\mathbf{M}} = \begin{pmatrix} \mathbf{M} & \Delta_{D(X)} \\ \Delta_{D(Y)} & 0 \end{pmatrix}$$

$$(\mathbf{M})_{ij} = d^p(\mathbf{x}_i, \mathbf{y}_j) \qquad (\mathbf{\Delta}_{D(X)})_i = d^p(\mathbf{x}_i, \partial\Omega)$$
$$(\mathbf{\Delta}_{D(Y)})_j = d^p(\mathbf{y}_j, \partial\Omega)$$

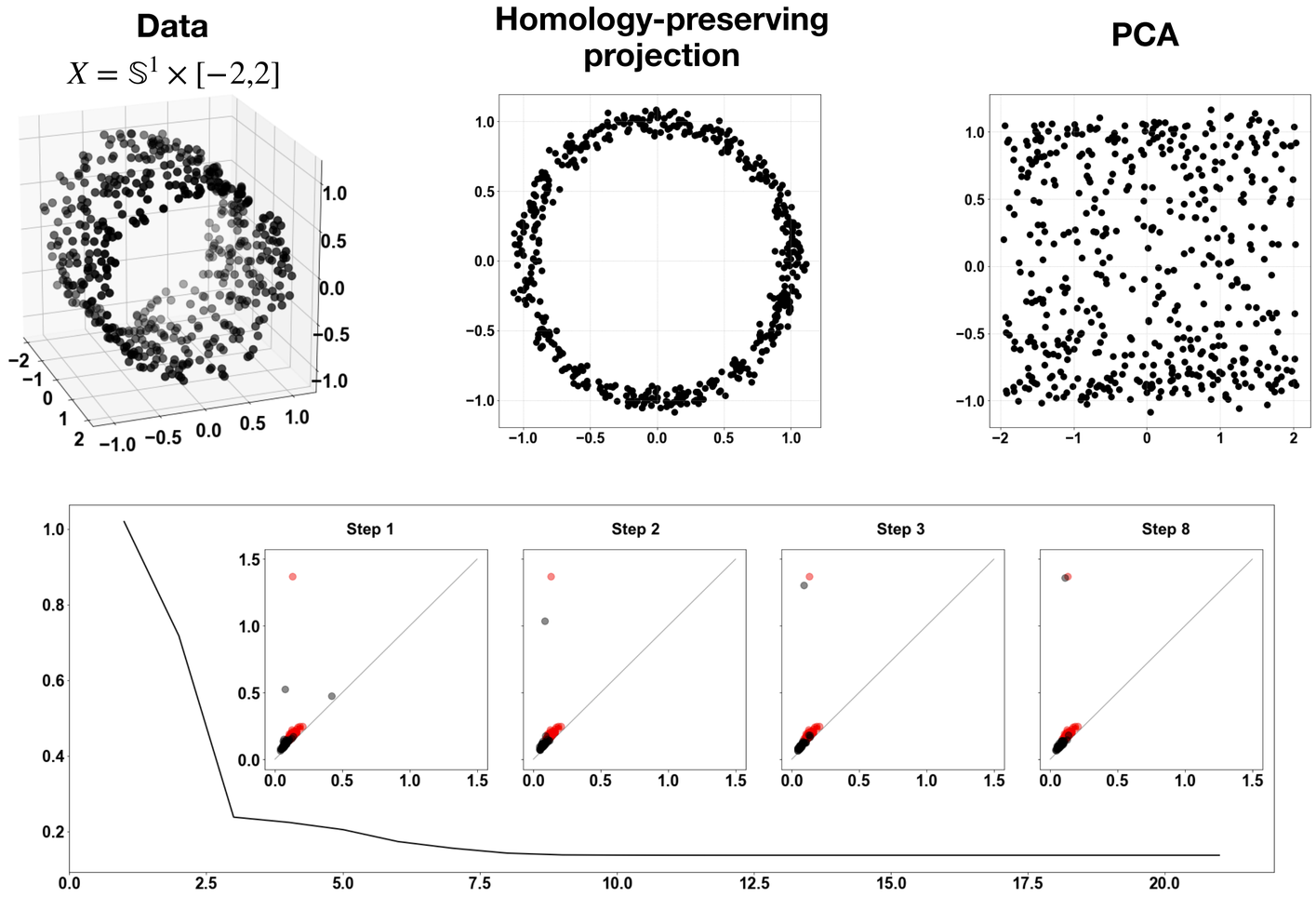


Approximations

- Sinkhorn divergence, O(n^2)
- Sliced Wasserstein distance, O(*n*log*n*)

Persistent Homology-based Projection Pursuit
Kachan O. *ICML Differential Geometry in Machine Learning Workshop (2020)*

Persistent homology-based projection pursuit



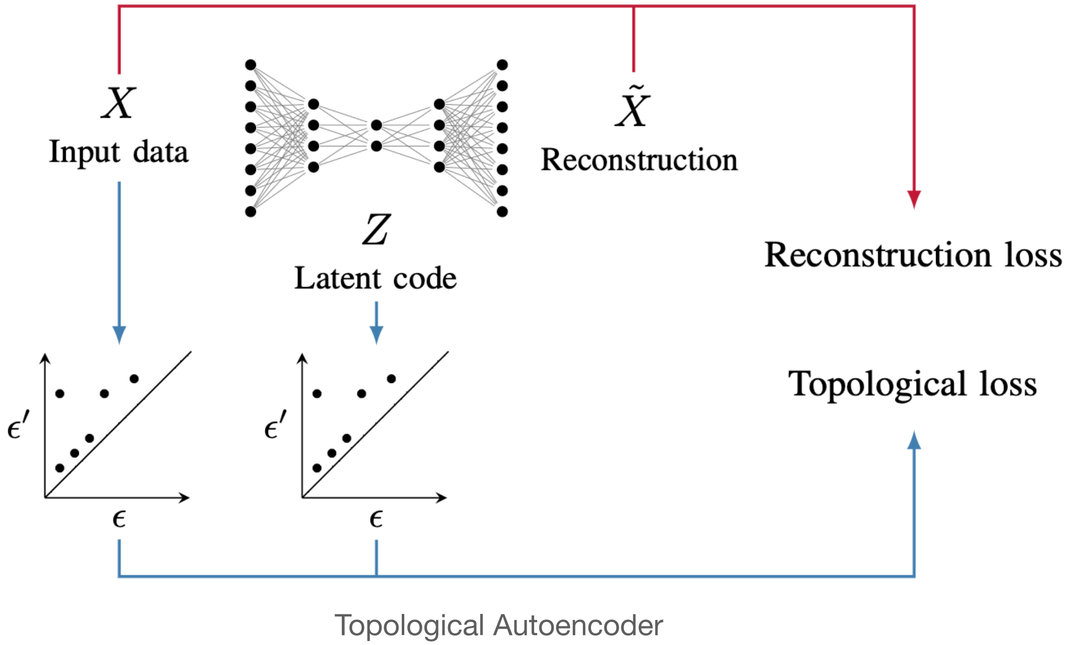
Loss function and persistence diagrams of data D(X) (red, fixed) and low-dimensional projection D(Z) (black).

Persistent Homology-based Projection Pursuit
Kachan O. *ICML Differential Geometry in Machine Learning Workshop (2020)*

Topological autoencoders

$$L = L_{REC}(X, (h \circ g)(X)) + \lambda L_{TOPO}$$
$$L_{TOPO} = L_{X \to Z} + L_{Z \to X}$$

$$L_{X \to Z} = \|\mathbf{A}_X[\pi_X] - \mathbf{A}_Z[\pi_X]\|^2$$
$$L_{Z \to X} = \|\mathbf{A}_Z[\pi_Z] - \mathbf{A}_X[\pi_Z]\|^2$$



Topological Autoencoder

$D_X = \{(10,20), (12,30), (15,50)\}$

$\pi_X = \{(1,12), (2,24), (4,34)\}$

	1	2	3	4
1		20		
2				30
3				50
4				

$A_X[\pi_X]$

	1	2	3	4
1		2		
2				3
3				5
4				

$A_Z[\pi_X]$

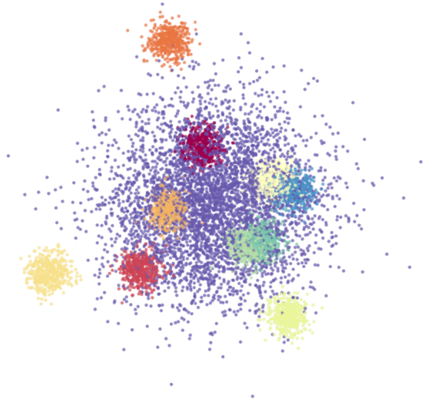
$$X \to Z \to \tilde{X}$$

$$Z \cong X$$

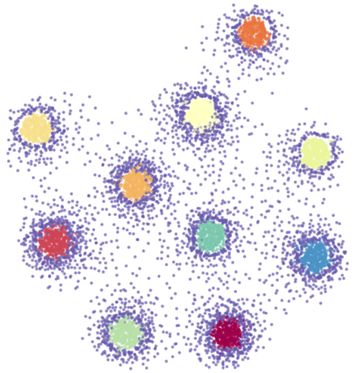
Topological Autoencoders
Moor et al. *International Conference on Learning Representations (2020)*

Topological autoencoders

SPHERES data set that consists of ten high-dimensional 100-spheres or radius R living in a 101-dimensional space that are enclosed by one larger sphere of radius 5R that consists of the same number of points as the total of inner spheres.

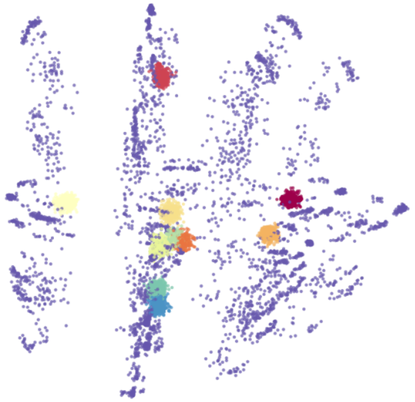


PCA

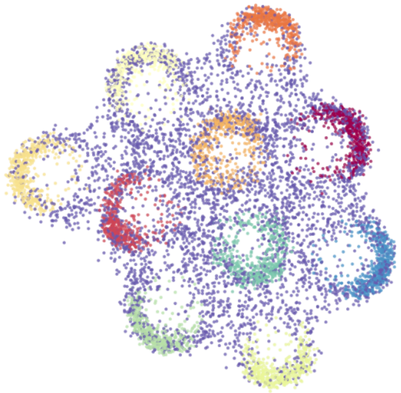


t-SNE

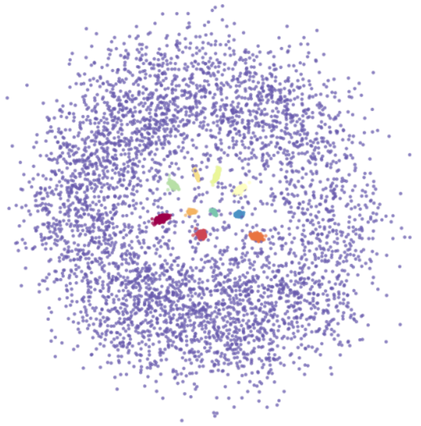
AE



Isomap



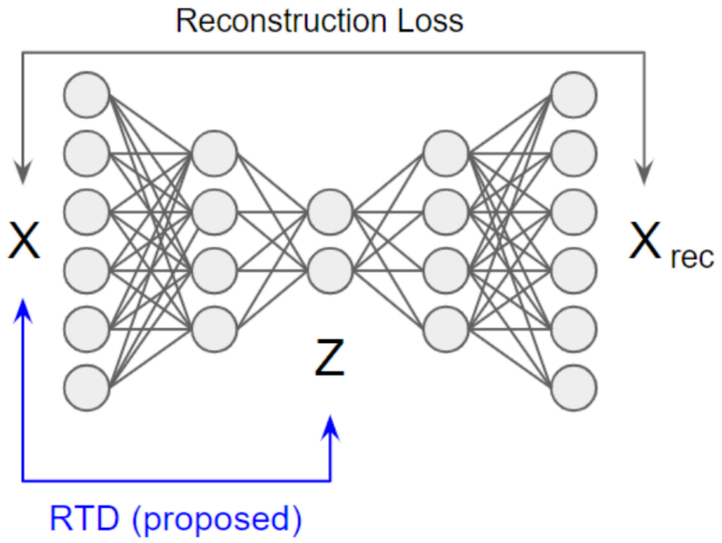
UMAP



TopoAE

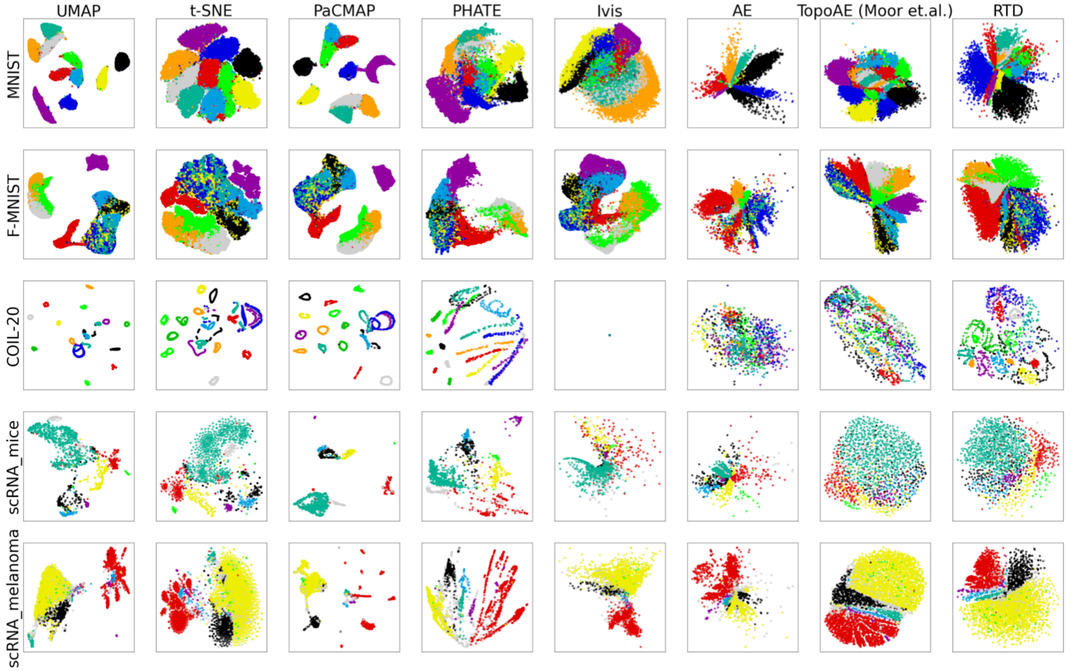
Topological Autoencoders
Moor et al. *International Conference on Learning Representations (2020)*

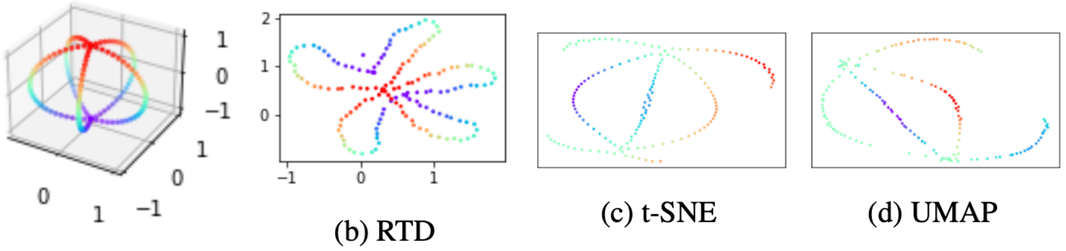
Topology-preserving representation learning



$$X \to Z \to \tilde{X}$$

$$Z \cong X$$

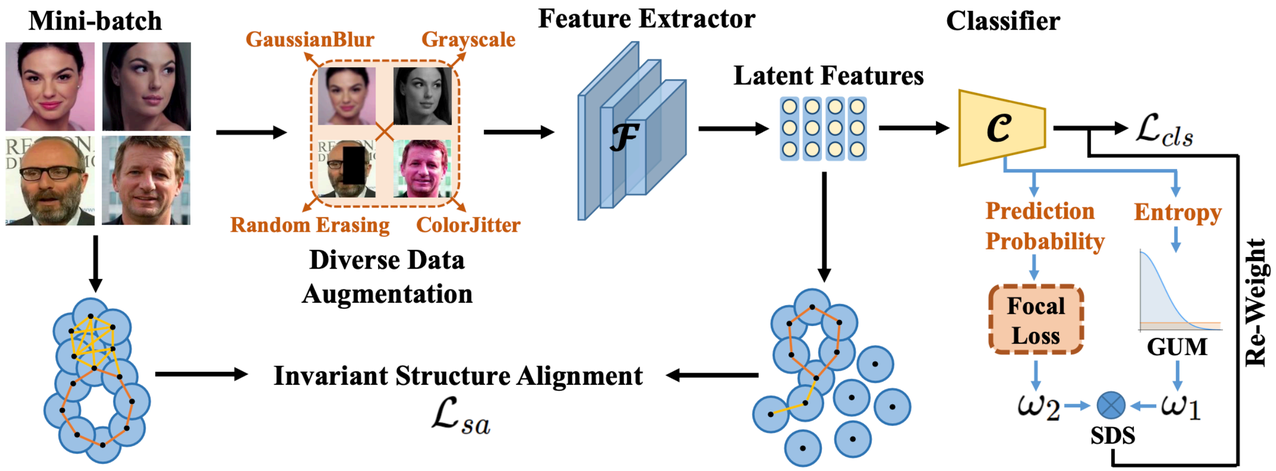




(b) RTD (c) t-SNE (d) UMAP

Learning Topology-preserving Data Representations
Trofimov et al. *International Conference on Learning Representations (2023)*

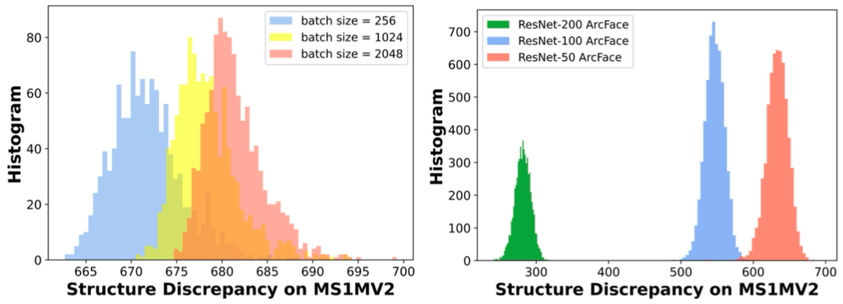
Topology alignment for face recognition



$$L = L_{CLS} + \lambda L_{SA}$$

$$X \rightarrow Z \rightarrow Y$$

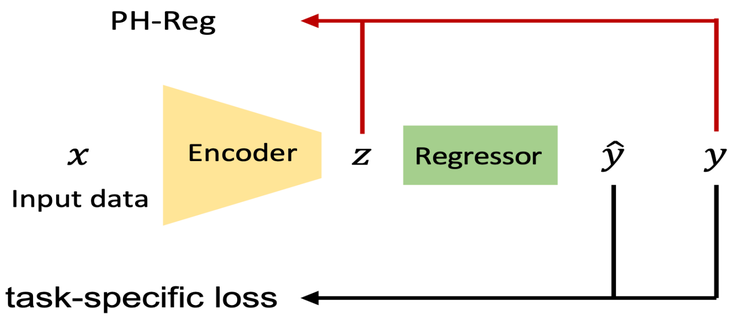
$$Z \cong X$$



Training Data	Method	IJB-C(1e-6)	IJB-C(1e-5)	IJB-C(1e-4)
MS1MV2	R50, ArcFace Deng et al. (2019)	80.52	88.36	92.52
	R50, MagFace Meng et al. (2021)	81.69	88.95	93.34
	R50, AdaFace Kim et al. (2022)	-	-	96.27
	R50, **TopoFR**†	89.32	**94.77**	96.40
	R50, **TopoFR**	**90.52**	94.71	**96.49**
	R100, CosFace Wang et al. (2018b)	87.96	92.68	95.56
	R100, ArcFace Deng et al. (2019)	85.65	92.69	95.74
	R100, MV-Softmax Wang et al. (2020)	-	-	95.20
	R100, CircleLoss Sun et al. (2020)	-	89.60	93.95
	R100, URL Shi et al. (2020)	-	95.00	96.60
	R100, BroadFace Kim et al. (2020)	85.96	94.59	96.38
	R100, CurricularFace Huang et al. (2020)	-	-	96.10
	R100, MagFace+ Meng et al. (2021)	**90.24**	94.08	95.97
	R100, SCF Li et al. (2021b)	-	94.78	96.22
	R100, DAM-CurricularFace Liu et al. (2021)			96.20
	R100, ElasticFace+ Boutros et al. (2022)	-	-	96.65
	R100, AdaFace Kim et al. (2022)	-	-	96.89
	R100, **TopoFR**†	87.90	**95.27**	96.90
	R100, **TopoFR**	90.21	95.23	**96.95**
	R200, ArcFace Deng et al. (2019)	85.75	94.67	96.53
	R200, **TopoFR**†	87.90	**95.19**	**97.12**
	R200, **TopoFR**	**88.39**	95.15	97.08
Glint360K	R50, ArcFace Deng et al. (2019)	88.40	95.29	96.81
	R50, **TopoFR**	**91.30**	**95.99**	**97.27**
	R100, ArcFace Deng et al. (2019)	86.97	96.09	97.04
	R100, **TopoFR**	**91.06**	**96.57**	**97.60**
	R200, ArcFace Deng et al. (2019)	**89.45**	95.71	97.20
	R200, **TopoFR**	86.84	**96.54**	**97.74**

TopoFR: A Closer Look at Topology Alignment on Face Recognition
Dan et al. https://openreview.net/forum?id=R4gqcDRJ9l *(2023)*

Deep regression representation learning with topology

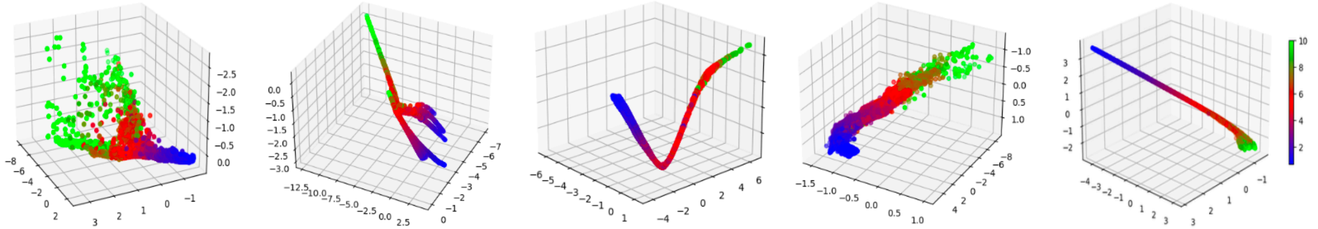


$$X \to Z \to Y$$

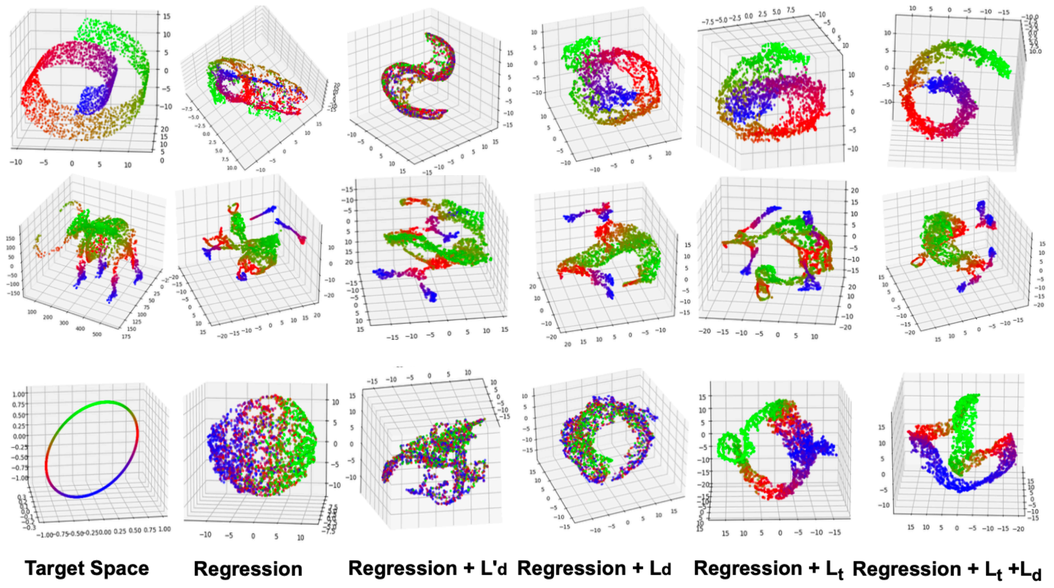
$$Z \cong Y$$

Table 3: Results on AgeDB

Method	MAE ↓		
	ALL	Many	Med.
Baseline (Yang et al., 2021)	7.77	6.62	9.55
$+\mathcal{L}'_d$	7.81	6.96	8.88
$+\mathcal{L}_d$	7.55	6.81	**8.43**
$+\mathcal{L}_t$	7.50	6.58	8.79
$+\mathcal{L}_d + \mathcal{L}_t$	**7.48**	**6.52**	8.71

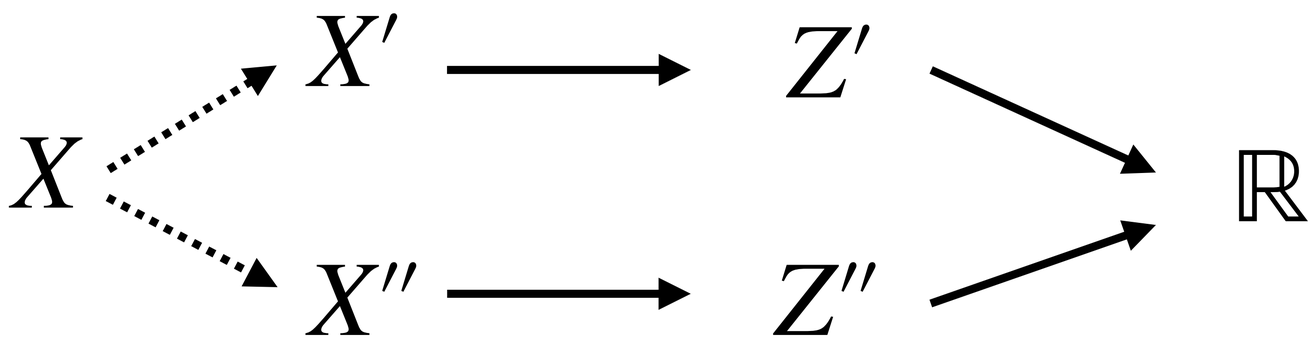


(a) Regression (b) Regression $+\mathcal{L}'_d$ (c) Regression $+\mathcal{L}_d$ (d) Regression $+\mathcal{L}_t$ (e) Regression $+\mathcal{L}_R$



Deep Regression Representation Learning with Topology
Zhang et al. https://arxiv.org/abs/2404.13904 *(2024)*

Self-supervised learning



$Z \cong X$ $Z \cong Y$