# Learning a manifold from a teacher's demonstrations

**Pei Wang**
Rutgers University - Newark
peiwang@rutgers.edu

**Arash Givchi**
Rutgers University - Newark
arash.givchi@gmail.com

**Patrick Shafto**
Rutgers University - Newark
patrick.shafto@gmail.com

## Abstract

We consider the problem of learning a manifold from a teacher's demonstration. Extending existing approaches of learning from randomly sampled data points, we consider contexts where data may be chosen by a teacher. We analyze learning from teachers who can provide structured data such as individual examples (isolated data points) and demonstrations (sequences of points). Our analysis shows that for the purpose of teaching the topology of a manifold, demonstrations can yield remarkable decreases in the amount of data points required in comparison to teaching with randomly sampled points. We also discuss the implications of our analysis for learning in humans and machines.

## 1 Introduction

In machine learning, learners are assumed to operate in a relatively simplified problem space: data points sampled by a random process. Humans learn from a richer, stronger context. Two aspects that have received the most attention are the fact that data may be chosen by a more knowledgeable informant, such as a teacher [Shafto et al., 2014, Zhu, 2015], and that the data points may themselves be structured into sequences as in demonstration [Kuhl et al., 1997, Brand et al., 2002a]. In this paper, we consider how having teachers who select structured data may affect the complexity of learning.

Our goal is to understand theoretical bounds on learning manifolds and their topology from teaching via structured data, which we expect to inform debates in machine learning and human learning. We particularly investigate the topology of manifolds for several reasons. First, the decomposition of learning into grounded and more abstract aspects parallels common wisdom across human and machine learning, which have converged on hierarchical ("deep") models of learning. Second, teaching topology will prove to be data-efficient for manifold learning applications, such as clustering where only information about the global structure of manifold (e.g. number of connected components etc.) is needed. Third, teaching will be able to proceed without full knowledge of the geometry, and the requirement for the teacher can be relaxed by just knowing the homotopy type of the manifold.

We begin with preliminaries in Section 2. Section 3 provides results related to teaching the topology of manifolds via data points and demonstrations, showing that demonstrations can yield vastly more efficient teaching. Section 4 provides concluding discussions.

## 2 Preliminaries

In machine learning, the manifold assumption states that high dimensional data in the real world are typically concentrated on a much lower dimensional manifold. Because the difficulty of inferring the

Topological Data Analysis and Beyond Workshop at the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

geometry of an arbitrary manifold is bounded by its worst local feature, quantified as the *reach* of the manifold, we may only aim to reduce the sample complexity of learning a manifold by focusing on its global properties, which are encoded by the topology.

In this paper, $\mathcal{M}$ is an orientable compact sub-manifold in $\mathbb{R}^n$. We mainly focus on low dimensional manifolds such as curves and surfaces. However, teaching methods developed in the following sections can be directly used to convey low dimensional topological features of any manifold.

For the formalism of teaching-learning algorithms, we consider $\mathcal{A}$ as a class of **learning algorithms** that construct approximations of $\mathcal{M}$ and/or identify the homotopy type of $\mathcal{M}$ from a set of data points sampled from $\mathcal{M}$. Examples of such algorithms are available in Cheng et al. [2005], Niyogi et al. [2008], Boissonnat and Ghosh [2014].

A collection of data points $\mathcal{D}_\mathcal{A} \subset \mathcal{M}$ is called a **teaching set** with respect to $\mathcal{A}$ if there exists a learning algorithm $A \in \mathcal{A}$ that recovers the homotopy type of $\mathcal{M}$ using $\mathcal{D}_\mathcal{A}$. $|\mathcal{D}_\mathcal{A}|$ denotes the size of $\mathcal{D}_\mathcal{A}$. A teaching set $\mathcal{D}^*_\mathcal{A}$ is said to be *minimal w.r.t.* $\mathcal{M}$ if $|\mathcal{D}^*_\mathcal{A}| \leq |\mathcal{D}_\mathcal{A}|$ for any teaching set $\mathcal{D}_\mathcal{A}$ of $\mathcal{M}$. Further, $\mathcal{D}^*_\mathcal{A}$ is a *minimal teaching set w.r.t. the homotopy type of* $\mathcal{M}$ if $\mathcal{D}^*_\mathcal{A}$ is a teaching set for some $\mathcal{M}'$ of the same homotopy type as $\mathcal{M}$ and $|\mathcal{D}^*_\mathcal{A}| \leq |\mathcal{D}_\mathcal{A}|$ for any teaching set $\mathcal{D}_\mathcal{A}$ of a manifold homotopy equivalent to $\mathcal{M}$. The size of a minimal teaching set is called the **minimal teaching number**.

## 3 Structured data and manifold teaching

Our approach is inspired by human teaching together with their corresponding class of learners: examples [Shepard et al., 1961] and demonstrations [Brand et al., 2002b]. Inspired by human teaching, in this section, we propose two corresponding styles of methods to teach the topology of a manifold using structured data: isolated data points (individual examples) and sequential data points (demonstrations). In particular, we provide lower bounds for teaching complexity of each method. pairs of points (comparisons),

### 3.1 Manifold teaching from sample points

The pioneering work in Niyogi et al. [2008] introduced a framework to reconstruct manifolds from random sampling. Their work can be rephrased as a manifold teaching problem. Suppose two agents, which we call a teacher and a learner, wish to communicate a manifold $\mathcal{M} \subset \mathbb{R}^n$. In their setting, the teacher passes a collection of *randomly* sampled data points $\mathcal{D}_{\mathcal{A}(\epsilon)} = \{x_1, \ldots, x_k\}$ to the learner, who then builds a manifold by a learning algorithm in the **class** $\mathcal{A}(\epsilon)$: the learner first picks a parameter $\epsilon \in \mathbb{R}^+$, then for each $x_i \in \mathcal{D}_{\mathcal{A}(\epsilon)}$, makes an n-dimensional ball $B_\epsilon(x_i)$ centered at $x_i$ of radius $\epsilon$. Here $n$ is the dimension of the ambient space which can be inferred from data points' coordinate size. The union of all these balls $U_\epsilon(\mathcal{D}_{\mathcal{A}(\epsilon)}) = \cup_{x \in \mathcal{D}_{\mathcal{A}(\epsilon)}} B_\epsilon(x)$ constitutes the learned space.

The main result in Niyogi et al. [2008] provides an estimation $N_\epsilon$ on the number of data that are needed to guarantee that the learned space $U_\epsilon$ and the target manifold $\mathcal{M}$ are homotopy equivalent with a given confidence. $N_\epsilon$ depends on the confidence level, the volume and the reach of $\mathcal{M}$, and also the learner's choice of $\epsilon$.

Considering $N_\epsilon$ as a sufficient bound on the minimal teaching number of $\mathcal{M}$, we seek a necessary condition. The calculation of $N_\epsilon$ proposed in Niyogi et al. [2008] requires knowledge about the geometric features of $\mathcal{M}$ (volume and reach), which translated to our context implies that either the teacher knows the true $\mathcal{M}$ or the teacher has observed a large amount of data points, which allows one to make good estimations of these geometric features. Hence we will start our analysis by assuming that the teacher has access to the true manifold. In Section 4, we will discuss how this assumption can be relaxed in many practical cases.

Suppose that the learner uses the class of algorithms $\mathcal{A}(\epsilon)$, what is a *minimal teaching set* to convey the homotopy type of $\mathcal{M}$? The case for non-contractible 1-dim manifolds is extremely neat. Since every such $\mathcal{M}$ is homotopy equivalent to a circle, at least three points are needed as explained below.

**Example 3.1.** Let $S^1$ be a unit circle embedded in $\mathbb{R}^2$, and $\mathcal{A}(\epsilon)$ be the class of learning algorithms described above. It is clear that any data set with only one or two points will result contractible $U_\epsilon$ for any choice of $\epsilon$. However, as illustrated in Figure 1(a), with three equidistant points sampled from $S^1$,

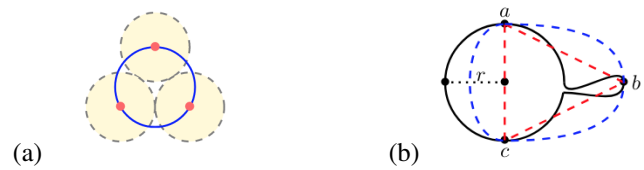



Figure 1: Teaching sets for 1-dim manifolds

any learner $\mathcal{A}(\epsilon)$ with $\frac{\sqrt{3}}{2} \leq \epsilon < 1$ will recover the correct topology of $S^1$ from the union of three connected disks with a hole in the middle. Thus the minimal teaching number for a circle is three.

Suppose that $\mathcal{M}$ is a closed orientable surface. An example is given below.

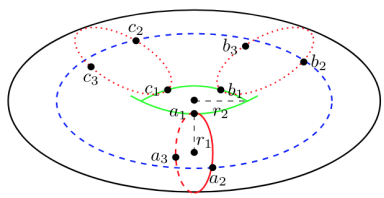



Figure 2: Teaching set for torus.

**Example 3.2.** Let $T^2$ be a torus embedded in $\mathbb{R}^3$ as shown in Figure 2. $T^2$ can be obtained by rotating the red circle $l_1$ around the green circle $l_2$. Denote the radii of $l_1$ and $l_2$ by $r_1$ and $r_2$ respectively. Two 1-dim holes of $T^2$ are represented by $l_1$ and $l_2$. As in Example 3.1, each $l_i$ needs at least three teaching points. Since the learner $\mathcal{A}(\epsilon)$ picks one $\epsilon$ for all data points, more data points are needed for $l_i$ when $\frac{\sqrt{3}}{2} r_i \geq r_j$, where $i, j \in \{1, 2\}$ and $i \neq j$. Hence to find the minimal teaching set for the homotopy type of $T^2$, we may assume that $r_1 = r_2 = r$. Suppose that any three data points sharing a circle in Figure (b) are equidistant points. Then $\mathcal{D}_1 = \{a_1, a_2, a_3, b_1, c_1\}$ can be used to teach $l_1$ and $l_2$. To recover the only 2-dim hole of $T^2$, it is natural to add $\{b_2, b_3, c_2, c_3\}$ into $\mathcal{D}_1$ to complete the red dotted circles going through $b_1$ and $c_1$. Ideally, $\epsilon$-balls centered at these 9 points should form a torus. However, there are large undesirable gaps left open between the red circles because the learner is restricted to pick $\frac{\sqrt{3}}{2} r < \epsilon < r$.

We now compute how many extra points are needed to fill in all these gaps. Direct calculation shows that the radius of the dashed blue circle $l_3$ is $2.5r$ and nine equidistant data points on $l_3$ are needed to teach it with $\frac{\sqrt{3}}{2} r < \epsilon < r$. If we rotate $l_1$ around $l_2$ nine times with each step $2\pi/9$, then the trace of $\{a_1, a_2, a_3\}$ produces 27 data points (including all 9 points in $\mathcal{D}_1$). With these 27 points, we almost form a torus but still have many small gaps. One may count that in total there are 27 such gaps. So 54 points are enough. Moreover, notice that the inner green circle $l_2$ is over taught, one may check that 3 teaching points can be removed from $l_2$. Hence we may teach $T^2$ with 51 points.

The approach we used in Example 3.2 can be generalized to all orientable surfaces.

**Proposition 3.3.** Let $\mathcal{M}_g \subset \mathbb{R}^3$ be a closed orientable surface with genus $g$. Then the *minimal teaching number* for the *homotopy type* of $\mathcal{M}_g$ with respect to $\mathcal{A}(\epsilon)$ is bounded by $49g + 2$.

*Proof.* We will proceed by induction. When $g = 1$, $\mathcal{M}_1$ is homotopy equivalent to $T^2$. So the homotopy type of $\mathcal{M}_1$ can be taught by 51 points. Suppose that the claim holds for any $\mathcal{M}_g$ with $g < n$. Notice that $M_{g,1}$, surface with genus $g$ and 1 boundary component, can be obtained from $M_g$ by removing a disk. So when $g < n$, based on the inductive assumption, $M_{g,1}$ can be taught by $49g + 1$ points. Moreover, there exists a $\mathcal{M}_n$ which can be obtained by gluing a $M_{1,1}$ with a $\mathcal{M}_{n-1,1}$. Hence we may teach $\mathcal{M}_n$ with $[49 + 1] + [49(n - 1) + 1] = 49n + 2$ data points. $\square$

**Remark 3.4.** Let $\mathcal{M}_{g,b} \subset \mathbb{R}^3$ be a genus $g$ orientable surface with $b$ boundary components. Note that $\mathcal{M}_{g,b}$ can be obtained from $\mathcal{M}_g$ by removing $b$ disconnect disks. Therefore, the *minimal teaching number* of $\mathcal{M}_{g,b}$ with respect to $\mathcal{A}(\epsilon)$ is bounded by $49g + 2 - b$.

The above analysis suggests that due to the locally Euclidean nature of manifolds, a local to global teaching procedure as described above is not always efficient: even a simple manifold as a regular torus requires a large set of teaching points. Below we propose a new class of teaching algorithms that teaches the topology directly from demonstrations where each demonstration is a sequence of data describing a loop.

### 3.2 Manifold teaching from demonstrations

A main task in manifold teaching is passing the correct topology. This task forces large amount of data for any local to global teaching procedure due to the locally Euclidean nature of manifolds. In this section we will describe a method that teaches the topology directly from demonstrations where each demonstration is a sequence of data describing a loop.

Teaching with sequential data is efficient because topologies of manifolds are intuitively captured by loops in various dimensions. As in Example 3.1, a unit circle can be taught by three points. In fact, any (oriented) [1] non-contractible 1-dim manifold can be effectively described by a sequence consisting three *randomly sampled* points. For example, the teacher may teach the black loop $\mathcal{M}_b$ in Figure 1(b) by $\mathcal{D} = \{[a, b, c, a]\}$. The sequential data informs the learner to connect consecutive points by a simple curve[2], which will form a space $U_{\mathcal{D}}$ that is homotopy equivalent to $\mathcal{M}_b$. The red and blue curves in Figure 1(b) are two examples of $U_{\mathcal{D}}$ obtained by different learners. If curves that connect data points further adopt some mild assumption, for example smoothness, polynomial, linear etc., then the obtained space $U_{\mathcal{D}}$ can be parameterized accordingly.

**Remark 3.5.** It is important to note that calculations in this section are done for a particular target manifold $\mathcal{M}$, whereas in the previous section were done for the homotopy type of $\mathcal{M}$. For instance, given an arbitrary non-contractible 1-dim manifold $\mathcal{M}^1$, any sequence with three points randomly sampled from $\mathcal{M}^1$ forms a teaching set for $\mathcal{M}^1$. However for learners using $\mathcal{A}(\epsilon)$ (Section 3.1), it is true that the minimal teaching size for the homotopy type of $\mathcal{M}^1$ w.r.t. $\mathcal{A}(\epsilon)$ is three, but it is possible that there does not exist three points on $\mathcal{M}^1$ form a teaching set for $\mathcal{M}^1$ based on $\mathcal{A}(\epsilon)$.

Higher dimensional manifolds can also be conveyed using sequential data. In this setting, the teacher passes a *sequence of sequences* to the learner. The learner builds a manifold by a learning algorithm in the **class** $\mathcal{A}(\mathbf{l})$ where $\mathbf{l}$ indexes all the choices: for each sequence $\mathbf{a} = [a_1, \ldots, a_m]$, the learner connects consecutive points by a curve; for each sequence of sequences $[\mathbf{a}, \mathbf{b}, \ldots, \mathbf{c}]$ where $\mathbf{a}, \mathbf{b}, \ldots, \mathbf{c}$ are sequences of data points with the same length, the learner connects consecutive sequences by curved planes (the curvature is not necessarily zero). In more details, as Figure (d), with $\mathbf{a} = [a_1, a_2, a_3, a_1], \mathbf{b} = [b_1, b_2, b_3, b_1]$, the learner first connects points in $\mathbf{a}$ and $\mathbf{b}$ separately to form two loops (shown as black circles). Furthermore, to join $\mathbf{a}$ and $\mathbf{b}$, the learner may link each distinct pair of $a_i$ and $b_i$ by a curve [3] (shown as dished lines). This completes a closed path through $a_i, a_{i+1}, b_{i+1}$[4] and $b_i$, for $i = 1, 2, 3$. Then the learner glues a curved plane along each of these closed path. For example, the gray area $C$ in Figure 3(a) shows a curved plane glued along the red closed path going through $a_1, a_2, b_2, b_1$.

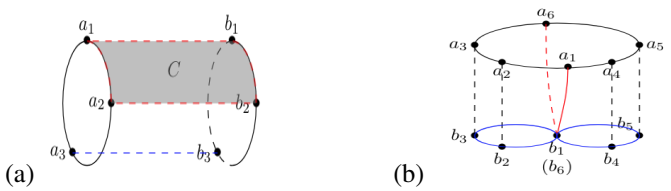



Figure 3: Teaching by sequential data

To make the learning algorithm $\mathcal{A}(\mathbf{l})$ robust to the choice of curves and planes, we further assume that the teacher and the learner agree that $(*)$: (1) there is no intersection between different connecting curves except end points; (2) two points are connected by at most one curve. For instance, a pair

[1] Without considering the orientation, a sequence with two points can also describe a loop.

[2] A curve is simple if it has no self-intersection.

[3] If $a_i$ and $b_i$ are the same point, do nothing.

[4] $a_4 = a_1$ and $b_4 = b_1$.

of pants can be taught by $\mathcal{D}_{\text{pants}} = \{_{[[a_1,a_2,a_3,a_6][b_1,b_2,b_3,b_6]],[[a_1,a_4,a_5,a_6][b_1,b_4,b_5,b_6]]}\}$ as shown in Figure 3(b). With the teaching set $\mathcal{D}_{\text{pants}}$, the learner needs to connect $a_1, b_1$ and $a_6, b_6$ multiple times. If the learner makes the connection by the red curves for the first time, assumption $(*)$ ensures the learner always picks the red curves during the entire learning process.

**Example 3.6.** The torus in Example 3.2 can be taught by a sequence of four sequences:

$\mathcal{D}_{\text{torus}} = [[a_1, a_2, a_3, a_1][b_1, b_2, b_3, b_1][c_1, c_2, c_3, c_1][a_1, a_2, a_3, a_1]]$ as shown in Figure (b). This teaching set only contains the 9 basic points which fits our initial intuition.

**Proposition 3.7.** Let $\mathcal{M}_g \subset \mathbb{R}^3$ be a closed orientable surface with genus $g \geq 2$. Then the *minimal teaching number* of $\mathcal{M}_g$ with respect to $\mathcal{A}(1)$ is bounded by $3g - 3$ sequences, where each sequence consists of at most $4$ data points.

*Proof.* A classical result of surfaces states that for any $\mathcal{M}_g$, there is a system of $3g-3$ disjoint simple closed curves which cut $\mathcal{M}_g$ into pairs of pants (see for example, Farb and Margalit [2011]). Note that each simple closed curve can be taught by a sequence of $4$ points; each pair of pants can be taught by a sequence that consists of three sequences representing its boundary curves. Moreover, two legs of a pair of pants can be glued along their boundary curves through a sequential data. For instance, two blue boundary curves in Figure 3(b) can be glued by $[[b_1, b_2, b_3, b_1][b_1, b_4, b_5, b_1]]$. Hence the claim holds. □

## 4 Conclusions

We considered the problem of teaching low-dimensional features of a manifold using structured data, which extends mathematical approaches to learning manifolds toward the contexts more consistent with the richness of human learning. Building on prior work in manifold learning, we formalize teaching manifolds from data points, observe that contrary to intuition, teaching does not facilitate learning as much as one would expect due to constraints imposed by the reach of the manifold. Considering learning from demonstrations—sequences of data points—we show that learning can be greatly facilitated by teaching. This approach relies on separating teaching the geometry of the manifold itself from teaching its topology. Focusing on teaching only the topology, we show that sequences of points can be used to represent the homology groups of the manifold, which compactly capture important abstract structure that can be used to facilitate future learning. Moreover, this relaxes the overly stringent and implausible requirement that the teacher must know the manifold exactly to an estimation of its homotopy type, which is almost always less stringent than the true manifold. A preliminary example where the teacher teaches with partial knowledge and unconstrained data is included in the Supplementary material (See Section A). Finally, we argue for a connection to human learning from teaching demonstrations, which are most naturally thought of as sequences. Future work may extend this approach toward more naturalistic learning problems faced by humans or solved by machine learning. The approaches are not restricted to manifold teaching and it would be interesting to explore teaching more general mathematical objects with low dimensional topological structures, such as graphs, CW-complexes and even groups.

## Acknowledgments and Disclosure of Funding

Research was sponsored by the Defense Advanced Research Projects Agency (DARPA) and the Army Research Office (ARO) and was accomplished under Cooperative Agreement Number W911NF-20-2-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the DARPA or ARO, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

The research was also supported by NSF MRI 1828528 to PS.

## References


Jean-Daniel Boissonnat and Arijit Ghosh. Manifold reconstruction using tangential delaunay complexes. *Discrete & Computational Geometry*, 51(1):221–267, 2014.

Rebecca J Brand, Dare A Baldwin, and Leslie A Ashburn. Evidence for 'motionese': modifications in mothers' infant-directed action. *Developmental Science*, 5(1):72–83, 2002a.

Susan Brand, Rita Dunn, and Fran Greb. Learning styles of students with attention deficit hyperactivity disorder: Who are they and how can we teach them? *The Clearing House*, 75(5):268–273, 2002b.

Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

Gunnar Carlsson, Tigran Ishkhanov, Vin De Silva, and Afra Zomorodian. On the local behavior of spaces of natural images. *International journal of computer vision*, 76(1):1–12, 2008.

Frédéric Chazal and Bertrand Michel. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *arXiv preprint arXiv:1710.04019*, 2017.

Frédéric Chazal, Leonidas J Guibas, Steve Y Oudot, and Primoz Skraba. Persistence-based clustering in riemannian manifolds. *Journal of the ACM (JACM)*, 60(6):41, 2013.

Siu-Wing Cheng, Tamal K Dey, and Edgar A Ramos. Manifold reconstruction from point samples. In *SODA*, volume 5, pages 1018–1027, 2005.

Benson Farb and Dan Margalit. *A primer on mapping class groups (pms-49)*. Princeton University Press, 2011.

Christoph Hofer, Roland Kwitt, Marc Niethammer, and Andreas Uhl. Deep learning with topological signatures. In *Advances in Neural Information Processing Systems*, pages 1634–1644, 2017.

Patricia K Kuhl, Jean E Andruski, Inna A Chistovich, Ludmilla A Chistovich, Elena V Kozhevnikova, Viktoria L Ryskina, Elvira I Stolyarova, Ulla Sundberg, and Francisco Lacerda. Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277(5326):684–686, 1997.

Chunyuan Li, Maks Ovsjanikov, and Frederic Chazal. Persistence-based structural recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1995–2002, 2014.

Clément Maria, Jean-Daniel Boissonnat, Marc Glisse, and Mariette Yvinec. The gudhi library: Simplicial complexes and persistent homology. In *International Congress on Mathematical Software*, pages 167–174. Springer, 2014.

Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1):419–441, 2008.

Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. A stable multi-scale kernel for topological machine learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4741–4748, 2015.

Patrick Shafto, Noah D. Goodman, and Thomas L. Griffiths. A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, 71:55–89, 2014. ISSN 00100285. doi: 10.1016/j.cogpsych.2013.12.004.

Roger N Shepard, Carl I Hovland, and Herbert M Jenkins. Learning and memorization of classifications. *Psychological monographs: General and applied*, 75(13):1, 1961.

Xiaojin Zhu. Machine teaching: An inverse problem to machine learning and an approach toward optimal education. In *AAAI*, pages 4083–4087, 2015.

# A Learning from teacher with partial knowledge

All teaching methods discussed in the main text assume teacher has full access to the true manifold. However, in reality, the teacher often does not know the underlying manifold and often does not have full control over which data can be used to teach. In this section, we consider teaching in a much more practical scenario that allows a teacher, who may have limited knowledge, to teach with unconstrained data. We illustrate how this would assist the learner to improve their estimation of the relevant topological and geometrical information from the data.

Following the standard setting of *topological data analysis* Carlsson [2009], Chazal and Michel [2017], we assume that the data $\mathcal{D}$ is a finite set of points sampled from the true manifold $\mathcal{M}$. Using an algorithm in **class** $\mathcal{A}(\epsilon)$ (Sec 3.1) with different $\epsilon$'s, the learner obtains a summary of estimations of $\mathcal{M}$ in form of *persistent homology*. Rather than picking a teaching set directly from $\mathcal{M}$, the teacher first selects a subset $\mathcal{D}_T$ from $\mathcal{D}$, then passes $\mathcal{D}_T$ to the learner in a proper sequential format according to algorithm $\mathcal{A}(\mathbf{l})$ (Sec 3.2) to demonstrate desired topological features of $\mathcal{M}$.

As discussed before, the difficulty of learning a manifold $\mathcal{M}$ increases dramatically as the reach of $\mathcal{M}$ drops. Now we illustrate how teaching helps in these situations by the following example.

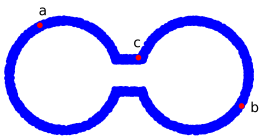



Figure 4: A barbell shaped annulus.

**Example A.1.** Let the true manifold $\mathcal{M}$ be the blue barbell shaped annulus shown in Figure 4 with reach $\tau = 0.26$. Assume that the learner analyzes randomly sampled data by TDA and the teacher knows that $\mathcal{M}$ contains a 1-dim hole. Based on $\mathcal{A}(\mathbf{l})$, three distinct points are required to form a teaching sequence for this hole. When fewer than three data points are observed by the learner, the teacher would simply wait until more data were collected. Suppose that the learner gets three data $\mathcal{D}_1 = \{a, b, c\}$ as shown. The corresponding persistence barcode of $\mathcal{D}_1$ is empty for $H_1$ (no 1-dim loop is ever formed for any choice of $\epsilon$). With $\mathcal{D}_1$, the teacher may teach by marking these points sequentially as for example $[a, b, c, a]$. Comparing the teacher's demonstration with the barcode, the learner would realize that $\mathcal{M}$ is homotopy equivalent to a circle and currently points gathered are not sufficient to extract any accurate geometrical information.

Further suppose that the learner intends to estimate the geometry of $\mathcal{M}$ and so more points are sampled. A given data set $\mathcal{D}$ is called *feasible*, if the learner is able to derive the true geometry of $\mathcal{M}$ from $\mathcal{D}$ with some $\epsilon$, i.e. if there exists $\epsilon^* < \tau = 0.26$ such that $U_{\epsilon*}(\mathcal{D})$ is homotopy equivalent to $\mathcal{M}$ [5]. To estimate the lower bound on size of a feasible data set, we randomly sample data sets from $\mathcal{M}$ with increasing sizes and 20 simulations for each size. Empirically it shows that feasible data sets appear only after $|\mathcal{D}| > 150$ and and appears in every simulation for $|\mathcal{D}| \geq 500$.

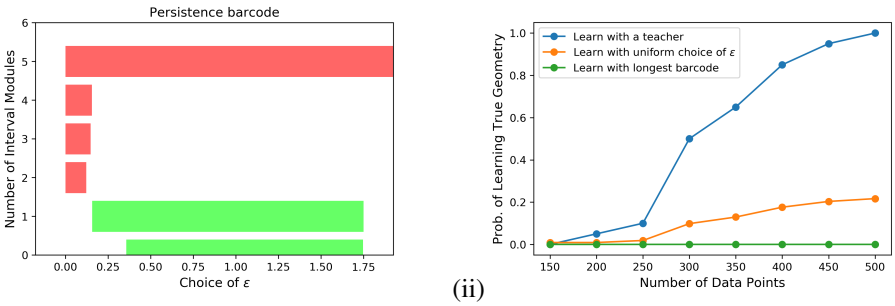
Figure 5(i) shows the persistence barcode for a data set of size 500.[6] The red bars are the longest four intervals for $H_0$, which reflects the number of connected components. After $\epsilon > 0.156$, only one red bar remains which indicates $U_\epsilon$ contains a single component for any $\epsilon > 0.156$. The green bars are the intervals for $H_1$ (ignoring intervals of length less than 0.05), which represents the number of 1-dim holes. The top green bar spans over $(0.158, 1.751)$ and indicates that there is a 1-dim loop forms at $\epsilon = 0.158$ and persists until $\epsilon = 1.751$. The bottom green bar spans over $(0.357, 1.747)$ and indicates that another 1-dim loop forms at $\epsilon = 0.357$ and persists until $\epsilon = 1.747$. All randomly sampled data sets of size 500 exhibit similar persistence barcode with two long intervals for $H_1$ as shown. Focusing on the range of $\epsilon$ where 1-dim holes exist, on average 78% choice of $\epsilon$ (with variance 0.0002) indicates two 1-dim loops over all simulations. Thus, without teaching, the learner would likely to conclude a wrong topological information, $H_1(\mathcal{M}) = \mathbb{Z}_2 \times \mathbb{Z}_2$, with high confidence. In contrast, with a teaching set of three points, the learner is able to not only infer the correct topology

[5] It is possible that $U_\epsilon(\mathcal{D})$ is homotopy equivalent to $\mathcal{M}$ for $\epsilon > \tau$. However the top and the bottom of the narrow middle part of $\mathcal{M}$ will be connected up in such $U_\epsilon(\mathcal{D})$, which leads to wrong geometry.

[6] The barcode was constructed using the GUDHI library Maria et al. [2014].

immediately after teaching but also accurately estimate the geometry of $\mathcal{M}$ by focusing on $U_\epsilon$ with $0.158 < \epsilon < 0.357$.

Figure 5(ii) plots the average learning accuracy of $\mathcal{M}$'s geometry for different types of learners. The blue curve shows the learners with a topological teacher who are assumed to follow a Bernoulli distribution since they are able to infer the correct geometry with every feasible data set. The orange curve is corresponding to learners who choose $\epsilon$ uniformly from the interval where barcode for $H_1(\mathcal{M})$ is not empty (Variances are omitted as their magnitudes are bounded above by $0.01$). The green curve shows learners, who approximate $\mathcal{M}$ by $U_\epsilon$ with the most persistent homology, stay incorrect on geometry even with increasing data size. Clearly learner's acquisition of geometry are accelerated by teaching topology.




(i) (ii)

Figure 5: Learning based on TDA

Persistent homology has started to attract attention in machine learning Carlsson et al. [2008], Chazal et al. [2013], Li et al. [2014], Reininghaus et al. [2015]. However, levering these topological features for learning poses considerable challenges because the relevant topological information is not carried by the whole persistence barcode but is concentrated in a small region of $\epsilon$ that may not be obvious Hofer et al. [2017]. Teaching by demonstration resolves these challenges by allowing the the learner to extract the most suitable topological information after the correct homology appears in the persistence barcode, and zooming the analysis of $\mathcal{M}$'s geometry into the most appropriate range of $\epsilon$ with high data efficiency.

More importantly, teaching by demonstration allows *accumulation* of information across learners, whereas other forms of teaching can only transmit information from an already knowledgeable teacher. As pointed out in Sec 3.1, the method of teaching by sampling points essentially assumes that the teacher knows the true manifold $\mathcal{M}$. However, given the intractability of manifold learning in general, there is no plausible way for the teacher to have access to $\mathcal{M}$. On such accounts, teaching does not resolve the true challenge of learning and instead passes off the problem to a teacher for whom the learning problem does not exist. The key advantage of teaching from demonstrations is that it allows the teacher to convey critical information of $\mathcal{M}$ without knowing the entire manifold. For example, let $\mathcal{M}$ be a torus as in Figure (b) with $r_1 << r_2$. The teacher may only have enough observations to conclude that there is a loop homotopy equivalent to the green circle $l_2$. With sequential data, the teacher could easily pass the only loop $l_2$ he observed, which allows the learner to focus on the region of $\epsilon$ where $l_2$ exists.

In addition, from a teacher's perspective, much less data is needed to learn the topology of an irregular manifold $\mathcal{M}$ than its geometry. For instance, let $\mathcal{M}$ be the 1-dim manifold shown in Figure (c). Denote the reach of $\mathcal{M}$ by $\tau$ and the radius of the left arc in $\mathcal{M}$ by $r$. Note that the teacher only needs $r$-dense data to learn the topology of $\mathcal{M}$, whereas $\tau$-dense data to learn the geometry. In fact, for any manifold $\mathcal{M}$, we may define its **topological reach** $\eta$ to be the largest number such that $U_\epsilon(\mathcal{M})$ is homotopy equivalent to $\mathcal{M}$ for any $\epsilon \leq \eta$, where $U_\epsilon(\mathcal{M}) = \cup_{\{p \in \mathcal{M}\}} B_\epsilon(p)$. According to Proposition 3.2 in Niyogi et al. [2008], for the same confidence level, points needed to achieve $\epsilon$-dense is polynomial increasing with $1/\epsilon$. Therefore when $\mathcal{M}$ is irregular, i.e. $\tau$ is significantly less than $\eta$, the amount of data needed to achieve $\eta$-dense is much fewer than $\tau$-dense. Since the topology of $U_\epsilon(\mathcal{M})$ remains the same for data beyond $\eta$-dense, it requires much less data to learn the topology of an irregular manifold than its geometry.