Topological Autoencoders

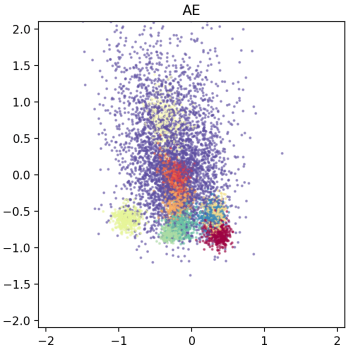
Michael Moor[†], Max Horn[†], Bastian Rieck[‡] and Karsten Borgwardt[‡]

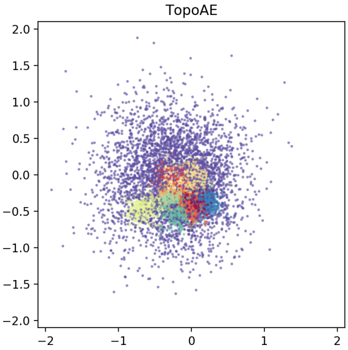
Machine Learning and Computational Biology Group, ETH Zurich

ICML 2020

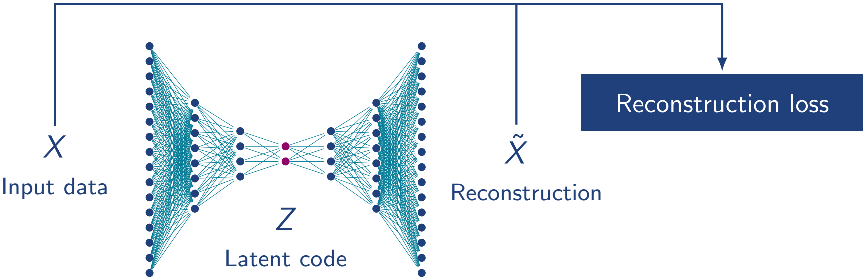
Motivation

Motivation

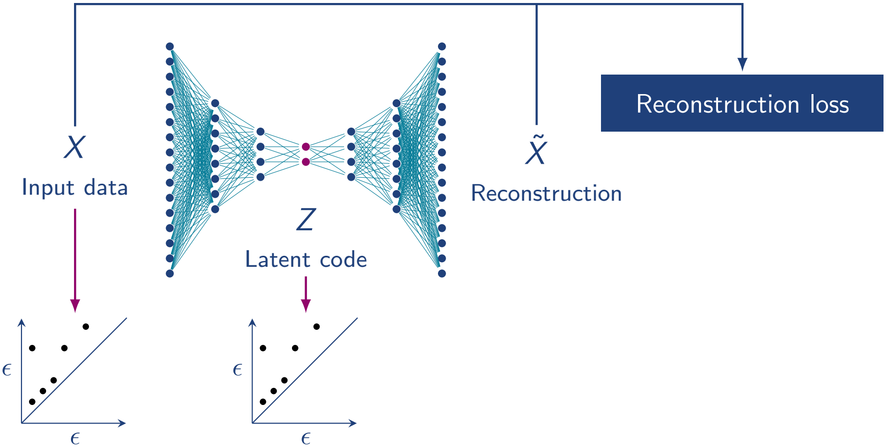

AE
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2
−1
0
1
2


TopoAE
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2
−1
0
1
2

Overview

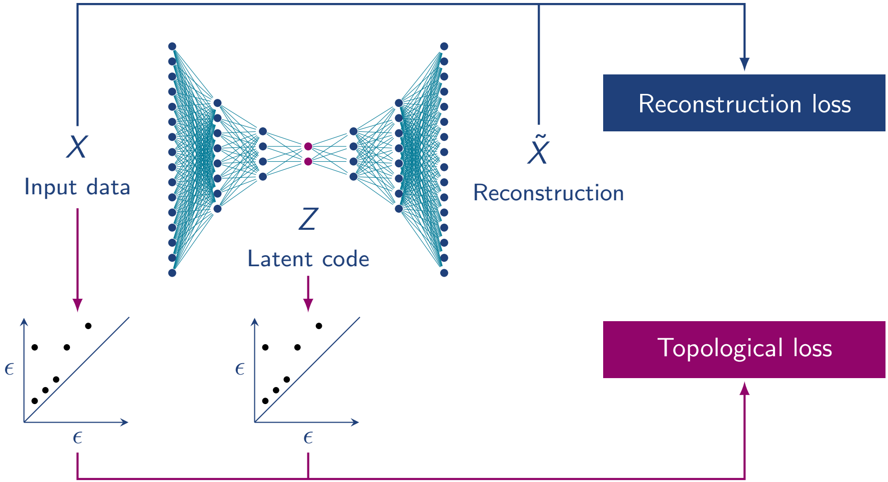

X
Input data
Z
Latent code
$\tilde{X}$
Reconstruction
Reconstruction loss

Overview



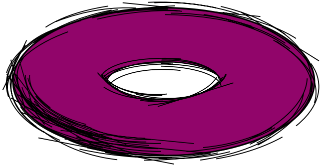
X
Input data
Z
Latent code
$\tilde{X}$
Reconstruction
Reconstruction loss
ϵ
ϵ
ϵ
ϵ

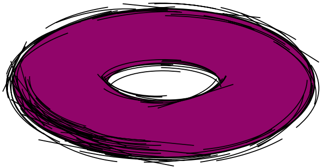
Overview



X
Input data
Z
Latent code
$\tilde{X}$
Reconstruction
Reconstruction loss
Topological loss
ϵ
ϵ
ϵ
ϵ

Topology - The study of connectivity

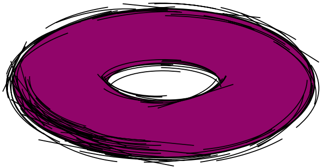


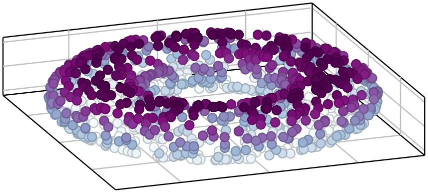


Betti numbers characterize topological spaces

- β_0 connected components
- β_1 cycles
- β_2 voids

Topology - The study of connectivity

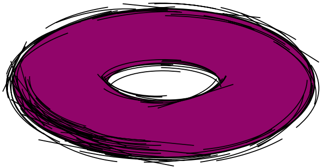


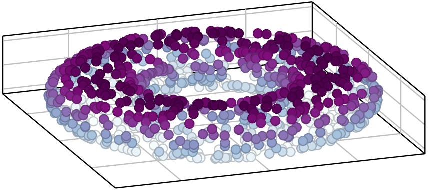


Betti numbers characterize topological spaces

- β_0 connected components
- β_1 cycles
- β_2 voids

Topology - The study of connectivity





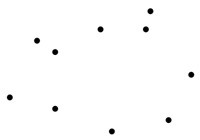
Betti numbers characterize topological spaces

- β_0 connected components
- β_1 cycles
- β_2 voids

Issues

- Great for manifolds (which are usually unknown)
- But instead *approximated* via samples
- Topology on samples is noisy

Vietoris-Rips Complex[1]: We 'grow' a neighbourhood graph (simplicial complex for higher dimensions) and keep track of the appearance and disappearance of topological features.



Filtration:

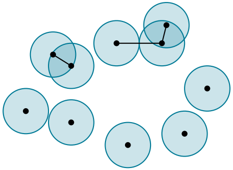
$$\emptyset = \mathrm{K}_0 \subseteq \mathrm{K}_1 \subseteq \cdots \subseteq \mathrm{K}_{n-1} \subseteq \mathrm{K}_n = \mathrm{K}$$

$$E := \left\{ (u, v) \mid \mathrm{dist}\,(p_u, p_v) \leq \epsilon \right\}$$

[1] Vietoris [1927]
[2] Edelsbrunner and Harer [2008]

Persistent Homology (PH)[2]

Vietoris-Rips Complex[1]: We 'grow' a neighbourhood graph (simplicial complex for higher dimensions) and keep track of the appearance and disappearance of topological features.



Filtration:

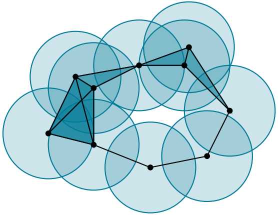
$$\emptyset = \mathrm{K}_0 \subseteq \mathrm{K}_1 \subseteq \cdots \subseteq \mathrm{K}_{n-1} \subseteq \mathrm{K}_n = \mathrm{K}$$

$$E := \left\{ (u, v) \mid \mathrm{dist}\left(p_u, p_v\right) \leq \epsilon \right\}$$

[1] Vietoris [1927]
[2] Edelsbrunner and Harer [2008]

Persistent Homology (PH)[2]

Vietoris-Rips Complex[1]: We ‘grow‘ a neighbourhood graph (simplicial complex for higher dimensions) and keep track of the appearance and disappearance of topological features.



Filtration:

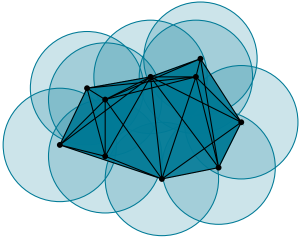
$$\emptyset = \mathrm{K}_0 \subseteq \mathrm{K}_1 \subseteq \cdots \subseteq \mathrm{K}_{n-1} \subseteq \mathrm{K}_n = \mathrm{K}$$

$$E := \left\{ (u, v) \mid \mathrm{dist}\,(p_u, p_v) \leq \epsilon \right\}$$

[1] Vietoris [1927]
[2] Edelsbrunner and Harer [2008]

Persistent Homology (PH)[2]

Vietoris-Rips Complex[1]**:** We 'grow' a neighbourhood graph (simplicial complex for higher dimensions) and keep track of the appearance and disappearance of topological features.



$$E := \left\{ (u, v) \mid \mathrm{dist}\left(p_u, p_v\right) \leq \epsilon \right\}$$

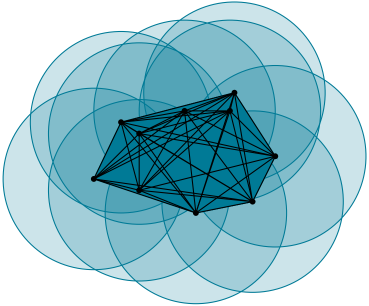
Filtration:

$$\emptyset = \mathrm{K}_0 \subseteq \mathrm{K}_1 \subseteq \cdots \subseteq \mathrm{K}_{n-1} \subseteq \mathrm{K}_n = \mathrm{K}$$

[1] Vietoris [1927]
[2] Edelsbrunner and Harer [2008]

Persistent Homology (PH)[2]

Vietoris-Rips Complex[1]: We ‘grow‘ a neighbourhood graph (simplicial complex for higher dimensions) and keep track of the appearance and disappearance of topological features.



$$E := \left\{ (u, v) \mid \mathrm{dist}\left(p_u, p_v\right) \leq \epsilon \right\}$$

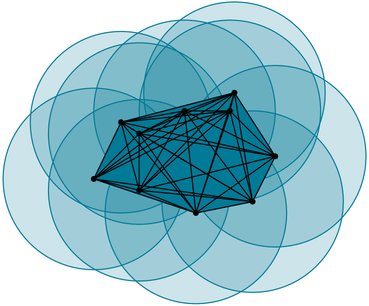
Filtration:

$$\emptyset = \mathrm{K}_0 \subseteq \mathrm{K}_1 \subseteq \cdots \subseteq \mathrm{K}_{n-1} \subseteq \mathrm{K}_n = \mathrm{K}$$

[1] Vietoris [1927]
[2] Edelsbrunner and Harer [2008]

Persistent Homology (PH)[2]

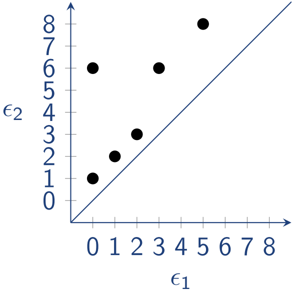
Vietoris-Rips Complex[1]: We ‘grow‘ a neighbourhood graph (simplicial complex for higher dimensions) and keep track of the appearance and disappearance of topological features.



$$E := \{ (u, v) \mid \mathrm{dist}(p_u, p_v) \leq \epsilon \}$$

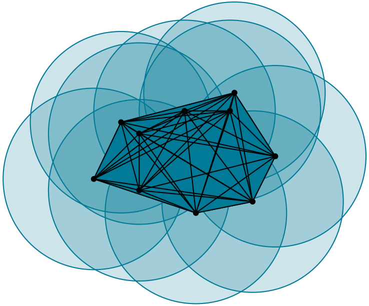
Filtration:

$$\emptyset = \mathrm{K}_0 \subseteq \mathrm{K}_1 \subseteq \cdots \subseteq \mathrm{K}_{n-1} \subseteq \mathrm{K}_n = \mathrm{K}$$



[1] Vietoris [1927]
[2] Edelsbrunner and Harer [2008]

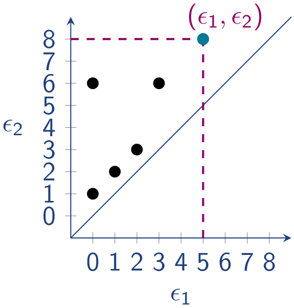
Vietoris-Rips Complex[1]: We 'grow' a neighbourhood graph (simplicial complex for higher dimensions) and keep track of the appearance and disappearance of topological features.



$$E := \{ (u, v) \mid \mathrm{dist}(p_u, p_v) \le \epsilon \}$$

Filtration:

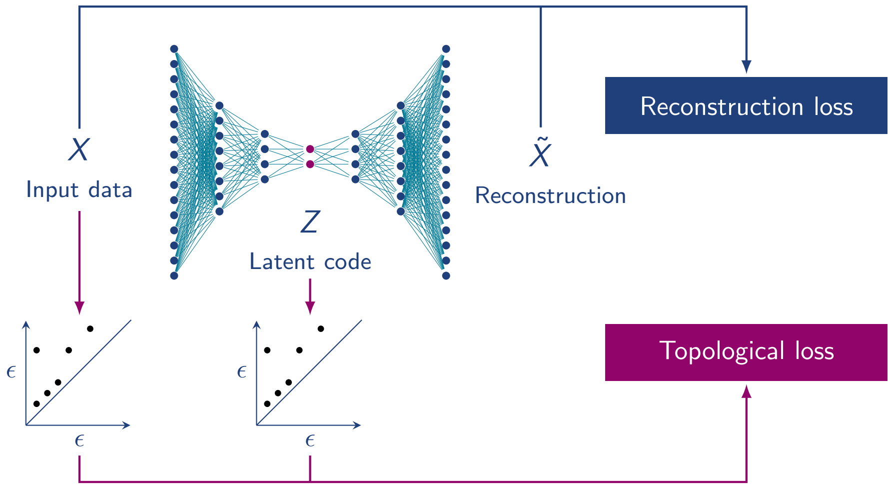
$$\emptyset = \mathrm{K}_0 \subseteq \mathrm{K}_1 \subseteq \cdots \subseteq \mathrm{K}_{n-1} \subseteq \mathrm{K}_n = \mathrm{K}$$



[1] Vietoris [1927]
[2] Edelsbrunner and Harer [2008]

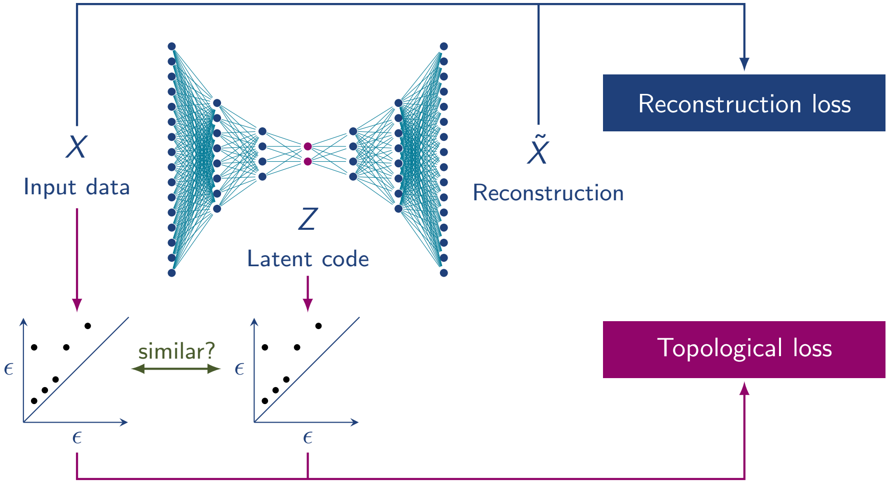
Persistent Homology II

Overview



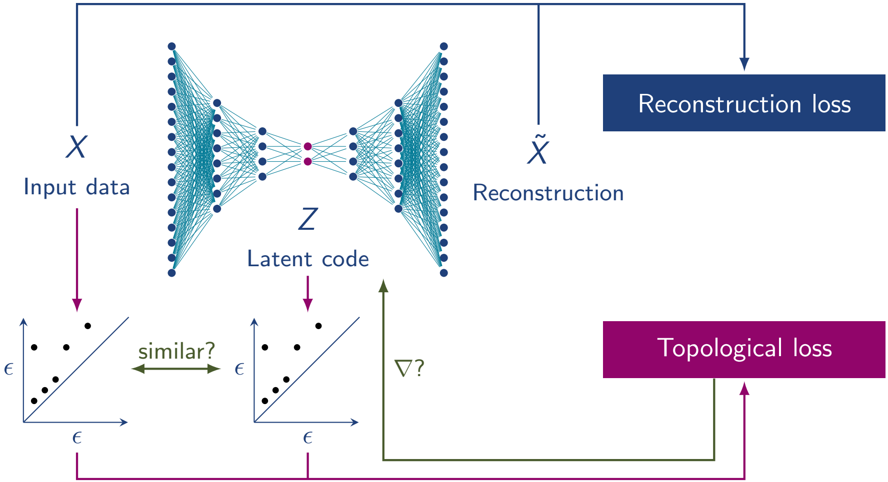
X
Input data
Z
Latent code
$\tilde{X}$
Reconstruction
Reconstruction loss
Topological loss
ϵ
ϵ
ϵ
ϵ

Overview



X
Input data
Z
Latent code
$\tilde{X}$
Reconstruction
Reconstruction loss
Topological loss
similar?
ϵ
ϵ
ϵ
ϵ

Overview



X
Input data
Z
Latent code
$\tilde{X}$
Reconstruction
Reconstruction loss
Topological loss
similar?
∇?
ϵ
ϵ
ϵ
ϵ

Distance matrix and relation to persistence diagrams

Distance matrix **A**

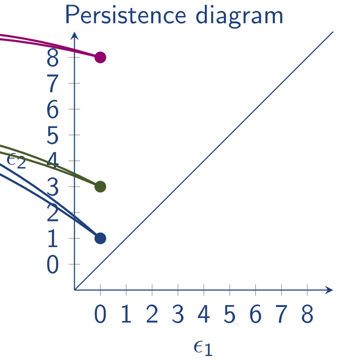
$$\begin{bmatrix} 0 & 1 & 2 & 10 \\ 1 & 0 & 8 & 2 \\ 2 & 8 & 0 & 3 \\ 10 & 2 & 3 & 0 \end{bmatrix}$$

Distance matrix and relation to persistence diagrams

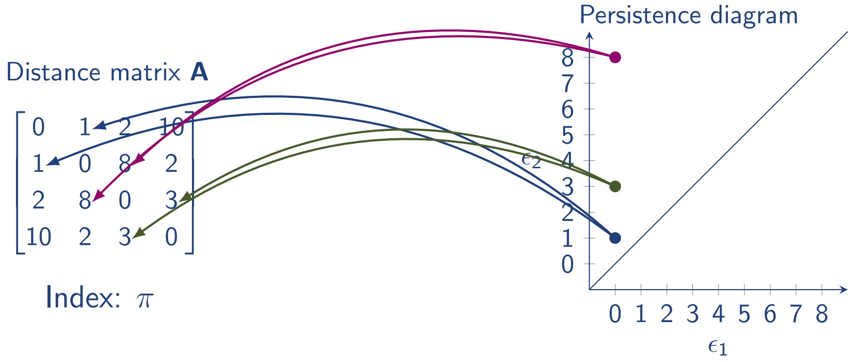
Distance matrix $\mathbf{A}$

$$\begin{bmatrix} 0 & 1 & 2 & 10 \\ 1 & 0 & 8 & 2 \\ 2 & 8 & 0 & 3 \\ 10 & 2 & 3 & 0 \end{bmatrix}$$

Index: π



Distance matrix and relation to persistence diagrams



Notation:

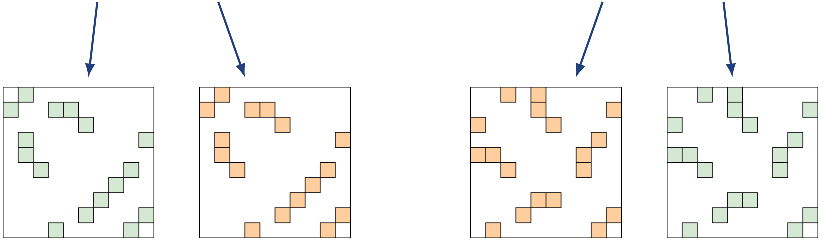
$\mathbf{A}^X$ = distance matrix of mini-batch in data space

π^X = index set resulting from PH calculation in data space

$\mathbf{A}^X\left[\pi^X\right]$ = vector of distances selected with π^X

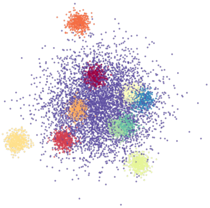
$$\mathcal{L}_t = \mathcal{L}_{\mathcal{X}\to\mathcal{Z}} + \mathcal{L}_{\mathcal{Z}\to\mathcal{X}}$$

$$\mathcal{L}_{\mathcal{X}\to\mathcal{Z}} := \frac{1}{2}\left\| \mathbf{A}^X\left[\pi^X\right] - \mathbf{A}^Z\left[\pi^X\right]\right\|^2 \qquad \mathcal{L}_{\mathcal{Z}\to\mathcal{X}} := \frac{1}{2}\left\| \mathbf{A}^Z\left[\pi^Z\right] - \mathbf{A}^X\left[\pi^Z\right]\right\|^2$$

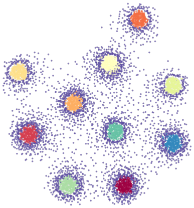


Experiments

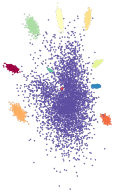
Spheres



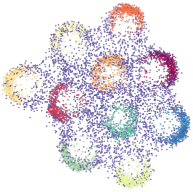
PCA



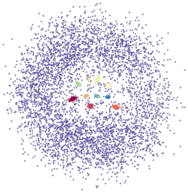
t-SNE



Autoencoder



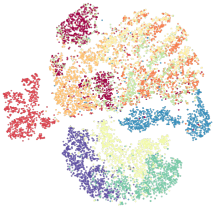
UMAP



Topo-AE

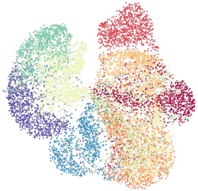
FashionMNIST [Xiao et al., 2017]

PCA



t-SNE

Autoencoder



UMAP

Topo-AE

Insights and Summary

- Novel method for preserving topological information of the input space in dimensionality reduction

Insights and Summary

- Novel method for preserving topological information of the input space in dimensionality reduction
- Under weak theoretical assumptions our topological loss term is differentiable and allowing the training of neural networks via backpropagation.

Insights and Summary

- Novel method for preserving topological information of the input space in dimensionality reduction
- Under weak theoretical assumptions our topological loss term is differentiable and allowing the training of neural networks via backpropagation.
- We prove that approximating topological features on the mini-batch level is robust.

Insights and Summary

- Novel method for preserving topological information of the input space in dimensionality reduction
- Under weak theoretical assumptions our topological loss term is differentiable and allowing the training of neural networks via backpropagation.
- We prove that approximating topological features on the mini-batch level is robust.
- Our method was uniquely able to capture spatial relationships of nested high-dimensional spheres

For further information, please check out our

Paper:

https://arxiv.org/abs/1906.00722

Code:

Credits:

- Aleph for TDA calculations https://github.com/Pseudomanifold/Aleph
- manim for animations https://github.com/3b1b/manim

References

H. Edelsbrunner and J. Harer. Persistent homology—a survey. In J. E. Goodman, J. Pach, and R. Pollack, editors, *Surveys on discrete and computational geometry: Twenty years later*, number 453 in Contemporary Mathematics, pages 257–282. American Mathematical Society, Providence, RI, USA, 2008.

L. Vietoris. Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen. *Mathematische Annalen*, 97(1):454–472, 1927.

H. Xiao, K. Rasul, and R. Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms, 2017.

Appendix

Bound of bottleneck distance between persistence diagrams on subsampled data

Theorem

Let X be a point cloud of cardinality n and $X^{(m)}$ be one subsample of X of cardinality m, i.e. $X^{(m)} \subseteq X$, sampled without replacement. We can bound the probability of the persistence diagrams of $X^{(m)}$ exceeding a threshold in terms of the bottleneck distance as

$$\mathbb{P}\left(\mathsf{d_b}\left(\mathcal{D}^X, \mathcal{D}^{X^{(m)}}\right) > \epsilon\right) \leq \mathbb{P}\left(\mathsf{d_H}\left(X, X^{(m)}\right) > 2\epsilon\right),$$

where $\mathsf{d_H}$ refers to the Hausdorff distance between the point cloud and its subsample.

Expected value of Hausdorff distance

Theorem

Let $\mathbf{A} \in^{n\times m}$ *be the distance matrix between samples of* X *and* $X^{(m)}$*, where the rows are sorted such that the first* m *rows correspond to the columns of the* m *subsampled points with diagonal elements* $a_{ii} = 0$*. Assume that the entries* a_{ij} *with* $i > m$ *are random samples following a distance distribution* F_D *with* $\operatorname{supp}(f_D) \in_{\geq 0}$*. The minimal distances* δ_i *for rows with* $i > m$ *follow a distribution* F_Δ*. Letting* $Z := \max_{1\leq i\leq n} \delta_i$ *with a corresponding distribution* F_Z*, the expected Hausdorff distance between* X *and* $X^{(m)}$ *for* $m < n$ *is bounded by:*

$$\mathbb{E}\left[\mathrm{d_H}(X, X^{(m)})\right] = \mathbb{E}_{Z\sim F_Z}[Z] \leq \int_0^{+\infty} \left(1 - F_D(z)^{(n-1)}\right) \mathrm{d}z \leq \int_0^{+\infty} \left(1 - F_D(z)^{m(n-m)}\right) \mathrm{d}z$$

Explicit Gradient Derivation

Letting $\boldsymbol{\theta}$ refer to the parameters of the *encoder*, we have

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{L}_{\mathcal{X} \to \mathcal{Z}} &= \frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{1}{2} \left\| \mathbf{A}^X \left[\pi^X \right] - \mathbf{A}^Z \left[\pi^X \right] \right\|^2 \right) \\
&= -\left(\mathbf{A}^X \left[\pi^X \right] - \mathbf{A}^Z \left[\pi^X \right] \right)^\top \left(\frac{\partial \mathbf{A}^Z \left[\pi^X \right]}{\partial \boldsymbol{\theta}} \right) \\
&= -\left(\mathbf{A}^X \left[\pi^X \right] - \mathbf{A}^Z \left[\pi^X \right] \right)^\top \left(\sum_{i=1}^{\left| \pi^X \right|} \frac{\partial \mathbf{A}^Z \left[\pi^X \right]_i}{\partial \boldsymbol{\theta}} \right),
\end{aligned}
$$

where $\left| \pi^X \right|$ denotes the cardinality of a persistence pairing and $\mathbf{A}^Z \left[\pi^X \right]_i$ refers to the ith entry of the vector of paired distances.

Density distribution error

Definition (Density distribution error)

Let $\sigma \in_{>0}$. For a finite metric space $\mathcal{S}$ with an associated distance $\operatorname{dist}(\cdot,\cdot)$, we evaluate the density at each point $x \in \mathcal{S}$ as

$$\mathrm{f}_\sigma^{\mathcal{S}}(x) := \sum_{y \in \mathcal{S}} \exp\left(-\sigma^{-1} \operatorname{dist}(x,y)^2\right),$$

where we assume without loss of generality that $\max \operatorname{dist}(x,y) = 1$. We then calculate $\mathrm{f}_\sigma^{X}(\cdot)$ and $\mathrm{f}_\sigma^{Z}(\cdot)$, normalise them such that they sum to 1, and evaluate

$$\mathrm{KL}_\sigma := \mathrm{KL}\left(\mathrm{f}_\sigma^{X} \parallel \mathrm{f}_\sigma^{Z}\right), \tag{1}$$

i.e. the Kullback–Leibler divergence between the two density estimates.

Quantification of performance

Data set	Method	$KL_{0.01}$	$KL_{0.1}$	KL_1	ℓ-MRRE	ℓ-Cont	ℓ-Trust	ℓ-RMSE	Data MSE
Spheres	Isomap	0.181	**0.420**	**0.00881**	**0.246**	**0.790**	**0.676**	10.4	–
	PCA	0.332	0.651	0.01530	0.294	0.747	0.626	11.8	0.9610
	TSNE	**0.152**	0.527	0.01271	<u>**0.217**</u>	0.773	<u>**0.679**</u>	<u>**8.1**</u>	–
	UMAP	0.157	0.613	0.01658	0.250	0.752	0.635	**9.3**	–
	AE	0.566	0.746	0.01664	0.349	0.607	0.588	13.3	<u>**0.8155**</u>
	TopoAE	<u>**0.085**</u>	<u>**0.326**</u>	<u>**0.00694**</u>	0.272	<u>**0.822**</u>	0.658	13.5	**0.8681**
F-MNIST	PCA	<u>**0.356**</u>	<u>**0.052**</u>	<u>**0.00069**</u>	0.057	0.968	0.917	<u>**9.1**</u>	0.1844
	TSNE	0.405	0.071	0.00198	<u>**0.020**</u>	0.967	**0.974**	41.3	–
	UMAP	0.424	0.065	0.00163	0.029	<u>**0.981**</u>	0.959	**13.7**	–
	AE	0.478	0.068	0.00125	**0.026**	0.968	<u>**0.974**</u>	20.7	<u>**0.1020**</u>
	TopoAE	**0.392**	**0.054**	**0.00100**	0.032	**0.980**	0.956	20.5	**0.1207**
MNIST	PCA	0.389	0.163	0.00160	0.166	0.901	0.745	<u>**13.2**</u>	0.2227
	TSNE	<u>**0.277**</u>	**0.133**	0.00214	<u>**0.040**</u>	0.921	<u>**0.946**</u>	22.9	–
	UMAP	**0.321**	0.146	0.00234	**0.051**	<u>**0.940**</u>	**0.938**	**14.6**	–
	AE	0.620	0.155	**0.00156**	0.058	0.913	0.937	18.2	<u>**0.1373**</u>
	TopoAE	0.341	<u>**0.110**</u>	<u>**0.00114**</u>	0.056	**0.932**	0.928	19.6	**0.1388**

Quantification of performance - 2

Data set	Method	$KL_{0.01}$	$KL_{0.1}$	KL_1	ℓ-MRRE	ℓ-Cont	ℓ-Trust	ℓ-RMSE	Data MSE
	PCA	**0.591**	**0.020**	**<u>0.00023</u>**	0.119	**<u>0.931</u>**	0.821	**<u>17.7</u>**	0.1482
	TSNE	0.627	0.030	0.00073	**<u>0.103</u>**	0.903	**0.863**	**25.6**	–
CIFAR	UMAP	0.617	0.026	0.00050	0.127	0.920	0.817	33.6	–
	AE	0.668	0.035	0.00062	0.132	0.851	**<u>0.864</u>**	36.3	**0.1403**
	TopoAE	**<u>0.556</u>**	**<u>0.019</u>**	**0.00031**	**0.108**	**0.927**	0.845	37.9	**<u>0.1398</u>**