

---

# Topological Data Analysis of Decision Boundaries with Application to Model Selection Supplemental Material

---

Karthikeyan Natesan Ramamurthy<sup>1</sup> Kush R. Varshney<sup>1</sup> Krishnan Mody<sup>1,2</sup>

In this supplementary material, the references to the main paper are suffixed by “-M”.

## 1. Proof of Theorem 1

*Proof.* Let us assume that among the  $n$  samples  $\bar{z}$  drawn,  $|\bar{x}| = n_x$  and  $|\bar{y}| = n_y$ , so that  $n = n_x + n_y$ . Let us denote the event  $x \notin A_i$  for any  $i$  as  $E_i^a$  and the event  $y \notin B_j$  for any  $j$  as  $E_j^b$ . The probability of these events are

$$P(E_i^a | |\bar{x}| = n_x) = (1 - \mu_x(A_i))^{n_x} \leq (1 - \alpha_x)^{n_x} \quad \text{and} \quad (1)$$

$$P(E_j^b | |\bar{y}| = n_y) = (1 - \mu_y(B_j))^{n_y} \leq (1 - \alpha_y)^{n_y}. \quad (2)$$

The probability bound on the composite event (4-M) is expressed as

$$P\left(\left(\bigcap_i \overline{E_i^a}\right) \cap \left(\bigcap_j \overline{E_j^b}\right)\right) > 1 - \delta, \quad (3)$$

which simplifies to

$$P\left(\overline{\left(\bigcup_i E_i^a\right) \cup \left(\bigcup_j E_j^b\right)}\right) > 1 - \delta. \quad (4)$$

This implies that

$$P(\bigcup_i E_i^a) + P(\bigcup_j E_j^b) \quad (5)$$

should be bounded from above by  $\delta$ . The individual conditional probabilities can be union-bounded as  $P(\bigcup_i E_i^a | |\bar{x}| = n_x) \leq l_a(1 - \alpha_x)^{n_x}$  and  $P(\bigcup_j E_j^b | |\bar{y}| = n_y) \leq l_b(1 - \alpha_y)^{n_y}$ . Hence, the upper bound on (5) is

$$\sum_{n_x=0}^n P(\bigcup_i E_i^a | |\bar{x}| = n_x) p(|\bar{x}| = n_x) + P(\bigcup_j E_j^b | |\bar{y}| = n - n_x) p(|\bar{y}| = n - n_x), \quad (6)$$

---

<sup>1</sup>IBM Research, Yorktown Heights, NY, USA <sup>2</sup>Courant Institute, New York University, New York City, NY, USA. Correspondence to: Karthikeyan Natesan Ramamurthy <knatesa@us.ibm.com>.

which, after some algebra, simplifies to

$$l_a(1 - q\alpha_x)^n + l_b(1 - (1 - q)\alpha_y)^n. \quad (7)$$

We need to find an  $n$  such that the expression in (7) is bounded above by  $\delta$ .

Since we know  $1 - \alpha q \leq \exp(-\alpha q)$  using Taylor approximation, when  $n > \frac{1}{\alpha q}(\log 2l + \log(\frac{1}{\delta}))$ ,  $l \exp(-\alpha q n) \leq \delta/2$ . Hence if we pick  $n$  according to (3-M), (5) will be  $\leq \delta$ , and with probability greater than  $1 - \delta$ , we can ensure (4-M).  $\square$

## 2. Necessity of Decision Boundary Topology

When understanding the decision surfaces of labeled data, using the approach presented in our paper is more accurate than using the unlabeled data topology for either of the classes as pursued in Bianchini & Scarselli (2014) and Guss & Salakhutdinov (2018). We use a simple counter-example in Figure 1 to demonstrate this. Clearly, neither of the classes reflect the true topology of the decision boundary.

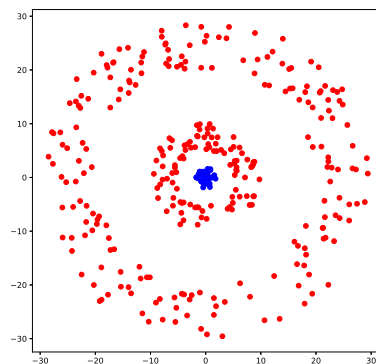


Figure 1. The blue class has Betti numbers  $\beta_0 = 1, \beta_1 = 0$ . The red class has Betti numbers  $\beta_0 = 2, \beta_1 = 2$ . Neither of them reflect the true topology of the decision boundary which has Betti numbers  $\beta_0 = 1, \beta_1 = 1$

### 3. Background on Persistent Homology for Unlabeled Data

Consider a set of  $n$  data points in  $\mathbb{R}^d$ :  $\mathcal{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ . A set of points by itself has no shape per se, but if the points are viewed as samples from some shape, then the set of points reveals the underlying shape. We would like to estimate and approximate the topology of that shape by constructing a simplicial complex from the points and examining the topology of the simplicial complex. A zero-dimensional simplex is a point, a one-dimensional simplex is a line segment, a two-dimensional simplex is a triangle, a three-dimensional simplex is a tetrahedron, and so on; a simplicial complex is a set of simplices glued together in a particular way. Specifically, a simplicial complex  $\mathcal{S} = (\mathcal{Z}, \Sigma)$ , where  $\Sigma$  is a family of non-empty subsets of  $\mathcal{Z}$  such that each subset  $\sigma \in \Sigma$  is a simplex. Furthermore, the following condition must also hold:  $\sigma \in \Sigma$  and  $\tau \subseteq \sigma$  implies that  $\tau \in \Sigma$ . In forming these non-empty subsets of points that form a simplex, we only consider subsets of points that are close to each other. There are various notions of closeness that we come back to later in this section.

Topology, being the study of shape, is primarily concerned with the number of connected components and the number and dimension of holes that an object has. The Betti numbers characterize the connectivity as follows. The zeroth Betti number  $\beta_0$  is the number of connected components, the first Betti number  $\beta_1$  is the number of one-dimensional holes or circles, the second Betti number  $\beta_2$  is the number of two-dimensional voids or cavities, and so on. For example, a torus or inner tube has  $\beta_0 = 1$  because it is just one component,  $\beta_1 = 2$  because of the main hole through the middle and the hole formed when looking at a cross-section, and  $\beta_2 = 1$  because of the cavity of the inner tube. Betti numbers for simplicial complexes are defined in the same way. Formally,  $\beta_k(\mathcal{S})$  is the dimension of the  $k$ th homology group of the complex  $H_k(\mathcal{S})$  (Carlsson, 2009).

Various approaches exist for constructing simplicial complexes from  $\mathcal{Z}$ . All of these depend on a scale parameter  $\epsilon$  (also referred to as *time*) which specifies the extent of closeness of points. In the Čech complex  $\check{\text{Cech}}(\mathcal{Z}, \epsilon)$ , a simplex is created between a set of points  $\mathcal{G}$  if and only if there is a non-empty intersection of the closed Euclidean balls  $B(\mathbf{z}_i, \epsilon/2)$ ,  $\forall i \in \mathcal{G}$ . In the Vietoris-Rips (VR) complex,  $\text{VR}(\mathcal{Z}, \epsilon)$ , a simplex is created if and only if the Euclidean distance between every pair of points is less than  $\epsilon$ . Efficient construction of the VR complex can proceed by creating an  $\epsilon$ -neighborhood graph, also referred to as the *one-skeleton* of  $\mathcal{S}$ . Then inductively, triplets of edges that form a triangle are taken as two-dimensional simplices, sets of four two-dimensional simplices that form a tetrahedron are taken as three-dimensional simplices, and so on.

Homological inference depends on the scale parameter

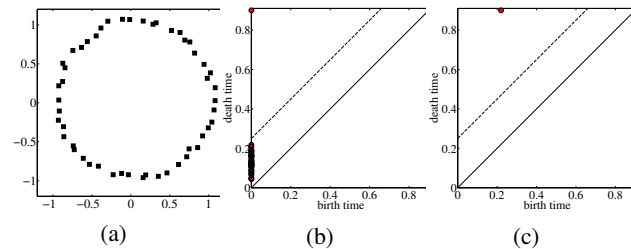


Figure 2. (a) Noisy data samples from an underlying circle in 2-dimensional space, (b) persistence diagram for  $H_0$ , (c) persistence diagram for  $H_1$ .

(time) at which the complexes are constructed. The topological features of the simplicial complex  $\mathcal{S}$  constructed from the data points  $\mathcal{Z}$  that are stable across scales, i.e. that are *persistent*, are the ones that provide information about the underlying shape. Topological features that do not persist are noise. *Persistence diagrams* are representations of the birth and death times of each homology cycle corresponding to each homology group  $H_k$ ,  $k = 0, 1, \dots$ , i.e. for increasing values of the scale parameter, the  $\epsilon$  value at which a topological feature begins to exist and ceases to exist.

As an example, let us consider the point cloud  $\mathcal{Z}$  shown in Figure 2(a), with noisy samples drawn from a circle, which has Betti numbers  $\beta_0 = 1$ ,  $\beta_1 = 1$ , and  $\beta_k = 0$  for  $k > 1$ . At the value  $\epsilon = 0$ , the simplicial complex that is formed from  $\mathcal{Z}$  is a collection of all the individual points not connected to any other point, resulting in the birth of  $n$  topological features in the  $H_0$  persistence diagram shown in Figure 2(b). As the scale increases, all of these little features die and only one persists until the largest scale under consideration; thus the persistent  $\beta_0 = 1$ . Looking at the  $H_1$  persistence diagram in Figure 2(c), we see that the only feature that is born persists until the largest scale and thus the persistent  $\beta_1 = 1$ . It is born at approximately a scale parameter of 0.2, which is when all of the points have been connected into a ring in the simplicial complex.

This section and Figure 2 (©2015 IEEE) are reprinted with permission, from Varshney & Ramamurthy (2015).

### 4. Decision Boundary Complexes for Demonstration in Section 3.3-M

We will display the decision boundary complexes corresponding to the demonstration in Section 3.3-M. The local scale multipliers for the LS-LVR filtration in Figure 3 are varied between 0.5 and 1.5 in 100 increments. The scale parameters for the P-LVR filtration in Figure 4 are varied between 0 and 10 in 100 increments. We only show 20 complexes for each filtration in evenly spaced increments.

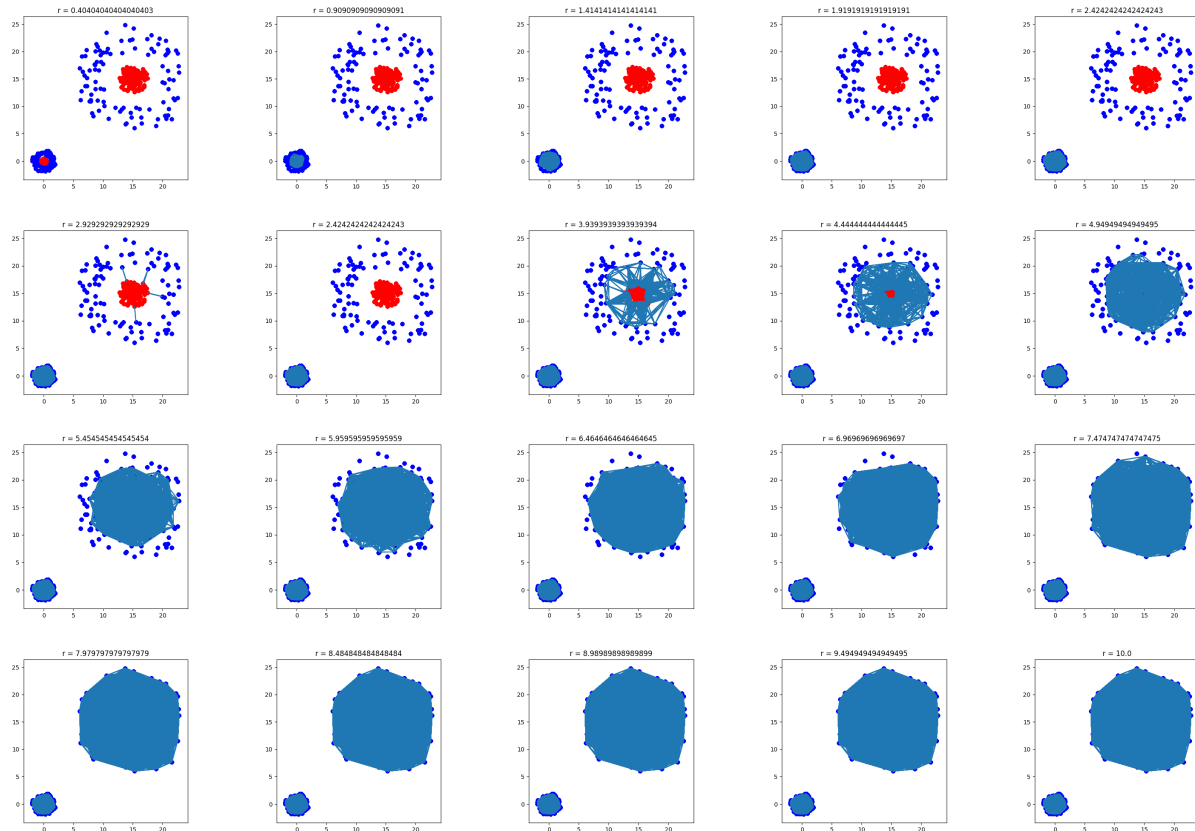


Figure 3. P-LVR complexes - the scales are given in the title of each image.

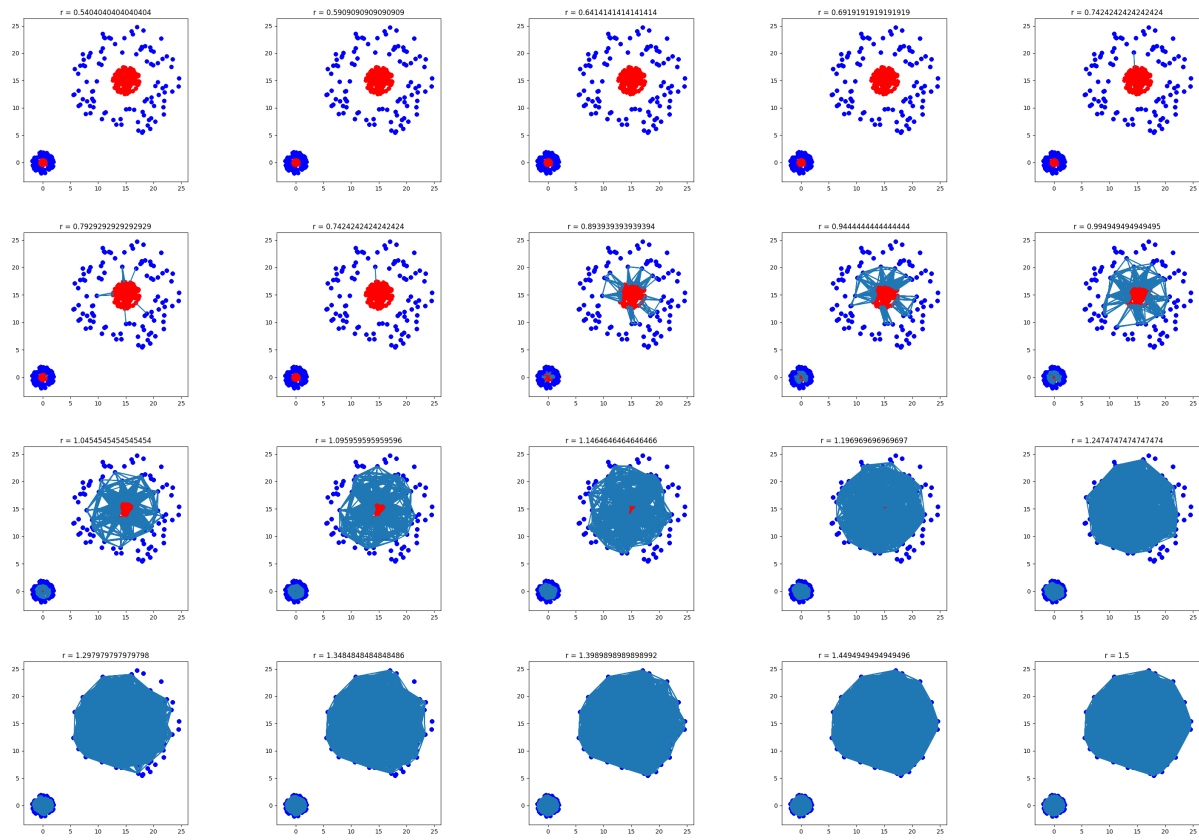


Figure 4. LS-LVR complexes - the local scale multipliers are given in the title of each image.

## **5. Data and Model Complexities - Total Lifetimes of Homology Groups**

We provide the data model complexities for all binary datasets and trained classifiers for the three applications in the following tables.









## **6. Performance of Task Classifier on Other Tasks**

We provide the classifier accuracies for the three applications in the following tables.















































































## References

- Bianchini, M. and Scarselli, F. On the complexity of neural network classifiers: A comparison between shallow and deep architectures. *IEEE Trans. Neural Netw. Learn. Syst.*, 25(8):1553–1565, August 2014.
- Carlsson, G. Topology and data. *Bull. Amer. Math. Soc.*, 46(2):255–308, April 2009.
- Guss, W. H. and Salakhutdinov, R. On characterizing the capacity of neural networks using algebraic topology. arXiv:1802.04443, 2018.
- Varshney, K. R. and Ramamurthy, K. N. Persistent topology of decision boundaries. In *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 3931–3935, Brisbane, Australia, April 2015.