$(\text{P-LVR}_t(\mathcal{D}))_{t\in\mathbb{R}}$ and $(\text{LS-LVR}_t(\mathcal{D}))_{t\in\mathbb{R}}$, called plain-LVR and locally scaled LVR, respectively, are constructed from a sample $\mathcal{D} = \{(x_1, y_1), \ldots, (x_m, y_m)\}$ containing elements from both classes. Specifically, given $\bullet \in \{\text{P-LVR}, \text{LS-LVR}\}$, the simplicial complexes of these filtrations are

$$\bullet_t(\mathcal{D}) = \{\sigma \subseteq \{x_i\}_{i=1}^m : \{\sigma_i, \sigma_j\} \in G_t^\bullet(\mathcal{D}) \text{ for all } \{\sigma_i, \sigma_j\} \subseteq \sigma\},$$

where $G_t^\bullet(\mathcal{D})$ is the simplicial complex of dimension 1 consisting of the vertices and edges of a bipartite graph $G_t^\bullet = (V_1, V_2, E_t)$ with $V_j = \{x_i : y_i = j\}$ for $j \in [2]$ satisfying $\{v_1, v_2\} \in E_t$ if and only if $d_\bullet(v_1, v_2) \leq t$, plus a set of new edges joining the vertices connected by paths of length two in $G_t^\bullet$. The dissimilarities $d_\bullet$ are given by

$$d_{\text{P-LVR}}(v_1, v_2) = \|v_1 - v_2\|_2, \qquad d_{\text{LS-LVR}}(v_1, v_2) = \frac{\|v_1 - v_2\|_2}{\sqrt{\rho_1 \rho_2}},$$

where ρ_i, $i \in [2]$ is the local scale of v_i, that is, the radius of the smallest sphere centered at v_i that encloses at least k points from the opposite class, where k is a fixed parameter.

A problem with the previous constructions is that they may require very large samples to robustly recover the topology of the decision boundaries. Obtaining such samples can sometimes be infeasible, especially when the cost of evaluating the value of $y \in \mathcal{Y}$ by the unknown function f given $x \in \mathcal{X}$ is high. Li et al. (2020) propose a method based on active learning (Settles, 2009) to sample a small but meaningful set to recover the homology of the decision boundaries using the labeled Čech and Vietoris–Rips complexes introduced in the article. Information extracted from persistence diagrams is used in both articles to perform model selection for basic datasets including Banknote (Lohweg, 2013), CIFAR-10 (Krizhevsky, 2009), MNIST (Deng, 2012), and Fashion-MNIST (Xiao et al., 2017), and basic machine and deep learning models.

Decision boundaries are directly related to the performance of neural networks. Given a binary classification task and a neural network $\mathcal{N}$ with $N_L = 1$ and projection function $\pi(x) = 2$ if $x > 0$ and $x \leq 0$, recall that the decision boundary of $\mathcal{N}$ is the set $\phi_{\mathcal{N}}^{-1}(0)$, as in Equation (5) for $b = 0$. Let $x \in \text{Dom}(\phi_{\mathcal{N}})$. For this kind of neural network, the higher the absolute value of $\phi_{\mathcal{N}}(x)$, the more likely it is that the same input, slightly perturbed, is classified with the same label as x, that is, $\pi(\phi_{\mathcal{N}}(x)) = \pi(\phi_{\mathcal{N}}(x + \varepsilon))$, for a small ε. In other words, the higher the absolute value of $\phi_{\mathcal{N}}$ for a given example x, the more robust the prediction of $\mathcal{N}$ for x.

Chen et al. (2019) propose a regularization term to remove *weak* connected components from the decision boundary, that is, connected components enclosing decision regions such that the predictions for their points are not robust. This regularization term for binary classification tasks can be computed from zero-dimensional zigzag persistence diagrams induced by $\phi_{\mathcal{N}}$ whenever the neural network function $\phi_{\mathcal{N}}$ is Morse, which occurs in several neural network architectures (Kurochkin, 2021), and $\mathcal{X}$ is a hypercube. This regularization term was the first to use the differentiability properties of persistent homology, studied and introduced in parallel by Leygonie et al. (2022) and by Carrière et al. (2021). The capacity of the term was studied for a simple kernel logistic regression, that was compared to other non-deep learning models such as KNN, logistic regression, or SVM, among others, trained with L^1 and L^2 regularization terms. The experiments were performed in synthetic and

real datasets, where the real datasets came from the UCI dataset bank (Kelly et al.) and from two biomedical datasets (Yuan et al., 2014; Ni et al., 2018). In most cases, the model trained with the topological regularizer outperformed the rest of the models. However, no deep learning models were involved, and a further experimentation is needed to see the efficacy of this approach for neural networks.

Decision boundaries of a hypothesis set

Previous articles have studied decision regions and boundaries for specific instances of neural networks, that is, pairs consisting of an architecture a and a particular set of parameters θ. Petri and Leitão (2020) study the topology of all possible decision boundaries given by a hypothesis set $\mathcal{F}_{a,\Theta}$ of a given fixed architecture a, and a given fixed set of possible parameters Θ. The objective of this study is to associate a *topological diversity measure* to a particular hypothesis set $\mathcal{F}_{a,\Theta}$ that characterizes the diversity in the topology of the different decision boundaries induced by the hypothesis set. To do this, a sample is first taken with a large number of possible parameters $\{\theta_1, \ldots, \theta_m\} \subseteq \Theta$. Then, for each parameter θ_i, $i \in [m]$, the zero and one-dimensional Vietoris–Rips persistence diagrams $D(\mathbb{V}_k(\mathrm{VR}(P_i, \|\cdot\|_2)))$ of a sample of points P_i from an approximation of the decision boundary of the neural network with architecture a and parameters θ_i are computed. Finally, topological diversity measures of the hypothesis set are given by the spread of the metric spaces (Willerton, 2015) whose points are the computed persistence diagrams for a fixed dimension and whose distance is a q-Wasserstein distance for a fixed $q \in \mathbb{N}$. To approximate spreads, which are computationally expensive to compute, Petri and Leitão saw that the averages of the 1-norms of the previously computed persistence diagrams, which are easier to compute, correlate with their spreads and thus also characterize the topological diversity of the hypothesis set.

Input space decomposition into convex polyhedra

Neural networks with a ReLU activation function $\varphi(x) = \max(0, x)$ determine a decomposition of the input space into convex polyhedra that assigns to each polyhedron in the decomposition a unique binary vector in such a way that two polyhedra share a facet if and only if their associated binary vectors differ exactly in one component. Liu et al. (2023b) use this polyhedral decomposition of the input space to infer the topology of manifolds embedded in the domain of a given ReLU neural network from a sample of their points using persistence diagrams. To compute persistence diagrams given the sample $S = \{x_i\}_{i=1}^m$ of the manifold $\mathbb{M}$, the proposed procedure computes the set of unique binary vectors $\mathfrak{B}(S) = \{b(x_i)\}_{i=1}^m$ associated with the points of the sample, where $b(x_i)$ is the binary vector associated with the convex polyhedra in which x_i lies, and then computes the Vietoris–Rips persistence diagrams $D(\mathbb{V}_k(\mathrm{VR}(\mathfrak{B}(S), h)))$ of $\mathfrak{B}(S)$ using the Hamming distance h, which counts the number of different bits in the two binary strings to compare. Although this is not directly related to neural network analysis, we firmly believe that this opens a way to study the structure and topology of the unknown data distribution for a specific learning problem by applying this method to large enough data samples.

Masden (2022) used polyhedral decompositions to study the topology of decision boundaries of small, non-trained, ReLU neural networks by attaching a polyhedral complex structure to the decomposition. Polyhedral complexes are similar to simplicial complexes, in that they are sets of polyhedra such that every face of a polyhedron of the complex is

also in the complex, and the intersection of any two polyhedra is either empty or a face of both. Masden found that polyhedral complexes derived from neural networks under some technical assumptions possess a dual structure known as *signed sequence cubical complexes*, and provided a polynomial time algorithm with respect to the number of non-input neurons to compute both complexes, yet exponential with respect to the input dimension. The decision boundary of the neural network corresponds to specific polyhedra, which in turn have dual representations in the signed sequence cubical complexes, from which homology groups for the decision boundary can be recovered under certain assumptions, such as the need of performing a one-point compactification of the decision boundary.

Differences were found in the structure of decision boundaries in terms of their neural network layer structures. In networks with a single hidden layer, the topology of decision boundaries exhibited remarkable consistency across various parameter ranges, including modifications to input data dimension and the number of hidden neurons. In contrast, networks with two hidden layers displayed more variability in the topological distribution of their decision boundaries.

Generative neural networks

TDA has also been used in generative deep learning. One of the most well-known models for generative deep learning is the generative adversarial network (GAN) (Goodfellow et al., 2014). Roughly, generative adversarial networks try to generate synthetic samples from a *real* data space $\mathcal{X} \subseteq \mathbb{R}^d$ distributed by an unknown function $\mathbb{P}_X$. To do this, generative adversarial models are composed of two neural networks, the generator $\mathcal{G}$, and the discriminator $\mathcal{S}$, with functions $\phi_{\mathcal{G}} \colon \mathcal{Z} \to \mathbb{R}^d$ and $\phi_{\mathcal{S}} \colon \mathbb{R}^d \to \{0, 1\}$, respectively, where $\mathcal{Z}$ is a space distributed by a known distribution $\mathbb{P}_Z$. We refer to $\mathcal{Z}$ as the noise space of the generative adversarial network. The generator is the neural network whose objective is to generate synthetic samples from samples from the noise space. On the other hand, the discriminator is the network that, given a sample, generated by $\mathcal{G}$ or sampled from the real data space $\mathcal{X}$, tries to distinguish between the two. The training of the generative adversarial network is performed by training the generator and the discriminator in an adversarial way, that is, the generator tries to fool the discriminator and the discriminator tries to distinguish between the synthetic and the real samples.

One of the problems with generative models is the evaluation procedure. Quantitative evaluation of the quality of a model is not straightforward. To this end, Khrulkov and Oseledets (2018) propose the *Geometry Score*, a persistent homology-based quality metric for generative adversarial networks. This metric is based on the following three assumptions: 1. The higher the quality of the generator, the more similar are the spaces $\operatorname{supp}(\mathbb{P}_X)$ and $\phi_{\mathcal{G}}(\operatorname{supp}(\mathbb{P}_Z))$; 2. $\operatorname{supp}(\mathbb{P}_X)$ is concentrated in a manifold $\mathcal{M}_X$ and $\phi_{\mathcal{G}}(\operatorname{supp}(\mathbb{P}_Z))$ is a manifold; 3. The higher the similarity between the spaces $\mathcal{M}_X$ and $\phi_{\mathcal{G}}(\operatorname{supp}(\mathbb{P}_Z))$, the higher the similarity between their topologies. Note that the first part of the second assumption is the manifold hypothesis, and it may not hold in general scenarios (Von Rohrscheidt and Rieck, 2023). However, this detail is irrelevant for persistent homology of point clouds, as it works for samples of spaces with or without a topology.

To compare both spaces $\mathcal{M}_X$ and $\phi_{\mathcal{G}}(\operatorname{supp}(\mathbb{P}_Z))$, Khrulkov and Oseledets compute persistence modules of witness filtrations (de Silva and Carlsson, 2004) for samples from $\mathcal{X}$ and $\phi_{\mathcal{G}}(\operatorname{supp}(\mathbb{P}_Z))$, asuming that the samples taken from $\mathcal{X}$ are also contained in $\mathcal{M}_X$. Witness