Hypothesis classes with a unique persistence diagram are nonuniformly learnable

Nicholas Bishop[1*], Thomas Davies[1*], and Long Tran-Thanh[2]

[1]University of Southampton, [2]University of Warwick

*NB and TD contributed equally to this work and should be considered joint first authors.

Motivation

- Persistence-based summaries are increasingly integrated into deep learning through topological loss functions or regularisers.
- The implicit role of a topological term in a loss function is to restrict the class of functions in which we are learning (the hypothesis class) to those with a specific topology.
- Although doing so has had empirical success, to the best of our knowledge there exists no result in the literature that theoretically justifies this restriction.
- Given a binary classifier in the plane with a Morse-like decision boundary, we prove that the hypothesis class defined by restricting the topology of the possible decision boundaries to those with a unique persistence diagram results in a nonuniformly learnable class of functions.
- In doing so, **we provide a statistical learning theoretic justification for the use of persistence-based summaries in loss functions.**

Topologically restricted hypothesis classes

In the context of machine learning, the *persistence map*, given by

$$\mathcal{PH}_k : X \mapsto D,$$

takes a set of points $X \subset \mathbb{R}^d$ and maps them to a *persistence diagram* D: a multiset in the extended plane that concisely represents the k-persistent homology (roughly, the topology of the points at all scales). Persistence diagrams and their embeddings, collectively referred to as *persistence-based summaries*, have been integrated into deep learning via topological loss or regularisation terms. For examples, see [1], [2], [3], [4], [5] and [6].

Given prior knowledge about the topology in the form of a persistence diagram D, topological loss terms implicitly restrict the decision boundaries to the class of functions

$$\mathcal{PH}^{-1}(D) = \{f : \mathcal{PH}(f) = D\}\,.$$

This class of functions has been studied by Curry [7]. Given a real-valued function $f : [0,1] \to \mathbb{R}$, we identify its graph as the decision boundary of a binary classifier $h_f : [0,1] \times \mathbb{R} \to \{0,1\}$ by

$$h_f((x,y)) = \begin{cases} 0, & f(x) \le y, \\ 1, & f(x) > y. \end{cases}$$

Thus we are studying the hypothesis class

$$\mathcal{H}_D = \{h_f : f \in \mathcal{PH}^{-1}(D)\}.$$

Nonuniform learnability

Consider the standard supervised machine learning setting, in which a learner, when given an input example $x \in \mathcal{X}$, from an input space $\mathcal{X}$, would like to accurately predict the corresponding target label $y \in \mathcal{Y}$, from an output space $\mathcal{Y}$. The learner has access to a finite sample, $S = \{(x_i, y_i)\}_{i=1}^m$, of training examples, sampled from a distribution, μ over $\mathcal{X} \times \mathcal{Y}$, of interest.

The goal of the learner is to select a hypothesis, $h : \mathcal{X} \to \mathcal{Y}$, belonging to the hypothesis class, $\mathcal{H}$, which achieves low expected error with respect to the distribution μ. The error of a given hypothesis rule h on a given example (x, y) is given via a loss function, $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$, by evaluating $\ell(h(x), y)$. More formally, the goal of the learner is to select a hypothesis which attains low risk:

$$L_\mu(h) = \mathbb{E}_{(x,y)\sim\mu}\left[\ell(h(x), y)\right].$$

Within the framework of Statistical Learning Theory, a hypothesis class which captures these goals is characterised by the following definition [8].

Definition 1 (Nonuniform learnability)**.** A hypothesis class $\mathcal{H}$ is nonuniformly learnable if there exists a learning algorithm, A, and a function $m_{\mathcal{H}}^{\mathrm{NUL}} : (0,1)^2 \times \mathcal{H} \to \mathbb{N}$ such that, for every $\epsilon, \delta \in (0,1)$ and for every $h \in \mathcal{H}$, if $m \ge m_{\mathcal{H}}^{\mathrm{NUL}}(\epsilon, \delta, h)$ then for every distribution μ, with probability at least $1 - \delta$ over the choice of $S \sim \mu^m$, it holds that $L_\mu(A(S)) \le L_\mu(h) + \epsilon$.

Main result

Following previous work on the fibre of the persistence map, we concern ourselves with binary classifiers on the plane with decision boundaries that are Morse-like real-valued functions on the unit interval with local minima at $x = 0, 1$, and the 0th persistence diagram [7]. Although this is clearly a restricted setting, it provides an initial justification for the use of persistence-based summaries in loss functions. We prove the following Theorem.

Theorem 2. Given a persistence diagram D, the hypothesis class of binary classifiers with decision boundaries defined by Morse-like real-valued functions on the interval with local minima at $x = 0, 1$ given by

$$\mathcal{H}_D = \{h_f : f \in \mathcal{PH}_0^{-1}(D)\}.$$

is nonuniformly learnable.

Conclusions

- Despite an increasing amount of work that integrates persistence-based summaries into loss functions for deep learning, before now there has been no theoretical justification for doing so.
- We use results on the fibre of the persistence map to show that restricting the hypothesis class for binary classifiers on the plane to functions with a unique 0th persistence diagram results in a nonuniformly learnable hypothesis class.

References

[1] C. Chen, X. Ni, Q. Bai, and Y. Wang, "Toporeg: A topological regularizer for classifiers," *CoRR*, vol. abs/1806.10714, 2018.

[2] R. B. Gabrielsson, B. J. Nelson, A. Dwaraknath, and P. Skraba, "A topology layer for machine learning," vol. 108, PMLR, 2020, pp. 1553–1563.

[3] C. Hofer, R. Kwitt, M. Niethammer, and M. Dixit, "Connectivity-optimized representation learning via persistent homology," vol. 97, PMLR, 2019, pp. 2751–2760.

[4] M. Moor, M. Horn, B. A. Rieck, and K. Borgwardt, "Topological autoencoders," *ArXiv*, vol. abs/1906.00722, 2019.

[5] J. Clough, N. Byrne, I. Oksuz, V. A. Zimmer, J. A. Schnabel, and A. King, "A topological loss function for deep-learning based image segmentation using persistent homology," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2020.

[6] X. Hu, F. Li, D. Samaras, and C. Chen, "Topology-preserving deep image segmentation," in *Advances in Neural Information Processing Systems 32*, 2019, pp. 5657–5668.

[7] J. Curry, "The fiber of the persistence map for functions on the interval," *Journal of Applied and Computational Topology*, vol. 2, pp. 301–321, 2018.

[8] G. M. Benedek and A. Itai, "Nonuniform learnability," in *Automata, Languages and Programming*, Springer Berlin Heidelberg, 1988, pp. 82–92.