



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Topological analysis of decision boundaries

Practical work

Maxim Mikhaylov

January 30, 2025

Supervisors: Prof. Dr. B. Gärtner, Dr. P. Schnider

Department of Computer Science, ETH Zürich

Abstract

Machine learning classifiers partition their input space into regions corresponding to different class labels. The boundaries between these regions, known as *decision boundaries*, play a crucial role in determining how a classifier generalizes to unseen data [9]. While traditional metrics like accuracy and loss provide aggregate measures of performance, they offer limited insight into how models make their decisions. Understanding the geometric and topological properties of these boundaries can provide valuable insights into classifier behavior, particularly regarding phenomena like overfitting and underfitting.

This project investigates the application of *topological data analysis* (TDA) to study decision boundaries in machine learning classifiers and how these boundaries evolve during training. We build upon the *labeled Vietoris-Rips* (LVR) complex framework, which was previously limited to binary classification, extending it to handle multiclass classification problems without the information loss inherent in decomposing multiclass problems into binary ones. Our multiclass extension of the LVR complex not only preserves more topological information but also proves more computationally efficient than binary decomposition, suggesting better scalability to complex datasets.

We analyze how topological features of decision boundaries change throughout the training process, establishing connections between topological metrics and model performance. Our results demonstrate that topological metrics strongly correlate with model accuracy across different architectures and datasets. We show that well-performing models develop similar topological structures in their decision boundaries, regardless of their architecture. Additionally, we find that topological metrics computed on training data correlate with test performance, suggesting potential applications in model evaluation without requiring a separate test set.

We propose and evaluate *circumcircle filtering*, a technique to improve the recovery of the LVR complex, and demonstrate why it becomes unnecessary in high-dimensional spaces of real-world datasets.

Contents

Contents	iii
1 Introduction	1
1.1 Related work	2
1.2 Overview	2
2 Theoretical background	3
2.1 Persistent homology	4
2.2 Simplicial complexes	4
3 Methodology	9
3.1 Datasets and models	9
3.2 Metrics and evaluation	11
3.3 Simplicial complex	11
4 Synthetic 2D data experiments	13
4.1 Initial results	13
4.2 Circumcircle filtering	14
4.3 Dowker complex	16
5 Results and discussion	19
5.1 Impact of circumcircle filtering	19
5.2 Results in the binary classification setting	21
5.3 Results on multiclass classification	24
5.4 Topology of the training decision boundary	27
5.5 Discussion and Future Work	29
A Visualization of LVR complex construction	31
B Visualization of Dowker complex construction	37

CONTENTS

C Binary classification topological metrics	41
D Multiclass classification topological metrics	45
Bibliography	49

Chapter 1

Introduction

Deep learning models have become increasingly complex, making it challenging to understand how they make their decisions. While metrics like accuracy and loss help evaluate model performance, they provide limited insight into the decision-making process.

One promising approach to understanding neural network behavior is to study the geometric and topological properties of their decision boundaries — the surfaces that separate different class regions in the input space. These boundaries encode fundamental information about how a model partitions the input space and how it will classify previously unseen data points. Research has demonstrated that geometric properties of decision boundaries, such as their margins, strongly influence a model’s ability to generalize [9].

Topological data analysis (TDA) offers a powerful framework for studying these boundaries. By capturing properties that are invariant under continuous deformation, topology can help us understand the essential structure of decision boundaries while ignoring irrelevant geometric details. Moreover, modern computational topology provides tools like persistent homology that can quantify topological features at different scales.

Our key contributions include:

1. We extend the Labeled Vietoris-Rips framework from binary to multi-class classification, avoiding the information loss inherent in decomposing multiclass problems into binary ones.
2. We propose and evaluate circumcircle filtering, a technique to improve topology recovery in low dimensions, and analyze why it becomes unnecessary in high-dimensional real-world datasets.
3. We analyze how topological features of decision boundaries change throughout the training process, establishing connections between topological metrics and model performance.

1. INTRODUCTION

4. We investigate whether topological metrics computed on training data can serve as indicators of model quality, potentially enabling detection of overfitting without requiring a separate validation set.

1.1 Related work

Topological data analysis has been applied to neural network research using a wide range of methods. Recent surveys have categorized these approaches according to which aspect of neural networks they analyze [1]:

1. *Structure of neural networks.* Some researchers have computed two special homology groups for graphs that represent feed-forward neural networks to analyze their structural properties.
2. *Decision regions and boundaries.* The labeled Čech and Vietoris-Rips complexes were developed to capture homology of neural network decision boundaries [19]. Later work enhanced this approach with active learning to improve sampling efficiency [13]. Several studies have also explored using the *graph-based topological data analysis* (GTDA) algorithm, which extends the Mapper algorithm to handle graph inputs and generate Reeb networks [15, 21].
3. *Activations and weights.* Methods in this category apply either the Mapper algorithm [4, 10, 16], or persistent homology [5, 11, 20] to neural network weights or activations.
4. *Training dynamics and loss functions.* Researchers have studied the number of connected components and local valleys of convex optimization targets in fully connected feed-forward networks [17]. Other work has investigated the fractal dimensions of weight trajectories during training [3, 25].

This research bridges the categories 2 and 4, presenting a novel investigation that combines topological analysis of decision boundaries with neural network training dynamics.

1.2 Overview

The remainder of this paper is organized as follows. Chapter 2 provides the necessary theoretical background on persistent homology and simplicial complexes. Chapter 3 describes our experimental setup, including datasets, models, and evaluation metrics. Chapter 4 presents our initial experiments on synthetic 2D data, which motivate and validate our methodological choices. Chapter 5 presents our main results on real-world datasets and discusses their implications. Finally, we conclude with a discussion of potential future research directions.

Chapter 2

Theoretical background

This chapter introduces the mathematical concepts and tools needed for topological analysis of decision boundaries. We begin with *persistent homology*, a fundamental tool for measuring topological features at different scales. We then present several types of simplicial complexes that can be used to capture the topology of decision boundaries, discussing their properties and relationships to each other. The ultimate goal is to compute persistent homology of decision boundaries using these complexes.

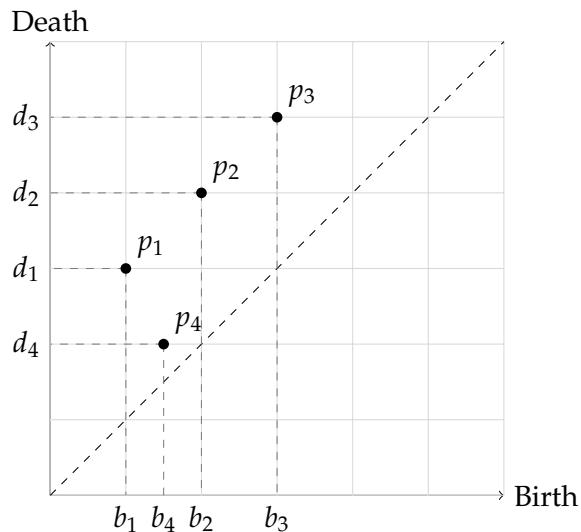


Figure 2.1: An example of a persistent diagram. A point p_i is born at b_i and dies at d_i .

2.1 Persistent homology

Persistent homology serves as a fundamental tool in TDA that tracks how topological features (connected components, holes) appear and disappear as a parameter is varied [8]. Given a filtered simplicial complex $(K_\varepsilon)_{\varepsilon \geq 0}$, persistent homology computes pairs (b_i, d_j) where a feature appears at $\varepsilon = b_i$ (birth) and disappears at $\varepsilon = d_i$ (death).

These birth–death pairs can be visualized in a *persistence diagram*, where each pair is shown as a point in the Euclidean plane. The distance of a point from the diagonal indicates the feature’s persistence, which is often interpreted as a measure of its significance. Figure 2.1 shows an example of a persistence diagram.

2.2 Simplicial complexes

To study the topology of decision boundaries, we need constructions that can capture their structure using only finite samples from the input space. Several types of simplicial complexes have been developed for this purpose, each with its own advantages and trade-offs.

One such complex is the *Dowker complex*, which has been used to study tumor microenvironments [22].

Definition 2.1 Let X, Y be sets. The Dowker complex $\mathcal{D}_{X,Y}^\varepsilon$ at parameter value ε is a simplicial complex with X as its vertex set, including the n -simplex (x_0, \dots, x_n) if and only if there exists $y \in Y$ such that $\|x_i - y\| \leq \varepsilon$ for all i .

A Dowker filtration can be obtained by varying ε .

Although the roles of X and Y may appear arbitrary, the following theorem shows that they can be interchanged without changing the homology groups, up to isomorphism:

Theorem 2.2 (Dowker, [7]) For any Dowker complex constructed from two sets X and Y , interchanging the roles of X and Y and keeping the same value of ε results in homology groups that isomorphic to the original ones:

$$H_k(\mathcal{D}_{X,Y}^\varepsilon) \cong H_k(\mathcal{D}_{Y,X}^\varepsilon). \quad (2.1)$$

The labeled Čech complex has been introduced as a theoretical tool to study decision boundaries of neural networks, offering probabilistic guarantees that it correctly captures the topology of these boundaries [19].

Definition 2.3 Given a set of points X , a reference set Y , and parameters ε and γ , the labeled Čech complex $L\check{C}_{X,Y}^{\varepsilon,\gamma}$ contains an n -simplex formed by points $x_0, \dots, x_n \in X$ if and only if:

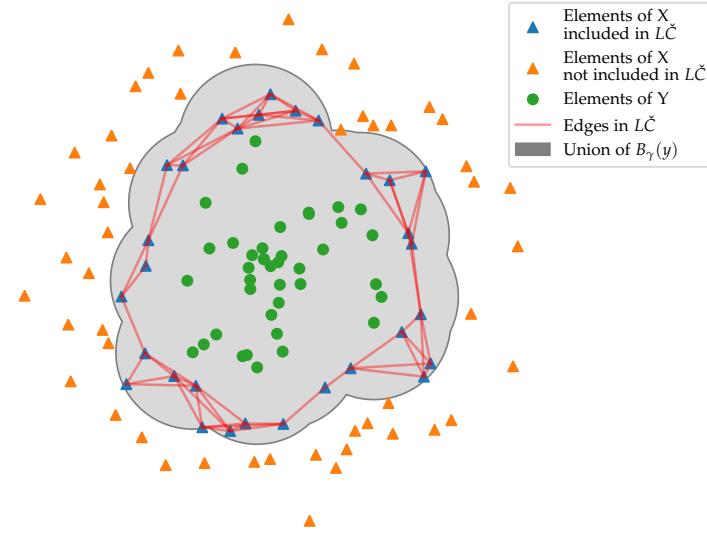


Figure 2.2: An example of the labeled Čech complex. Only 0- and 1-simplices are shown for clarity.

1. $\cap_{i=0}^n B_\varepsilon(x_i) \neq \emptyset$
2. For each $i \in (0, \dots, n)$, there exists $y \in Y$ such that $\|x_i - y\| \leq \gamma$.

The *labeled Čech filtration* is obtained by varying ε , while keeping γ fixed. Figure 2.2 shows an example of a labeled Čech complex constructed on points from two different classes.

The parameter ε encodes a distance threshold within X , similarly to the Čech complex, while the parameter γ encodes a distance threshold from points in X to Y , similarly to the Dowker complex. More concretely,

$$L\check{C}_{X,Y}^{\varepsilon,\gamma} = \check{C}_X^\varepsilon \cap D_{X,Y}^\gamma, \quad (2.2)$$

where \check{C}_X^ε is the Čech complex.

Both the labeled Čech complex and the Dowker complex only capture simplices with vertices in X , i.e. on one side of the decision boundary. This makes them not directly applicable to the multiclass setting, as they would only capture simplices on the boundary of one class, ignoring the decision boundary between the other classes. Moreover, both of these complexes are computationally expensive. An alternative approach is to use the *labeled Vietoris-Rips (LVR) complex* [19], which can be applied to multiple classes and is more tractable at the cost of not having known recovery guarantees.

Definition 2.4 Given a set of points X with associated labels $c : X \rightarrow \{1, \dots, k\}$ and a parameter $\varepsilon > 0$, the labeled Vietoris-Rips complex is constructed in three steps:

2. THEORETICAL BACKGROUND

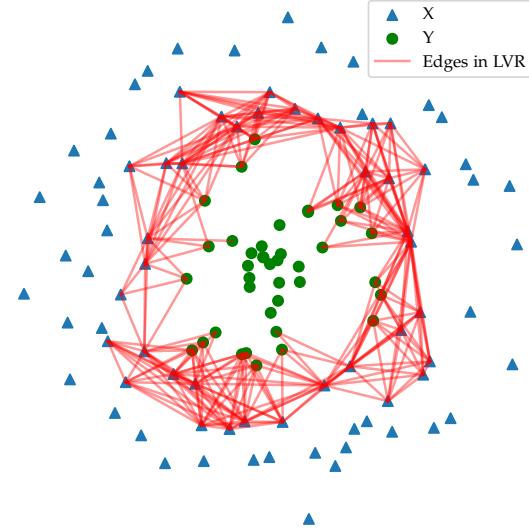


Figure 2.3: An example of the labeled Vietoris-Rips complex. Only 0- and 1-simplices are shown for clarity.

1. Create a graph G_ϵ with vertex set X by adding an edge between points $x_i, x_j \in X$ if and only if:
 - $\|x_i - x_j\| \leq \epsilon$ (points are close enough), and
 - $c(x_i) \neq c(x_j)$ (points belong to different classes)
2. Add edges between all 2-hop neighbors in G_ϵ , i.e. add an edge between x_i and x_k if there exists x_j such that both (x_i, x_j) and (x_j, x_k) are edges in G_ϵ .
3. Build higher-dimensional simplices by adding an n -simplex whenever all of its $(n-1)$ -dimensional faces are present in the complex, similarly to the standard Vietoris-Rips complex.

The *labeled Vietoris-Rips filtration* can be obtained by varying ϵ . An example of the LVR complex constructed on the same point set is shown in Figure 2.3. Comparing it to Figure 2.2, we can see that the LVR complex captures simplices that cross the boundary between the two classes, while the labeled Čech complex only captures simplices on one side of the boundary.

While it is known that the VR and Čech complexes are log-interleaved through inclusions [8]:

$$\check{C}_X^\epsilon \subseteq VR_X^\epsilon \subseteq \check{C}_X^{2\epsilon}, \quad (2.3)$$

it is not clear how the labeled versions compare.

The *locally scaled labeled Vietoris-Rips complex (LS-LVR complex)* aims to recover homology groups of datasets with non-uniform density more accurately compared to the LVR complex.

Definition 2.5 *The LS-LVR complex is constructed similarly to the LVR complex, but with the distance constraint for an edge (x_i, x_j) being given by $\|x_i - x_j\| \leq \epsilon\sqrt{\rho_i, \rho_j}$, where ρ_i is the distance from x_i to its k -th closest neighbor, with k being a fixed parameter.*

Chapter 3

Methodology

This chapter describes our pipeline for analyzing the topology of neural network decision boundaries. For a given neural network and dataset, our pipeline consists of the following steps:

1. Sample a subset of the dataset and obtain model predictions on this subset
2. Construct the LVR complex using the predicted labels
3. Construct the LVR complex using ground truth labels
4. Compare the resulting persistence diagrams using topological metrics

We first present the datasets used in our experiments, including both synthetic 2D data for interpretable validation and real-world image datasets. We then detail the neural network architectures used for classification. Finally, we describe our choices for the key components of the pipeline: the construction of the Labeled Vietoris-Rips complex for both binary and multiclass settings, and the metrics used for comparing persistence diagrams.

3.1 Datasets and models

To evaluate our approach we have used the MNIST [6], FashionMNIST [23] and CIFAR10 [12] datasets with the predefined training and testing splits. These datasets are commonly used as simple problems, providing a robust foundation for evaluating model performance. Additionally, we have also used synthetic 2D binary classification datasets for more easily interpretable experiments. In the 2D setting, we can visualize the simplicial complex, as well as know the ground truth decision boundary and its homology groups. For subsampling, we use 2000 points unless specified otherwise.

The synthetic datasets consist of points uniformly sampled from the $[-1, 1]^2$ square, with a total of 2000 points per dataset. Figure 3.1 shows the two

3. METHODOLOGY

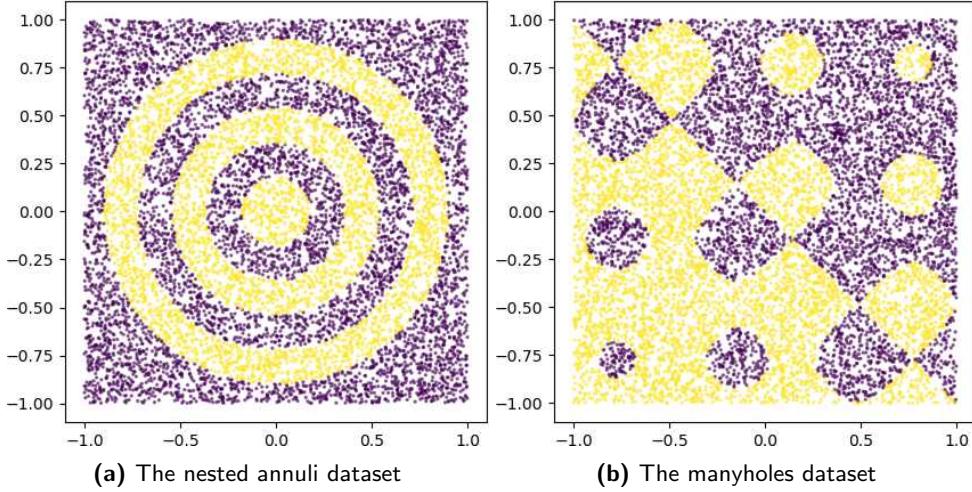


Figure 3.1: Synthetic 2D datasets

datasets used. The nested annuli dataset, also referred to as simply the nested dataset, has the decision boundary of 5 concentric circles, with the i -th circle having the radius of $0.18 \cdot (i + 1)$. The manyholes dataset assigns a point (x, y) class 1 if

$$\sin(10x) + \sin(10y) - x - y > 0, \quad (3.1)$$

and class 0 otherwise. This results in multiple “holes” in the decision boundary at various scales.

For the image datasets, we have used the following models:

1. **FC:** A fully connected neural network with layer sizes $w \cdot h, 512, 128, 16, 10$, where w and h are input image width and height. The network uses ReLU activations after each layer, and a softmax activation after the last layer.
2. **CNN:** A convolutional neural network with the architecture given in the PyTorch MNIST example¹, with the following parameters added:
 - a) **in_size** and **in_channels**: To account for CIFAR-10 having 3 channels and 32×32 images, as opposed to 1 channel and 28×28 images in MNIST and FashionMNIST, **in_channels** sets the number of input channels, and **in_size** adjusts the number of neurons in the first fully connected layer.
 - b) **scale**: A multiplicative factor for the number of convolutional filters and number of neurons in the fully connected layers.

¹Available at <https://github.com/pytorch/examples/tree/main/mnist>.

- c) `extra_cnn`: Adds convolutional layers with $64 \cdot \text{scale}$ channels, a 3×3 kernel size, 1-pixel padding and stride, as well as a ReLU activation.
- d) `extra_linear`: Adds fully connected layers with $128 \cdot \text{scale}$ neurons, as well as a ReLU activation.

3.2 Metrics and evaluation

For estimating similarity between two persistence diagrams, we use the following metrics:

Definition 3.1 For $q \geq 1$, the Wasserstein distance between two persistence diagrams $\text{Dgm}_p(\mathcal{F})$ and $\text{Dgm}_p(\mathcal{G})$ is defined as

$$d_{W,q}(\text{Dgm}_p(\mathcal{F}), \text{Dgm}_p(\mathcal{G})) = \left[\inf_{\pi \in \Pi} \left(\sum_{x \in \text{Dgm}_p(\mathcal{F})} \|x - \pi(x)\|_\infty^q \right) \right]^{1/q} \quad (3.2)$$

Definition 3.2 The total bar length (TBL) of a persistence diagram $\text{Dgm}_p(\mathcal{F})$ is defined as

$$\sum_{(b,d) \in \text{Dgm}_p(\mathcal{F})} d - b \quad (3.3)$$

The Wasserstein distance and TBL offer different approaches to comparing persistence diagrams. The Wasserstein distance directly measures the similarity between two persistence diagrams by comparing them to each other. In contrast, total bar length is a single-diagram measure that quantifies the complexity of an individual persistence diagram. To compare two diagrams using TBL, we calculate the TBL for each diagram separately and then take the absolute difference between these two values.

3.3 Simplicial complex

We use the labeled Vietoris-Rips complex, as presented in previous research [19]. To improve computational performance, we used the `ripser++` [24] and the `giotto-ph` [18] packages for computing Vietoris-Rips persistence barcodes. Both of these methods outperform the original `ripser` [2] package in terms of runtime. Due to GPU memory limitations, we use `giotto-ph` in all experiments.

In previous research [14, 19], classification datasets with $n > 2$ classes were split into $\binom{n}{2}$ binary classification problems. However, this process loses topological information, as shown in an example dataset in Figure 3.2. The decision boundary for all three classes (3.2a) has a non-trivial first homology

3. METHODOLOGY

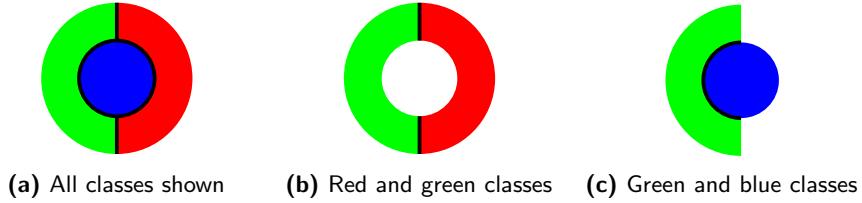


Figure 3.2: A 2D example that demonstrates the loss of topological information when binarizing a multiclass dataset. Red, green and blue shapes denote different classes, the black line is the decision boundary.

group H_1 , but this is not the case for any binary classification problem for this dataset (3.2b, 3.2c). Moreover, the decision boundary is disconnected for the red and green subdataset, but connected for the full dataset.

To avoid this information loss, we propose applying the Labeled Vietoris-Rips complex to multiclass classification problems. The construction of the complex is exactly the same as in the binary setting, with neighbors of different classes being connected and 2-hop neighbors being connected afterwards. Notably, the labeled Čech complex cannot be applied to the multiclass setting without modifications, as it uses elements of only one class as its vertices.

Consistent with previous research, we only consider 0 and 1-dimensional homology groups, and thus only compute simplices up to dimension 3. Unless specified otherwise, all homology groups and persistence diagrams discussed are in dimension 1.

Chapter 4

Synthetic 2D data experiments

This chapter evaluates our topological analysis methods on synthetic 2D datasets where we can visually verify our results and understand the challenges that arise. We first demonstrate that the LVR complex can produce spurious topological features by allowing simplices to cross decision boundary multiple times. We then propose and evaluate two potential solutions to this limitation: circumcircle filtering and the Dowker complex.

4.1 Initial results

To assess how well the LVR complex captures the topology of the decision boundary, we have created two synthetic datasets. Because the points in

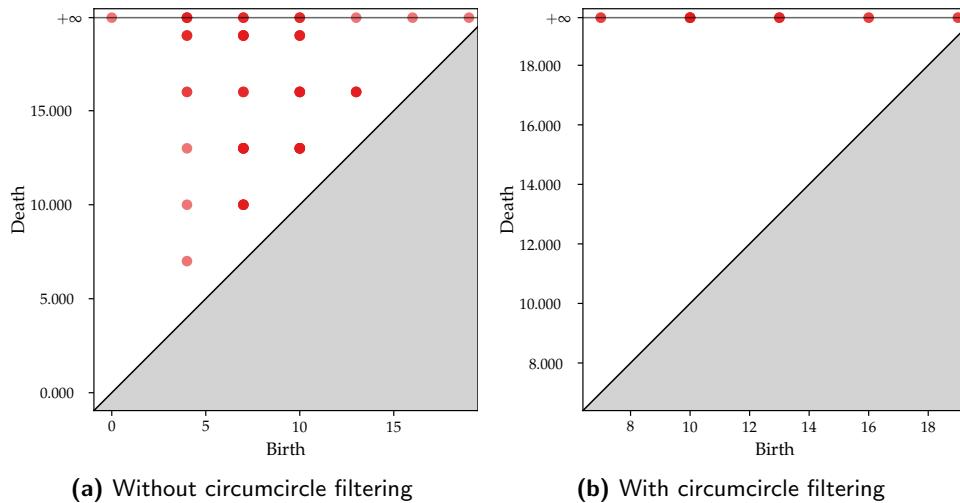


Figure 4.1: Persistence diagrams of the LVR complex on the nested dataset with and without circumcircle filtering.

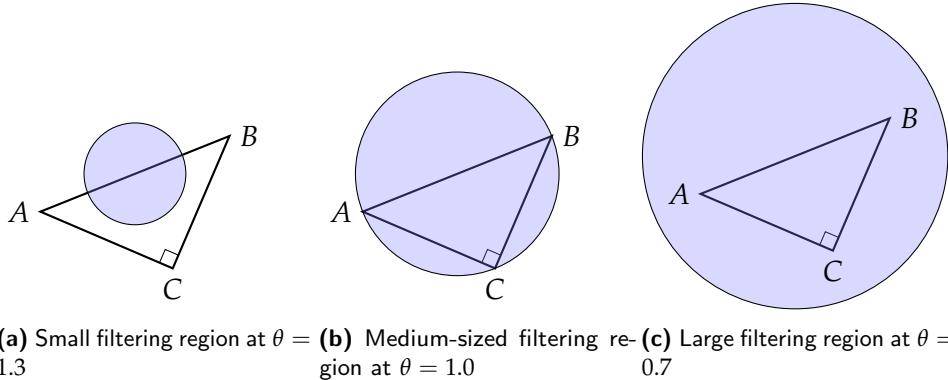


Figure 4.2: Circumcircle filtering with different values of θ . The light blue circle shows the region where any point's presence would cause edge AB to be removed. A larger value of θ leads to a smaller filtering region, while a smaller θ increases the region and thus leads to more aggressive filtering.

these datasets are distributed uniformly, we used the LVR complex rather than the LS-LVR complex.

On the nested dataset, we would expect to see five 1-dimensional homology classes that persist indefinitely. However, as shown in Figure 4.1a, the persistence diagram reveals more such classes than expected, and some of them do not persist until infinity. By examining the 2-skeleton of the complex in Figure 4.3a, we see that the LVR complex includes edges and triangles that cross the decision boundary multiple times. This is a problem, because this splits a true homology class into several classes, some of which may not persist until infinity.

4.2 Circumcircle filtering

To address this drawback, we propose *circumcircle filtering* (CC). This method removes an edge between vertices A and B (and all higher simplices containing this edge) from the simplicial complex if there exists a vertex C such that $\|A - B\|^2 > (\|A - C\|^2 + \|B - C\|^2)\theta$, where $\theta \in [0, 2]$ is a parameter. Intuitively, this approach aims to make a simplicial complex more similar to the Alpha complex [8] by removing some (but not all) non-Delaunay edges. The set of edges that should be removed with CC can be computed in $\mathcal{O}(n^2 \cdot d)$ time and $\mathcal{O}(n^2)$ memory by computing all pairwise distances of points. These computational costs are negligible compared to building the LVR complex and computing its persistent homology.

The CC method can be visualized by considering the plane with the points A, B and C in Figure 4.2. At $\theta = 1$, an edge AB is removed if $\|A - B\|^2 > \|A - C\|^2 + \|B - C\|^2$, i.e. if C lies inside a circle which has AB as its diameter, hence the name of *circumcircle filtering*. Increasing or decreasing the value of

4.2. Circumcircle filtering

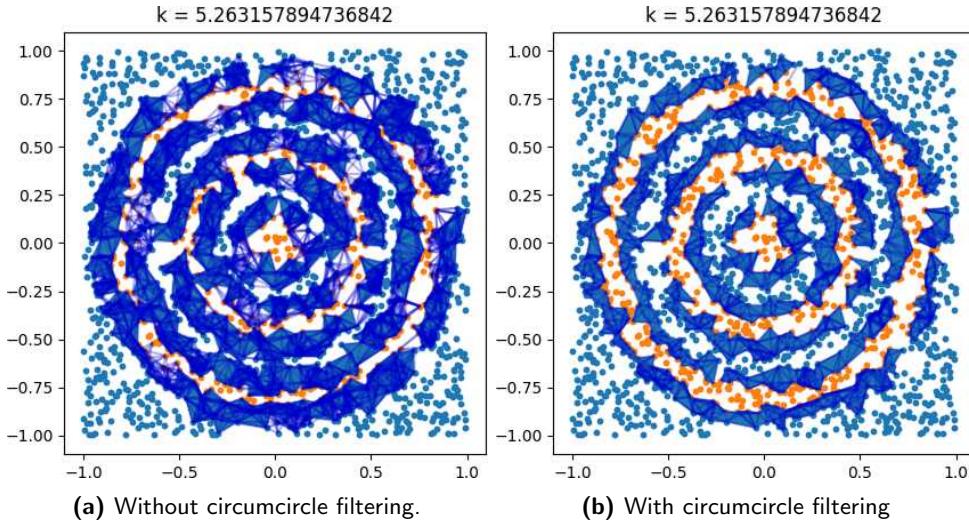


Figure 4.3: A visualisation of the 2-skeleton of the LVR complex on the nested dataset with and without circumcircle filtering. Not all 2-simplices are shown for clarity.

θ decreases or respectively increases the size of the circle, keeping the center the same. At $\theta = 0$, all edges are removed, because $\|A - B\|^2 > 0$ always holds, and at $\theta = 2$, no edges are removed, because

$$2 \cdot (\|A - C\|^2 + \|B - C\|^2) \geq \|A - C\|^2 + \|B - C\|^2 + 2\|A - B\|\|B - C\| \quad (4.1)$$

$$= (\|A - C\| + \|B - C\|)^2, \quad (4.2)$$

and

$$\|A - B\|^2 \leq (\|A - C\| + \|B - C\|)^2 \quad (4.3)$$

by the triangle inequality.

For the 2D datasets, we set $\theta = 1.4$, as with lower values almost no simplices were being filtered, and higher values filtered out too many simplices and the complex became very disconnected.

In Figure 4.1b, we see that applying CC lowers the number of homology classes and removes all classes that do not persist until infinity. Examining the 2-skeleton of the complex in Figure 4.3b, we see that almost all simplices that cross the decision boundary multiple times are removed. We can also see that CC has an added benefit of reducing the number of simplices in the complex, which can help reduce the computational cost of constructing the complex and computing its persistence diagram.

In Figure 4.4 and Figure 4.5, we observe a similar effect on the manyholes dataset, where without CC the complex contains spurious holes, such as the one around the point $(0.5, -0.1)$. Applying CC removes these spurious holes.

4. SYNTHETIC 2D DATA EXPERIMENTS

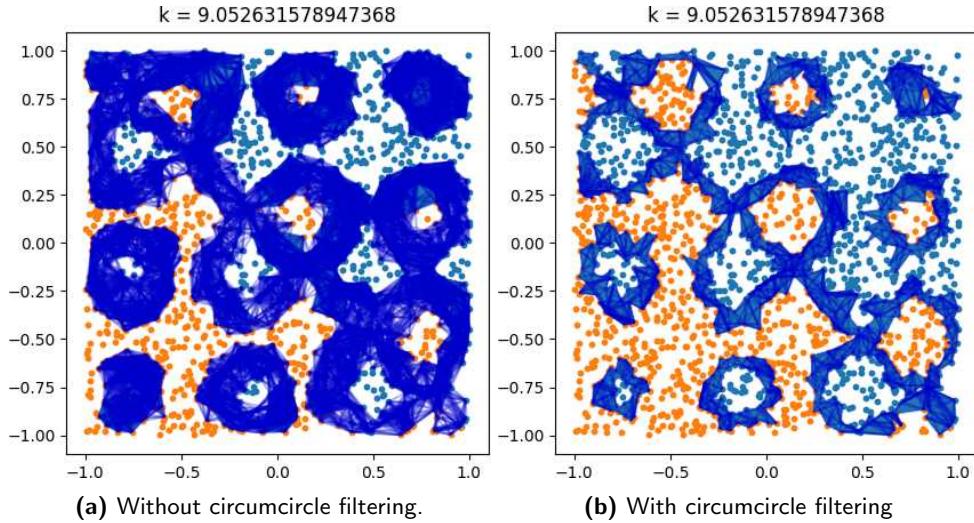


Figure 4.4: A visualisation of the 2-skeleton of the LVR complex on the manyholes dataset with and without circumcircle filtering. Not all 2-simplices are shown for clarity.

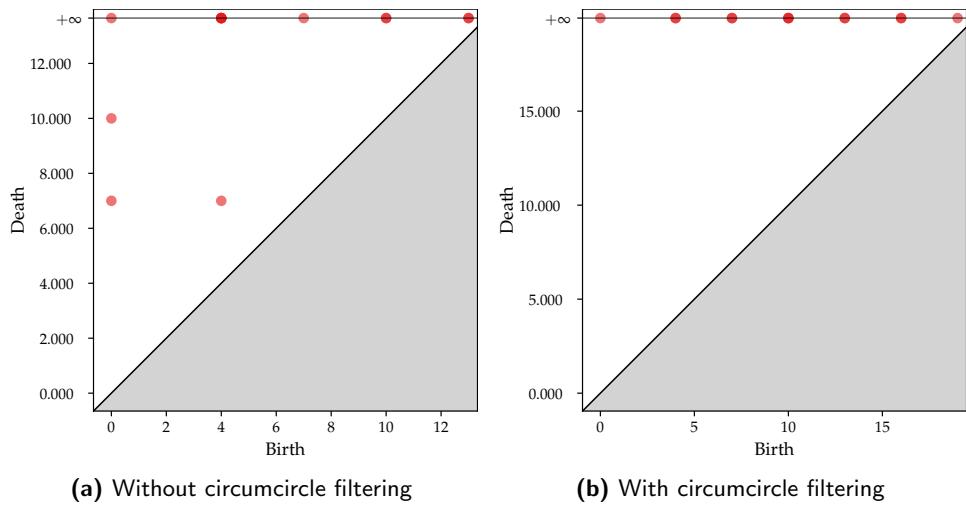


Figure 4.5: Persistence diagrams of the LVR complex on the manyholes dataset with and without circumcircle filtering.

A visualisation of how LVR changes as the filtration parameter is varied with and without CC can be found in Appendix A.

4.3 Dowker complex

Alternatively, the *Dowker complex* can be used to combat the same drawback. However, computing it naïvely up to 3-simplices requires $\mathcal{O}(n^3 \cdot d)$ runtime and $\mathcal{O}(n^3)$ memory for n points in \mathbb{R}^d , which is prohibitively expensive

4.3. Dowker complex

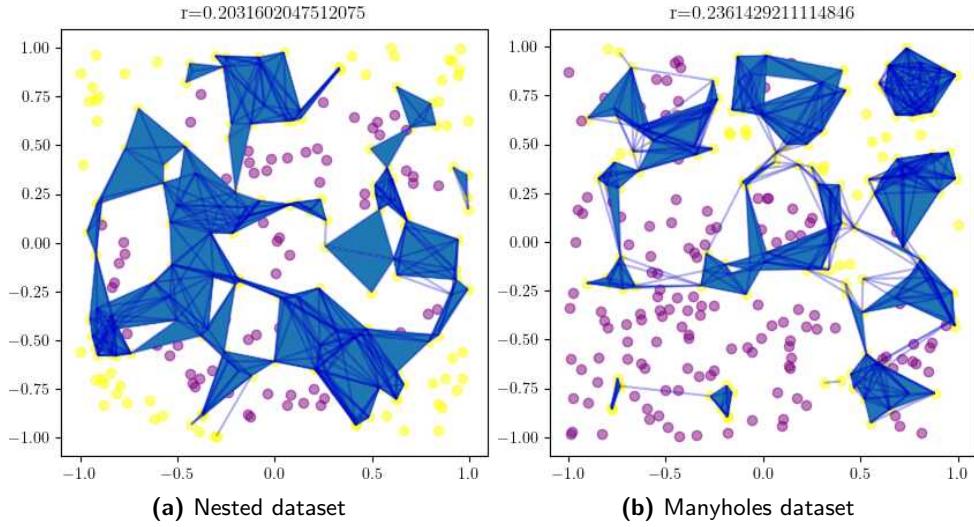


Figure 4.6: A visualisation of the 2-skeleton of the Dowker complex on the nested and manyholes datasets.

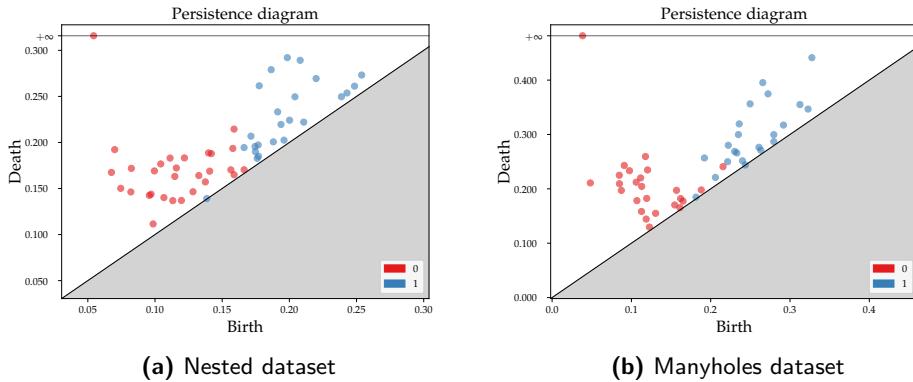


Figure 4.7: Persistence diagrams of the Dowker complex on the nested and manyholes datasets.

for the image datasets we consider. This approach may be more tractable within a more sample-efficient framework, such as one proposed in [14]. Additionally, the Dowker complex is not applicable to the multiclass setting.

In Figure 4.6, we show the Dowker complex on the nested and manyholes datasets. Due to the computational cost of constructing the Dowker complex, we only use 300 points from each dataset. We can see that the problem with simplices crossing the decision boundary multiple times is present in the Dowker complex as well. A visualisation of how the Dowker complex changes as the filtration parameter is varied can be found in Appendix B. The persistence diagrams shown in Figure 4.7 similarly demonstrate that the Dowker complex does not capture the topology of the decision boundary well.

Chapter 5

Results and discussion

This chapter presents our experimental results on real-world datasets. We begin by examining the impact of circumcircle filtering in high dimensions, showing why this technique, while useful in 2D, becomes unnecessary for real-world data. We then analyze how topological features of decision boundaries evolve during training in both binary and multiclass settings. Finally, we investigate whether topological metrics computed on training data can serve as indicators of model quality, potentially enabling detection of overfitting without requiring a separate validation set.

5.1 Impact of circumcircle filtering

Our previous two-dimensional experiments established an appropriate value of the circumcircle filtering parameter θ to be 1.4. However, applying this value to real-world datasets presents two challenges. First, these datasets exist in much higher dimensions: MNIST and FashionMNIST in 784 dimensions, and CIFAR-10 in 3072 dimensions. Second, while the two-dimensional datasets are relatively dense with 2000 points distributed over a 2D plane, the same number of points (used due to computational constraints) become sparse when distributed over hundreds or thousands of dimensions. These differences require us to determine the appropriate value of θ again.

To determine the appropriate value of θ in high dimensions, we analyzed persistence diagrams of the LS-LVR complex constructed using ground truth labels, comparing versions with and without CC filtering. The Wasserstein distance between these diagrams and the absolute distance between their total bar lengths provide a measure of CC's impact.

First, we examine the binary classification setting in the MNIST dataset with $\theta = 1$. Figure 5.1 illustrates that the Wasserstein distance remains largely unchanged when applying CC, despite CC removing 20–50% of

5. RESULTS AND DISCUSSION

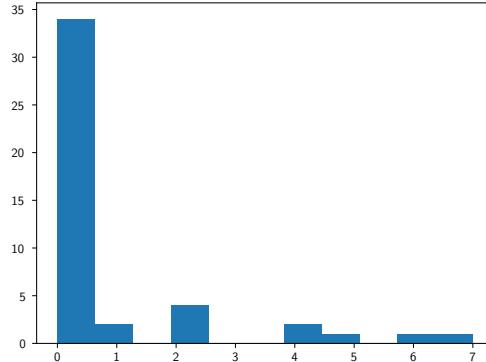


Figure 5.1: Distribution of changes in Wasserstein distance with and without CC filtering ($\theta = 1$) for binary classification on the MNIST dataset.

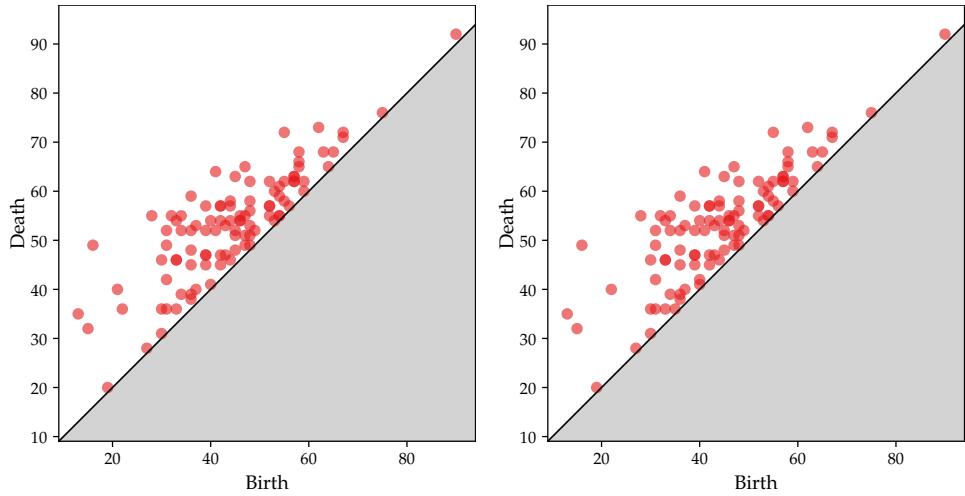


Figure 5.2: Comparison between persistence diagrams computed without CC filtering (left) and with CC filtering (right) for the binary classification pair that showed the largest difference in Wasserstein distance in MNIST.

edges. When changes do occur, they are minor — the maximum difference in distances is 7. This is generally a small difference, given that we consider 100 filtration values, and there are hundreds of homology classes in the persistence diagrams. Figure 5.2 shows the two most different persistence diagrams, which are visually very similar.

This lack of significant impact from CC is consistent across datasets and classification settings, as summarized in Table 5.1. Note that the values for the multiclass setting are lower because it involves only one experiment per dataset, as opposed to 45 experiments for the binary classification setting.

To increase the effect of CC, the value of θ would need to be reduced. However, values lower than 1 would remove an edge even when there are no points between its vertices. Figure 5.3 illustrates this: with $\theta < 1$, the edge

5.2. Results in the binary classification setting

Dataset	Binary	Multiclass
MNIST	7	4
FashionMNIST	14	1
CIFAR10	7	0

Table 5.1: Maximum differences in Wasserstein distance between persistence diagrams with and without CC for different datasets and classification settings.

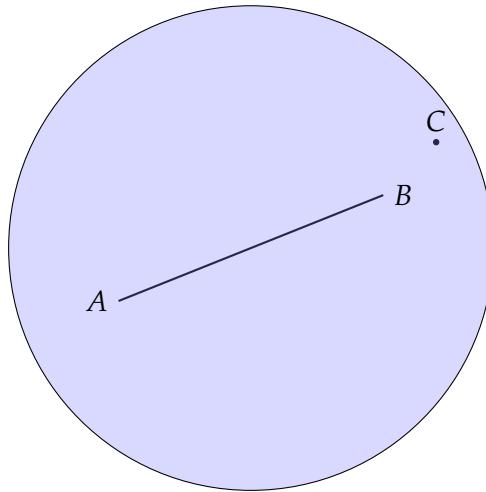


Figure 5.3: Illustration of circumcircle filtering with $\theta = 0.5$. The blue circle indicates the area where any point's presence would cause edge AB to be removed.

AB would be removed due to presence of point C in the circle, even though C is not between A and B . This goes against the intended purpose of CC.

Based on these findings, we conclude that circumcircle filtering is unnecessary for our experiments on real-world datasets. The most restrictive feasible value of $\theta = 1$ does not result in a significant change in the persistence diagrams, making CC unnecessary for our experiments.

5.2 Results in the binary classification setting

First, we analyze decision boundaries on the test subset of the MNIST dataset. We use the CNN model with default hyperparameters trained on the problem of distinguishing between the classes 0 and 1.

In Figure 5.5 we observe that the persistence diagram of the LS-LVR complex computed using predicted labels evolves during training to match the PD computed using ground truth labels. This is supported quantitatively by both the Wasserstein distance between the two diagrams and their total bar length, which are shown in Figure 5.4. The same trend can be observed for the other pairs of classes, shown in Figure C.2 and Figure C.3.

5. RESULTS AND DISCUSSION

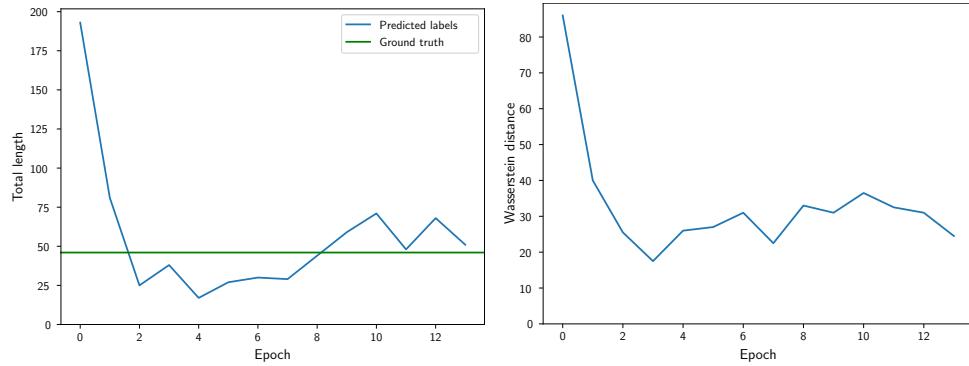


Figure 5.4: Evolution of topological metrics during training for the classes 0 and 1 in the MNIST dataset.

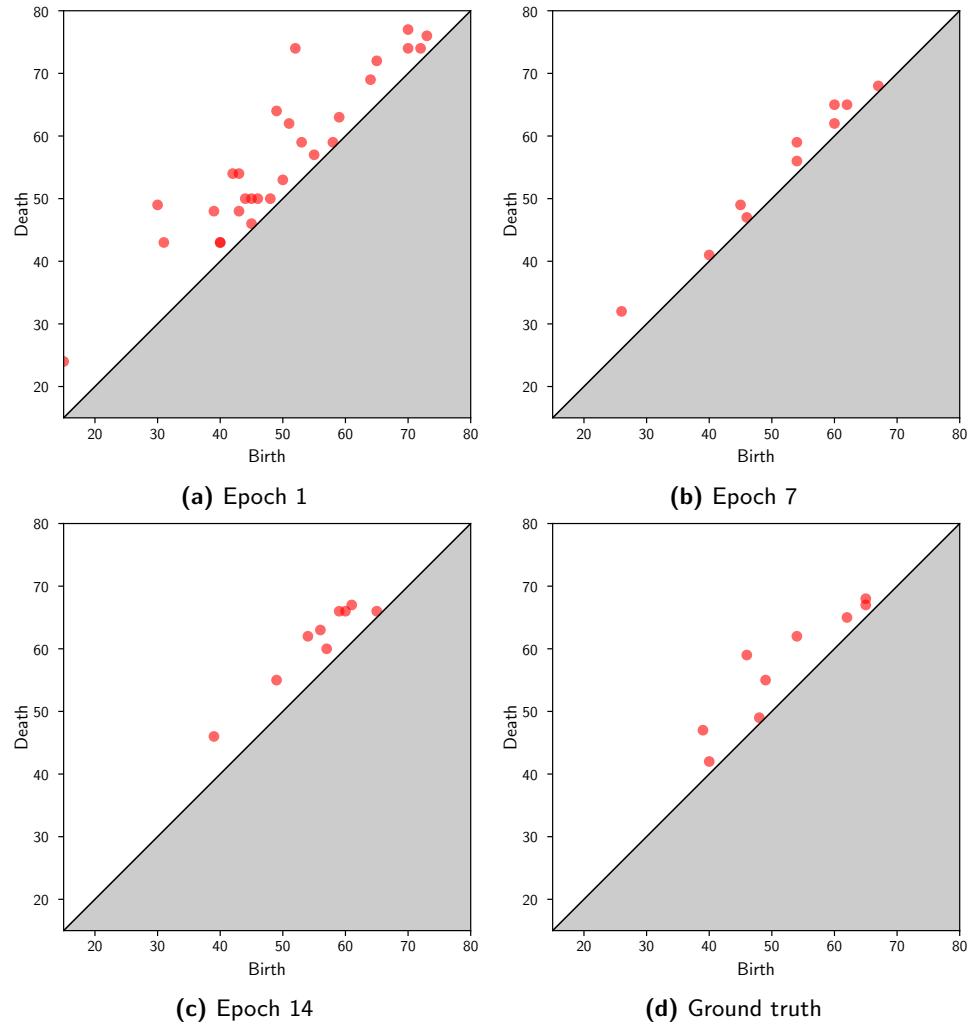


Figure 5.5: Evolution of persistence diagrams during training for the classes 0 and 1 in the MNIST dataset.

5.2. Results in the binary classification setting

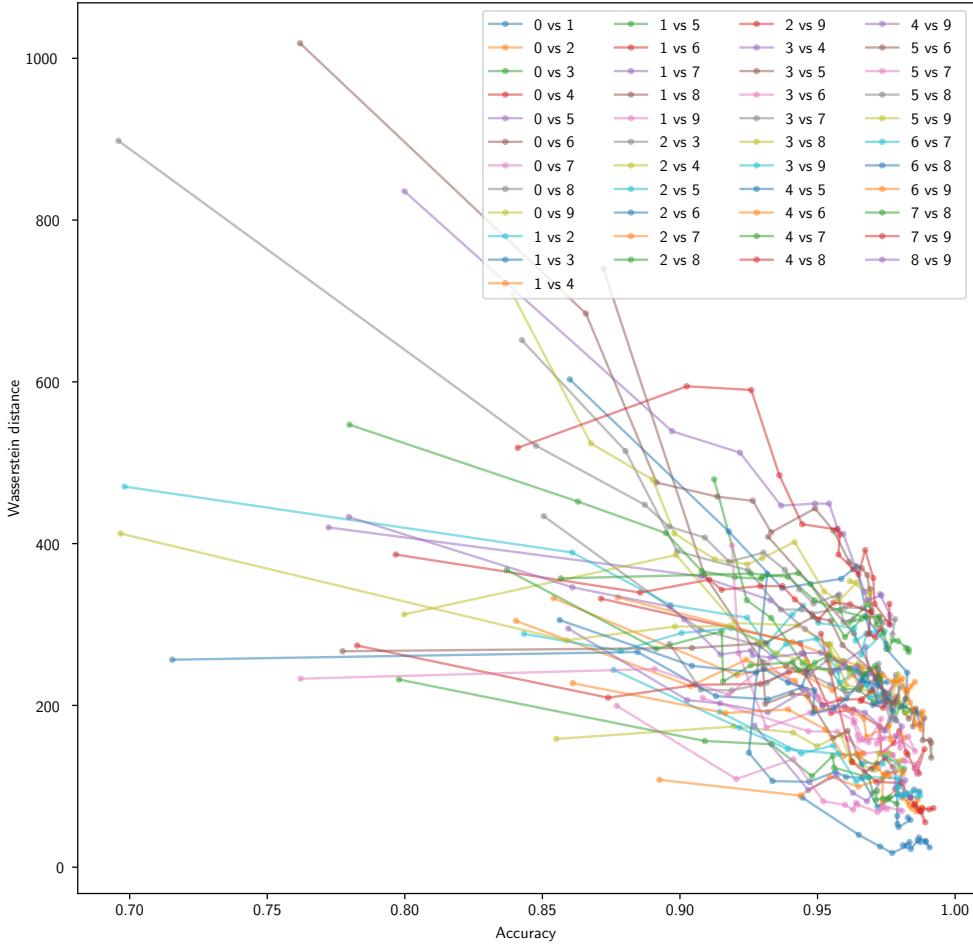


Figure 5.6: Relationship between Wasserstein distance and model accuracy across all binary classification pairs. Connected points represent consecutive epochs for the same class pair.

A notable “elbow” pattern emerges in these metrics: the initial epochs show significant changes, followed by slower, more stable changes. This pattern can be explained by the fact that as the model is trained, its loss decreases, and so does the magnitude of the changes in its weights. This, in turn, leads to less pronounced changes in the topology of the decision boundary.

The topological metrics also correlate with model accuracy. To illustrate this, we plot the Wasserstein distance as a function of accuracy across all class pairs in Figure 5.6, and do the same for the Wasserstein distance in Figure C.1. The Wasserstein distance shows stronger correlation with an average R^2 value of 0.72 compared to 0.49 for total bar length.

5. RESULTS AND DISCUSSION

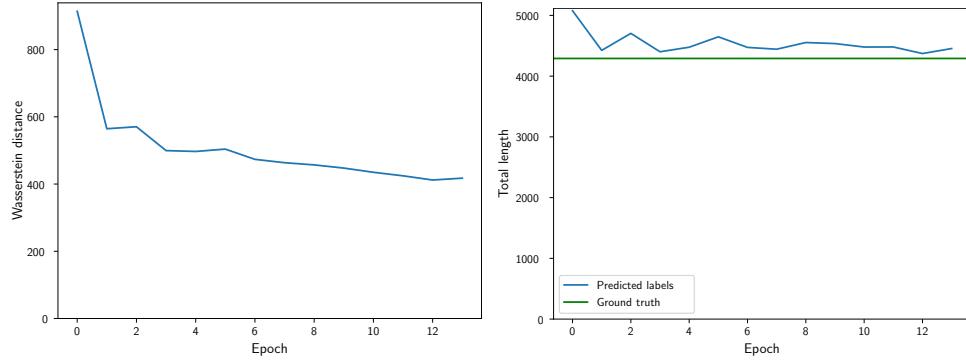


Figure 5.7: Evolution of topological metrics in multiclass training.

5.3 Results on multiclass classification

The results from binary classification extend naturally to the multiclass setting, but with additional topological information. As shown theoretically in Section 3.3, binarization loses topological information. This is evident in Figure 5.8, where the persistence diagrams contain significantly more points than their binary counterparts, reflecting more homology classes.

While the increased complexity makes qualitative analysis of the PDs more challenging, both topological metrics behave similarly to the binary case, as shown in Figure 5.7. Notably, the Wasserstein distance decreases monotonically, showing more stable behavior than in the binary setting.

Because we only have one measurement per epoch of a given model, to examine correlation of topological metrics with accuracy, we need multiple models. To this end, we train the following model variants:

1. *Baseline*: CNN with default hyperparameters
2. *Small*: CNN with reduced layer sizes (`scale = 0.25`)
3. *Wide*: CNN with increased layer sizes (`scale = 4`)
4. *Deep*: CNN with additional layers (4 extra CNN and 4 extra linear layers)
5. *Overfit*: CNN with aggressive learning rate growth ($\gamma = 1.9$) to simulate overfitting
6. *Underfit*: CNN with small learning rate (10^{-6}) to achieve underfitting
7. FC with default hyperparameters

Figure 5.9 reveals a strong correlation between Wasserstein distance and model accuracy across all model variants. A similar correlation exists for total bar length (Figure D.1), though with more noise in the measurements. The relationship appears even stronger than in binary classification, with an

5.3. Results on multiclass classification

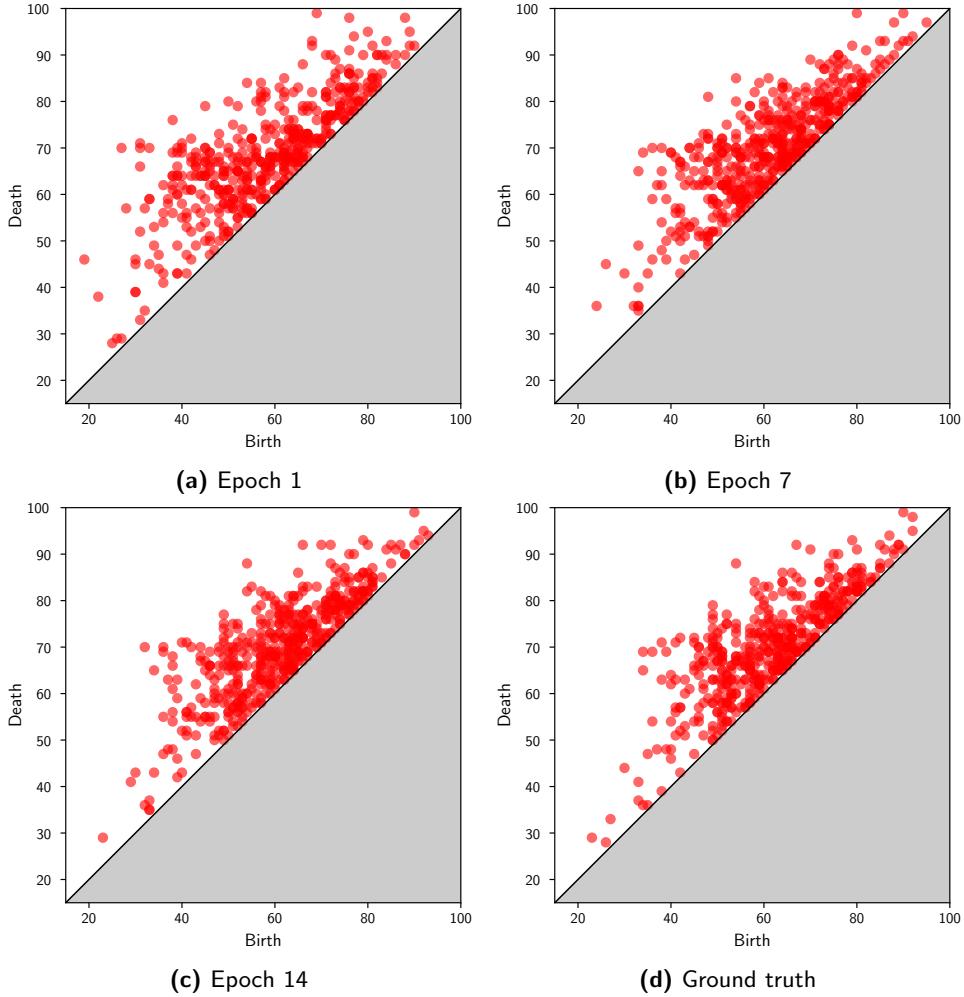


Figure 5.8: Evolution of persistence diagrams in the multiclass setting.

average R^2 value of 0.81 for Wasserstein distance across all model variants. This stronger correlation likely results from the richer topological information captured in the multiclass setting. Notably, the Wasserstein distance for the overfit model increases when the accuracy of the model decreases. This suggests that overfitting results in topological changes in the decision boundary, which would imply that overfitting can be detected by looking for increases in the Wasserstein distance.

The multiclass approach offers several advantages over binary decomposition. Beyond preserving more topological information, it provides more efficient evaluation, requiring only one analysis per epoch instead of $\binom{N}{2}$ binary comparisons. The higher point density in persistence diagrams also enables more robust statistical analysis.

5. RESULTS AND DISCUSSION

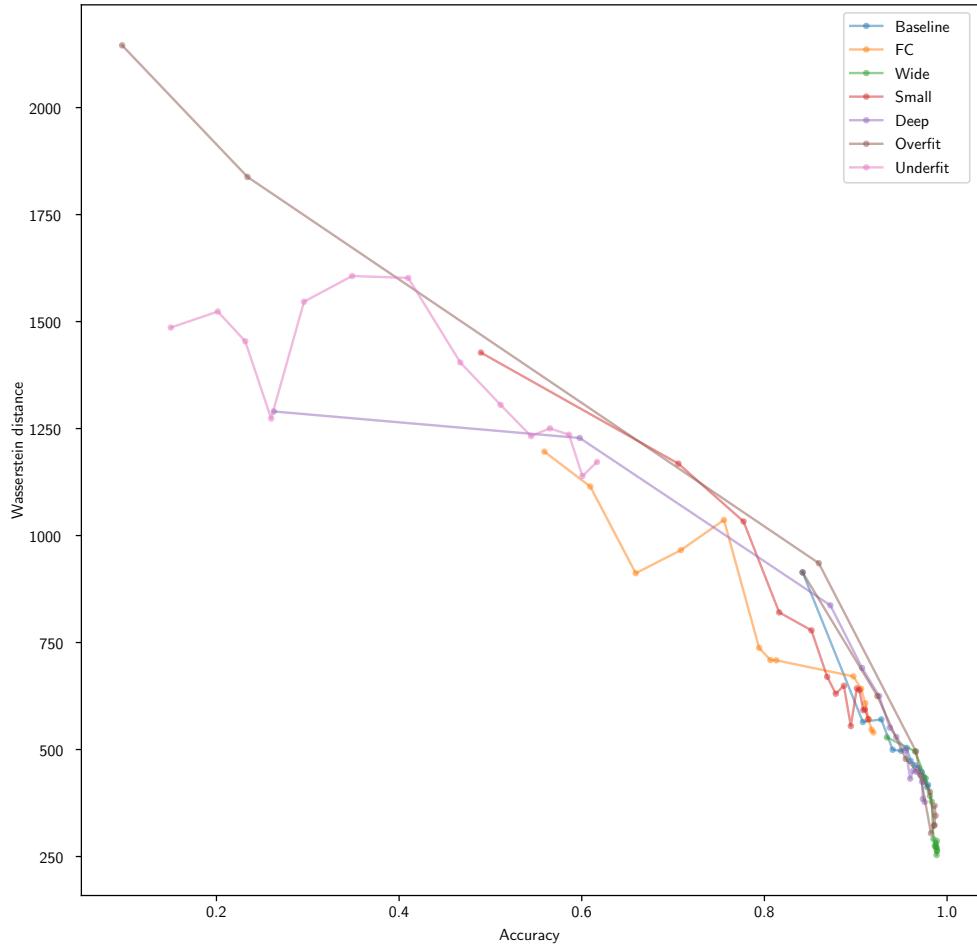


Figure 5.9: Relationship between Wasserstein distance and accuracy across different model architectures in multiclass classification. Connected points show consecutive epochs for each model. Note the particularly clear indication of overfitting in the overfit model’s trajectory.

An important consideration for our approach is how well it scales with dataset dimensionality. While MNIST and FashionMNIST images have 784 dimensions, CIFAR-10 presents a more challenging scenario with 3072 dimensions due to its larger resolution and three color channels. Initial experiments on CIFAR-10 using 2000 points (the same sample size used for MNIST) showed significantly weaker correlations between topological metrics and model accuracy, as shown in Figure 5.10 and Figure D.3. This degradation can be attributed to the increased sparsity of points in the higher-dimensional space — the same number of points must cover nearly four times as many dimensions. However, when we increased the sample size to 8000 points, the correlation became much stronger ($R^2 = 0.69$), as shown in Figure 5.11, approaching the levels observed for MNIST. This suggests that our method remains effective in higher dimensions, but requires more

5.4. Topology of the training decision boundary

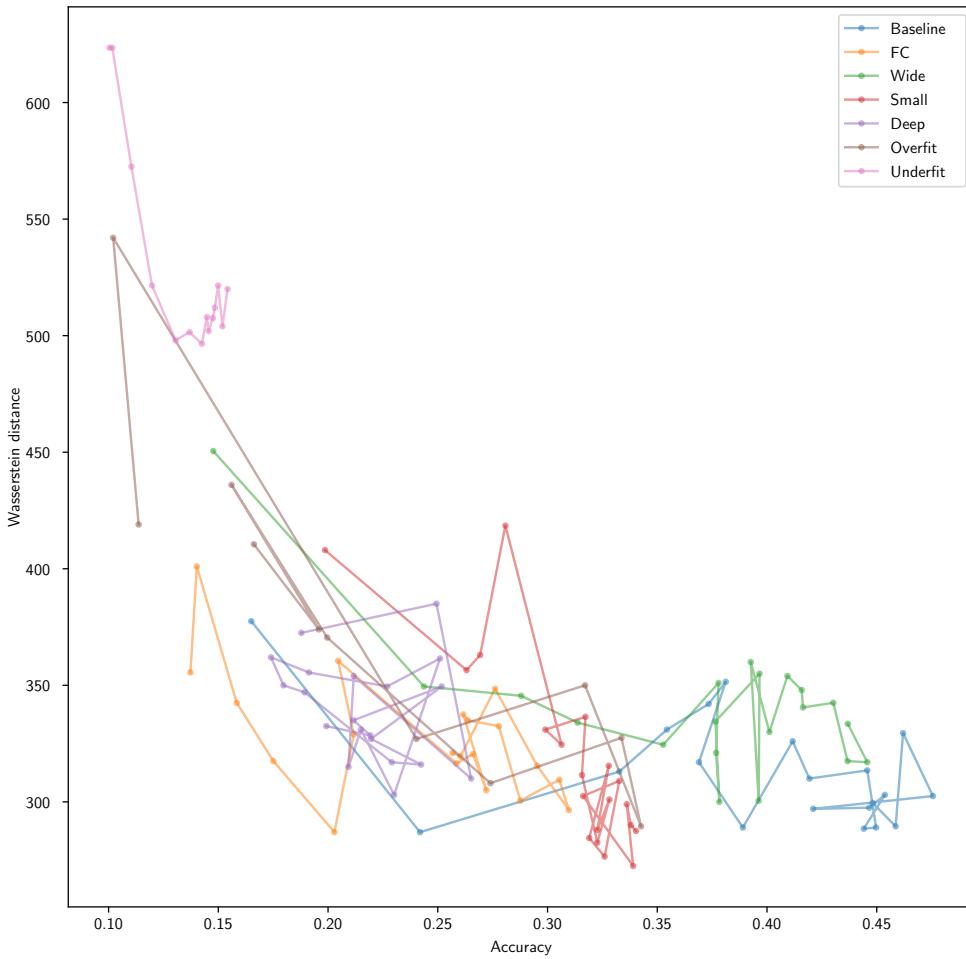


Figure 5.10: Relationship between Wasserstein distance and accuracy across different model architectures in multiclass classification on the CIFAR-10 dataset with 2000 points. Connected points show consecutive epochs for each model.

points to build a representative simplicial complex. This relationship between dimensionality and required sample size aligns with known challenges in high-dimensional data analysis, often referred to as the *curse of dimensionality*.

5.4 Topology of the training decision boundary

While our previous analysis in sections 5.2 and 5.3 focused on test data, examining the topology of decision boundaries on training data could provide valuable insights. Given the strong correlation between topological metrics and test accuracy, we investigate whether these metrics computed on training data could help assess model quality without requiring a separate test set, potentially enabling detection of overfitting and underfitting.

5. RESULTS AND DISCUSSION

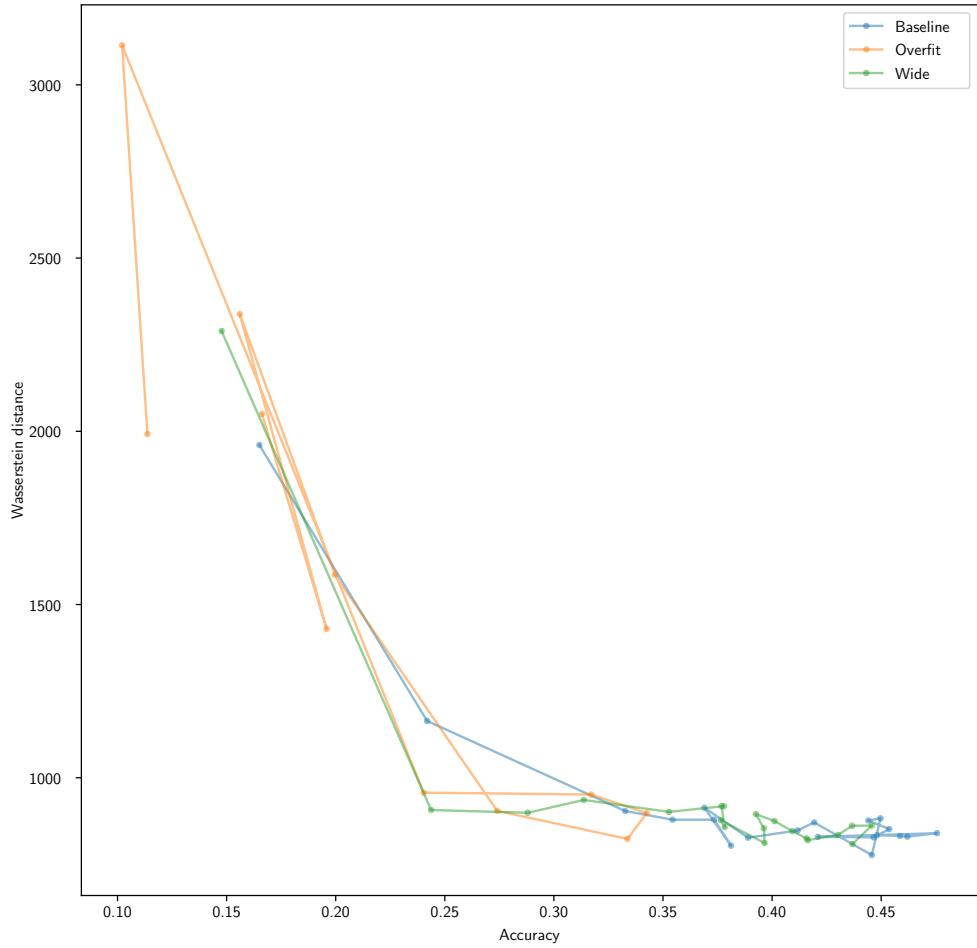


Figure 5.11: Relationship between Wasserstein distance and accuracy across different model architectures in multiclass classification on the CIFAR-10 dataset with 8000 points. Connected points show consecutive epochs for each model.

Figure 5.12 shows the relationship between model test accuracy and the Wasserstein distance computed on the *training* dataset. Comparing it to Figure 5.9, we see that while the correlation is not as strong and the measurements show more variance, there remains a clear relationship between training set topology and test performance. This suggests that topological metrics computed during training could serve as early indicators of model quality, without the need to use a separate test set. This would allow training models on all available data without sacrificing the ability to evaluate their generalization.

As in our previous experiments, the total bar length metric (shown in Figure D.2) exhibits similar trends but with significantly more noise, making Wasserstein distance a more reliable indicator of model quality.

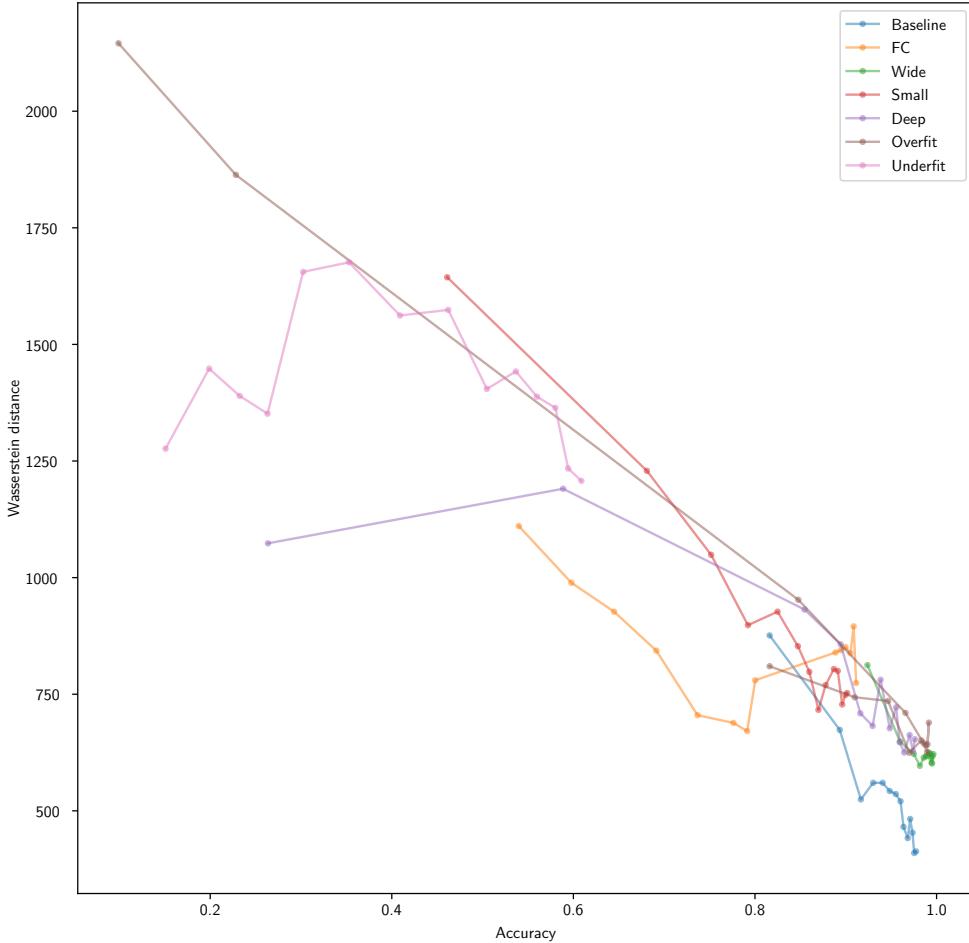


Figure 5.12: Relationship between test accuracy and Wasserstein distance computed on training data across different model architectures.

5.5 Discussion and Future Work

Our results demonstrate that topological analysis provides valuable insights into the behavior of neural network classifiers. The strong correlation between topological metrics and model accuracy suggests that well-performing models develop similar topological structures in their decision boundaries. This is particularly evident in our multiclass analysis, where high-accuracy models converge to similar Wasserstein distances despite having different architectures.

The observation that topological metrics computed on training data correlate with test performance, albeit more weakly, suggests potential applications in model evaluation without requiring a separate test set. Furthermore, the increase in Wasserstein distance during overfitting indicates that topological analysis could help detect when a model begins to overfit, potentially before

5. RESULTS AND DISCUSSION

it becomes apparent in validation metrics.

Our extension of the LVR complex to multiclass classification not only preserves more topological information but also proves more computationally efficient than binary decomposition. This efficiency gain becomes particularly significant for problems in high dimensions, suggesting that our approach could scale better to complex real-world applications.

Our experiments with CIFAR-10 revealed important insights about scaling topological analysis to higher dimensions. While the method’s effectiveness initially degraded when moving from MNIST to higher-dimensional CIFAR-10 data, we found that increasing the sample size restored the strong correlation between topological metrics and model performance. This suggests that topological analysis of decision boundaries remains viable even in high-dimensional spaces, provided we have enough points.

Several directions for potential future research emerge from this work:

1. *Topological regularization*: Given the correlation between topology and generalization, incorporating topological metrics into the training objective could help guide models toward better decision boundaries.
2. *Early stopping criteria*: The relationship between topological metrics and overfitting could be developed into practical stopping criteria for training.
3. *Architecture selection*: The convergence of successful models to similar topological structures suggests that topological analysis could aid in comparing and selecting model architectures.
4. *Relationship between input dimensionality and sample size*: Investigating this relationship could help determine optimal sample sizes for datasets of varying dimensionality, allowing to balance computational efficiency with topological accuracy.

Appendix A

Visualization of LVR complex construction

The following figures illustrate how the LVR complex captures the topology of example datasets. All visualizations show the 2-skeleton (vertices, edges, and triangles) of the complexes. Some 2-simplices are not shown for clarity.

A. VISUALIZATION OF LVR COMPLEX CONSTRUCTION

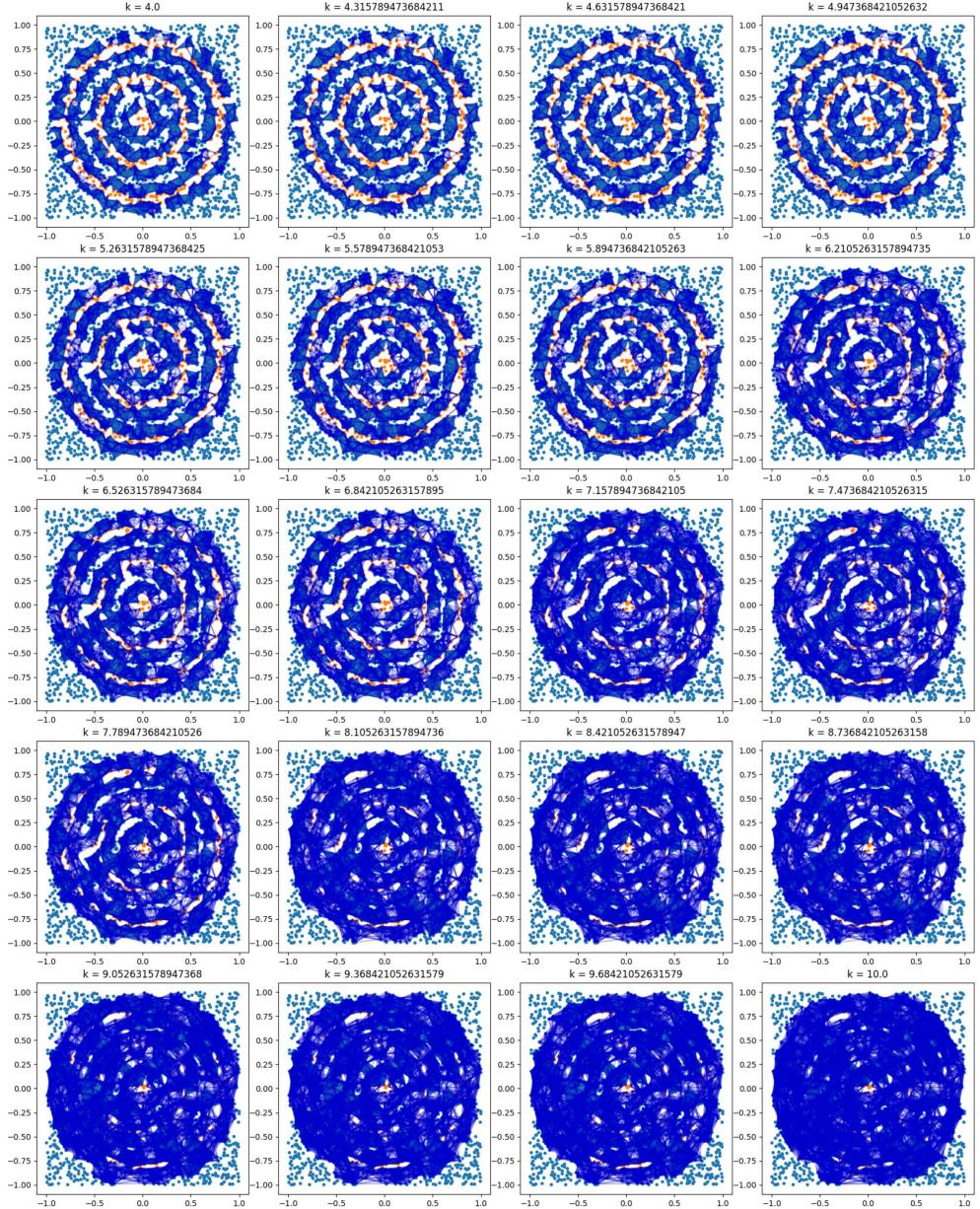


Figure A.1: Evolution of the LVR complex without CC on the nested dataset as the filtration parameter increases (left to right, top to bottom).

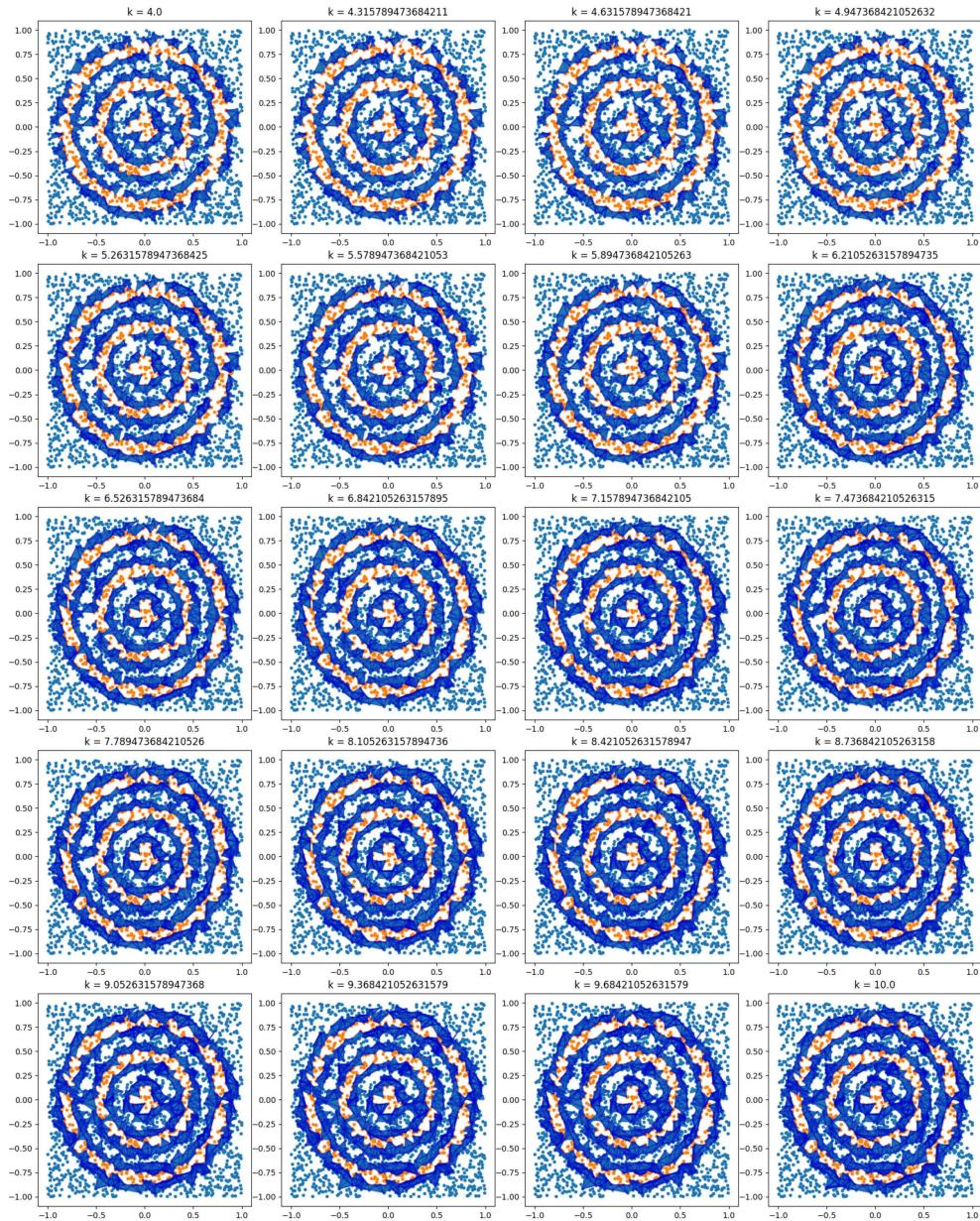


Figure A.2: Evolution of the LVR complex with CC on the nested dataset as the filtration parameter increases (left to right, top to bottom).

A. VISUALIZATION OF LVR COMPLEX CONSTRUCTION

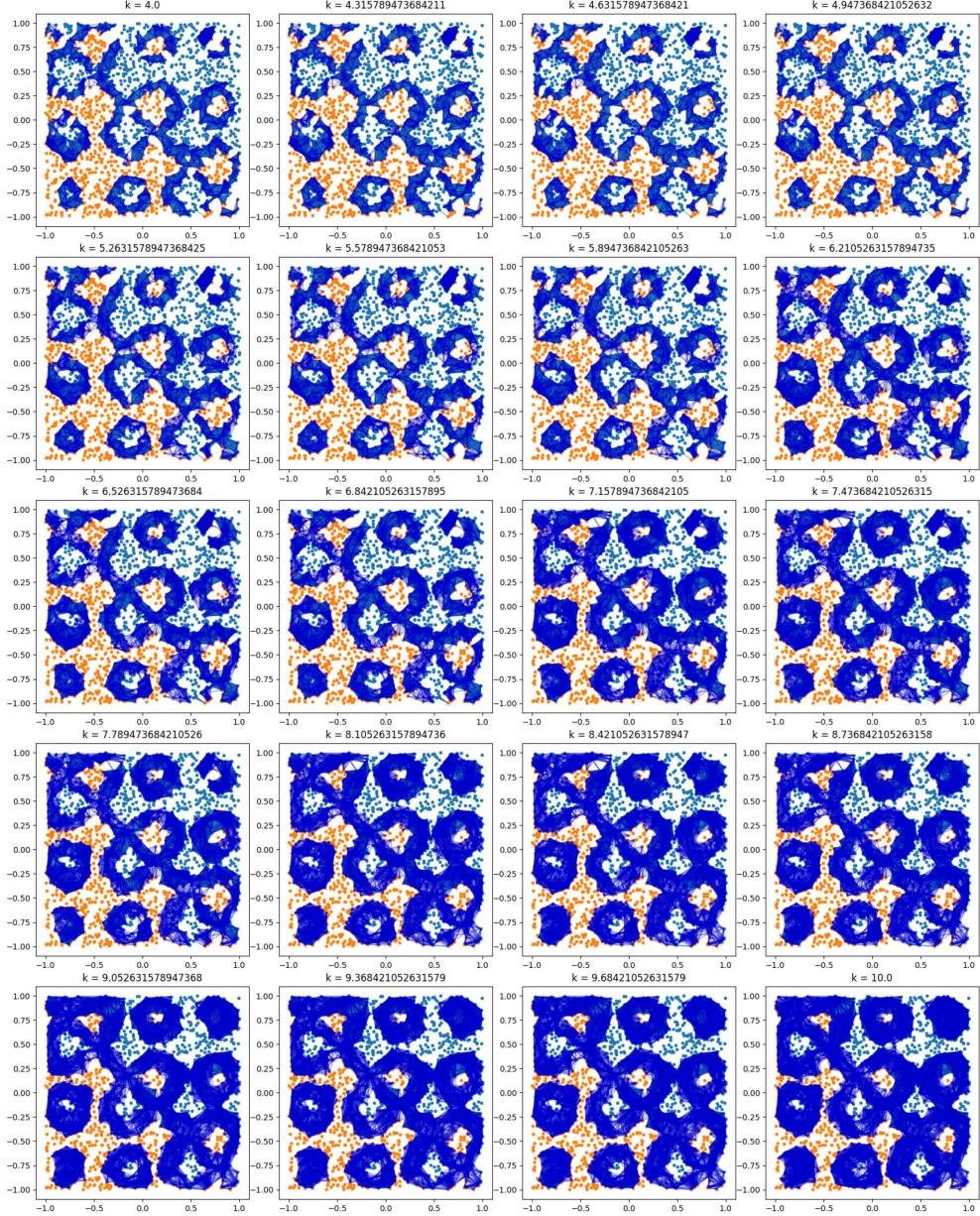


Figure A.3: Evolution of the LVR complex without CC on the manyholes dataset as the filtration parameter increases (left to right, top to bottom).

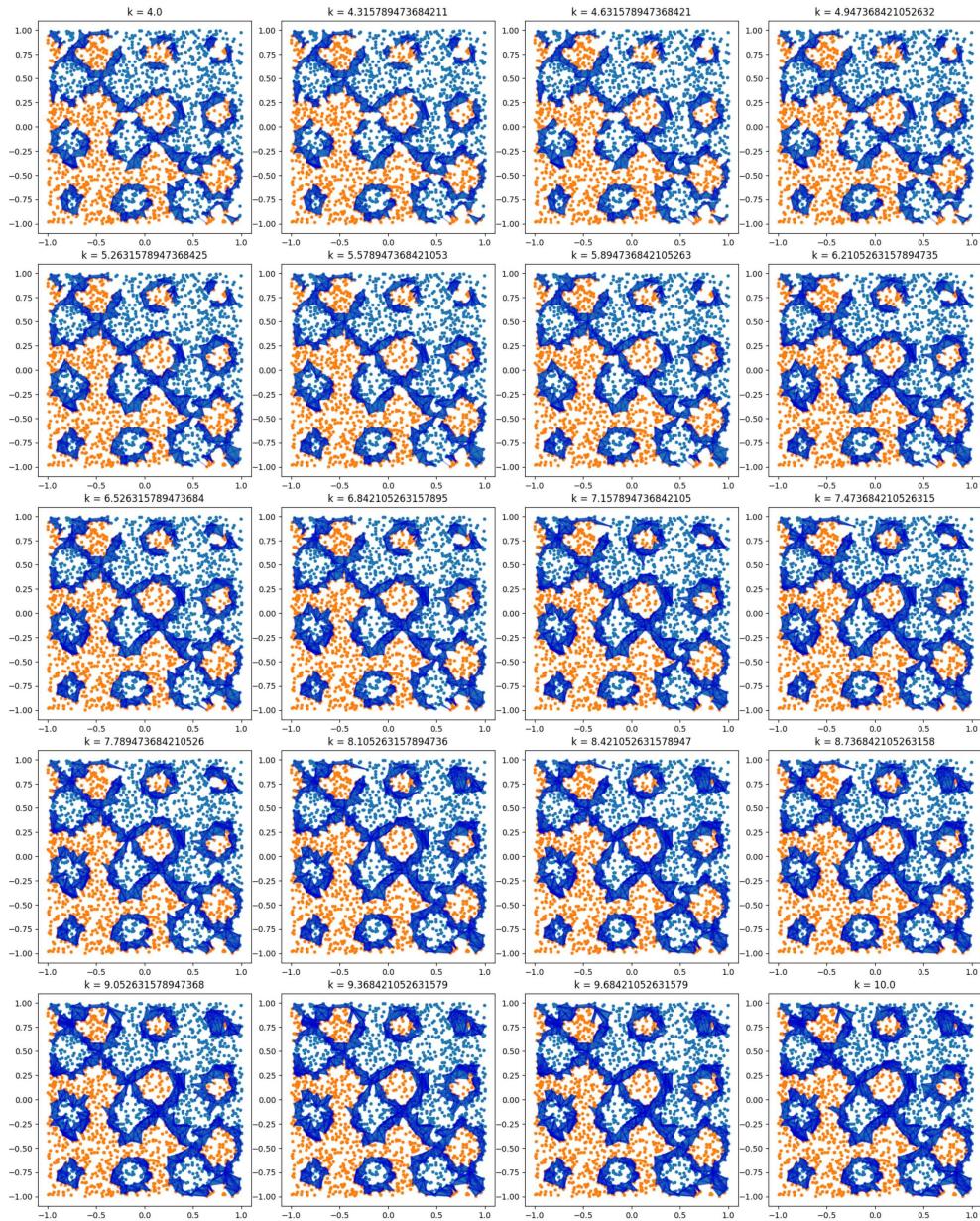


Figure A.4: Evolution of the LVR complex with CC on the manyholes dataset as the filtration parameter increases (left to right, top to bottom).

Appendix B

Visualization of Dowker complex construction

The following figures illustrate how the Dowker complex captures the topology of example datasets. All visualizations show the 2-skeleton (vertices, edges, and triangles) of the complexes. Some 2-simplices are not shown for clarity.

B. VISUALIZATION OF DOWKER COMPLEX CONSTRUCTION

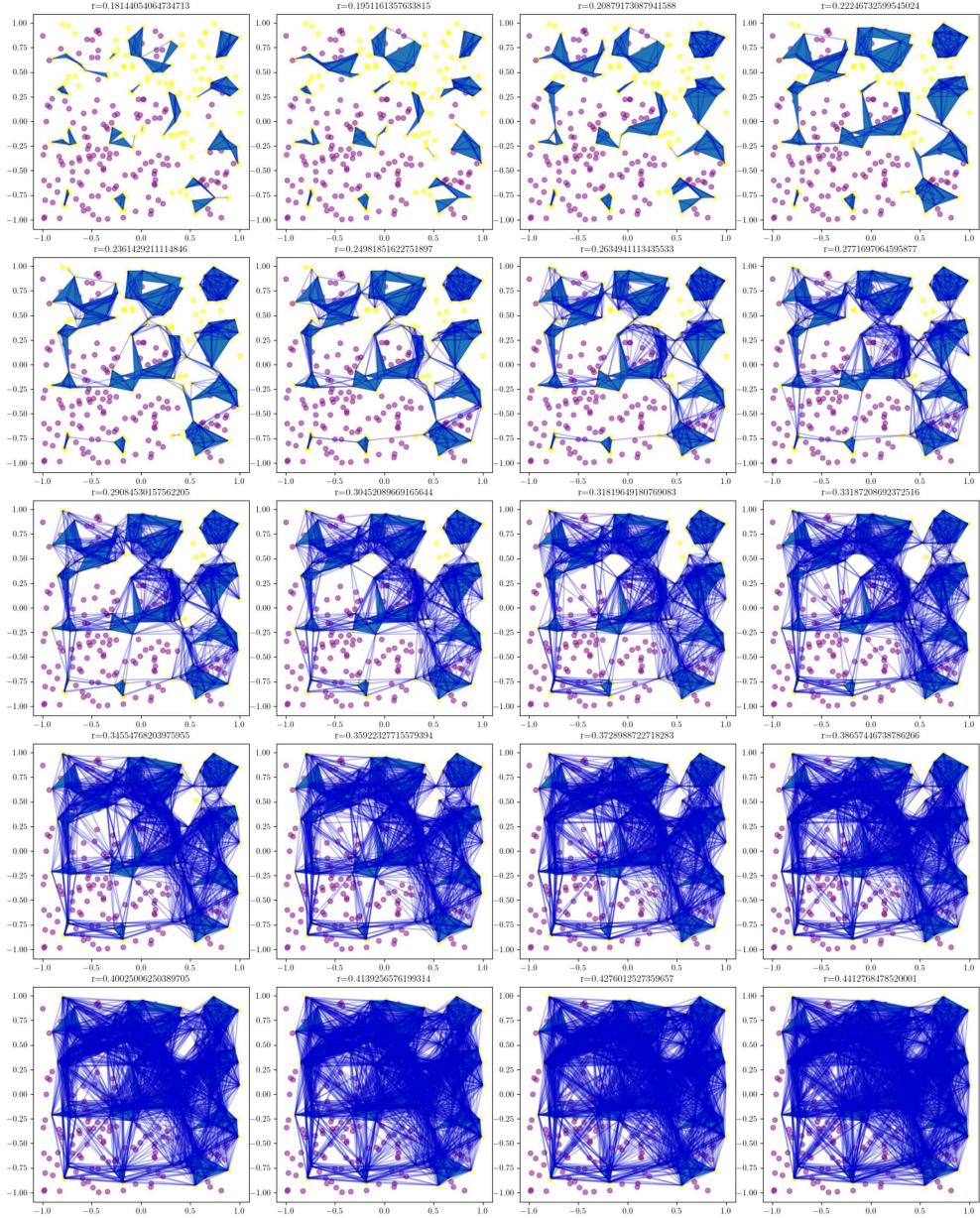


Figure B.1: Evolution of the Dowker complex without CC on the manyholes dataset as the filtration parameter increases (left to right, top to bottom).

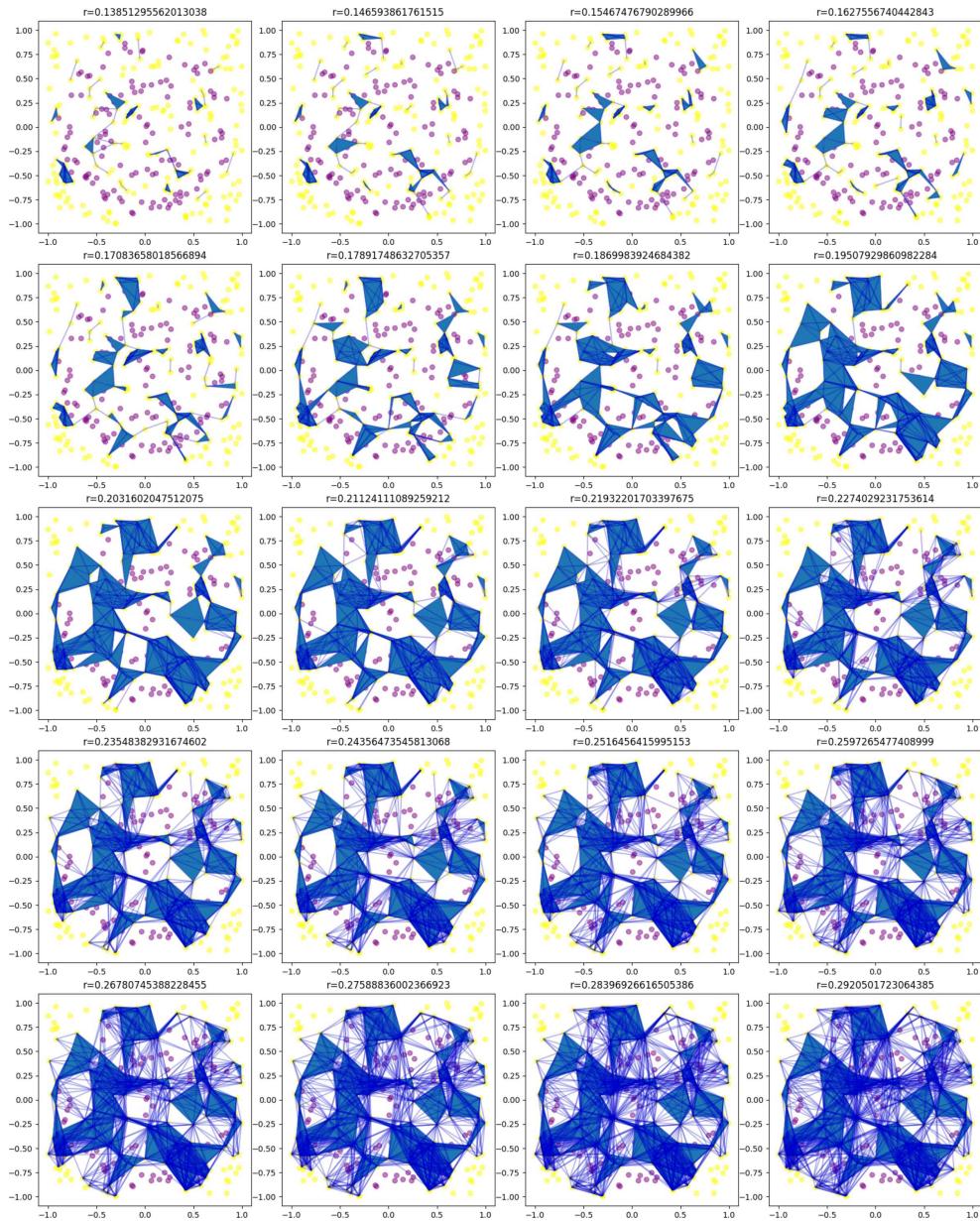


Figure B.2: Evolution of the Dowker complex with CC on the manyholes dataset as the filtration parameter increases (left to right, top to bottom).

Appendix C

Binary classification topological metrics

The following figures provide comprehensive results for binary classification experiments across all class pairs.

C. BINARY CLASSIFICATION TOPOLOGICAL METRICS

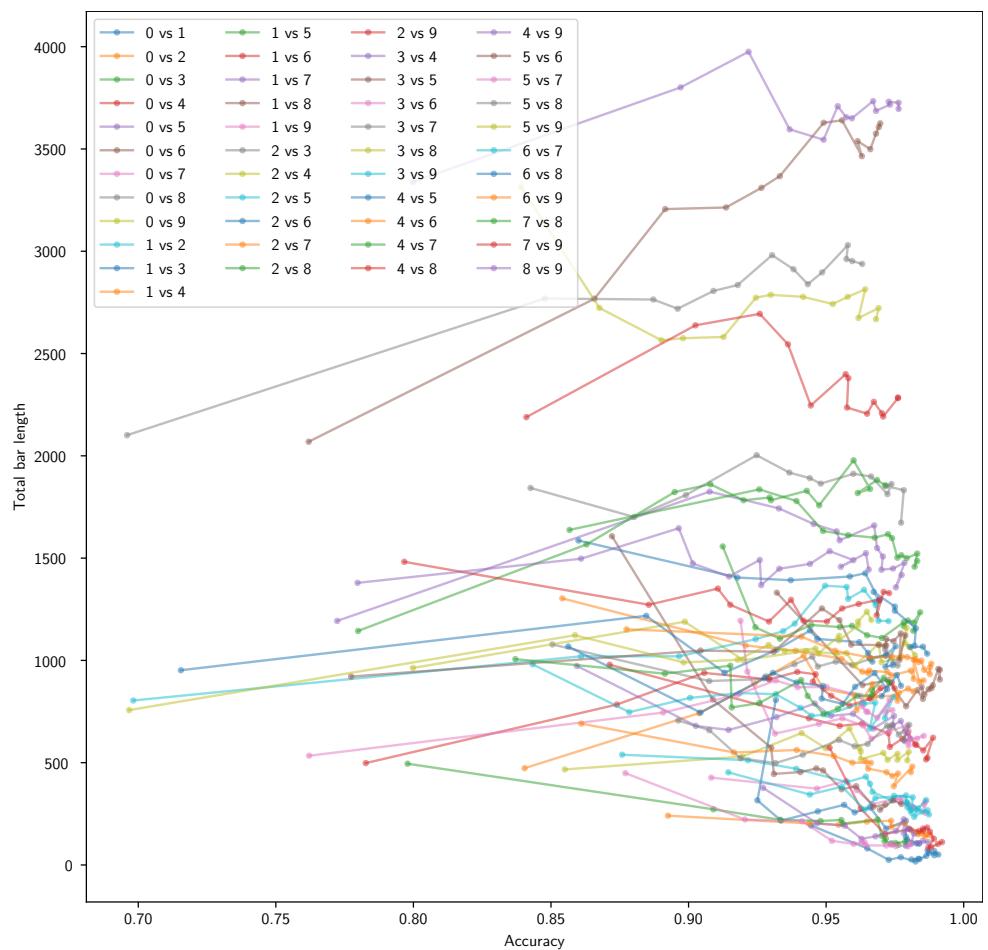


Figure C.1: Relationship between accuracy and total bar length difference across different class pairs.

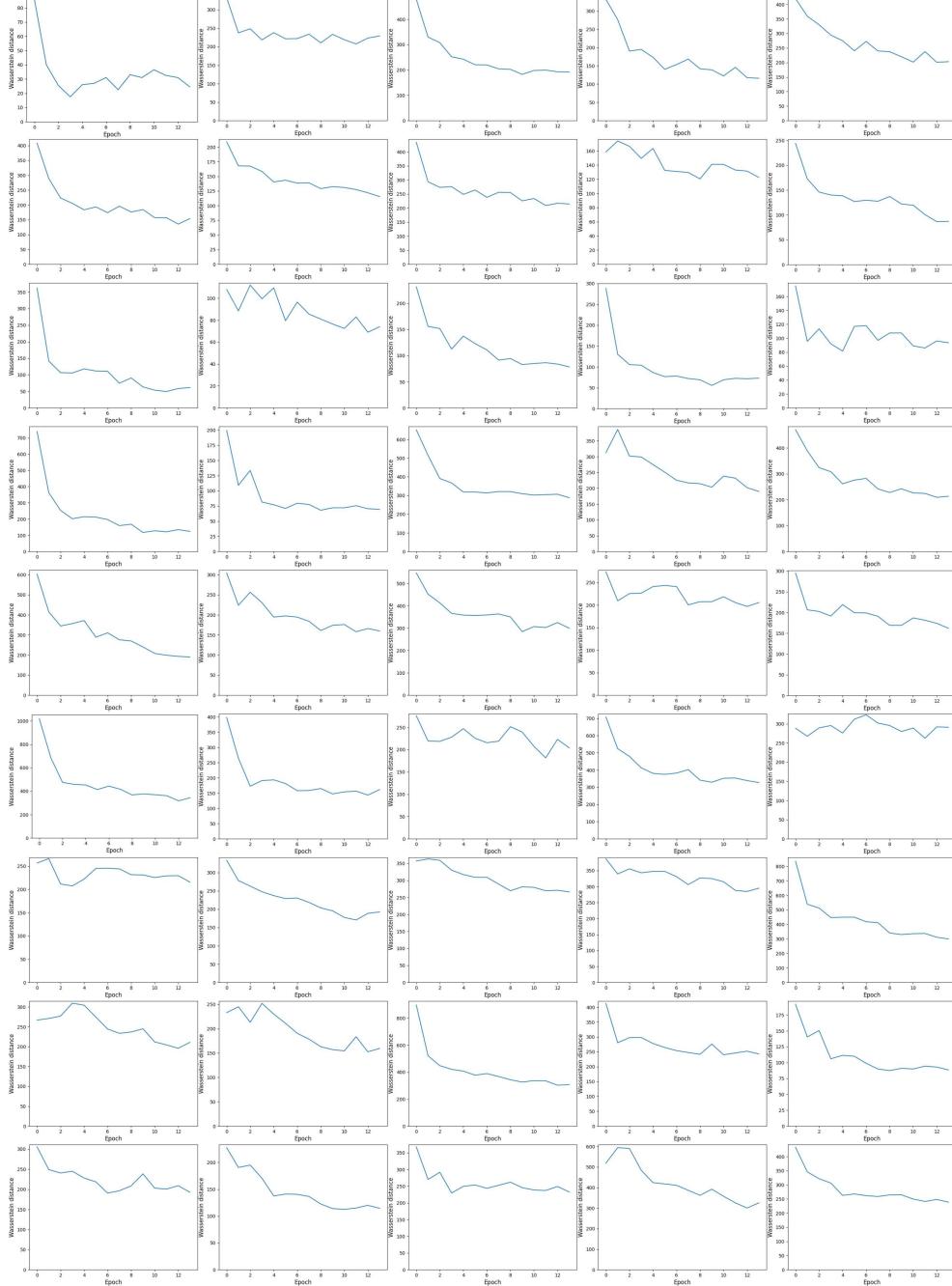


Figure C.2: Evolution of Wasserstein distances during training for all binary classification pairs.

C. BINARY CLASSIFICATION TOPOLOGICAL METRICS

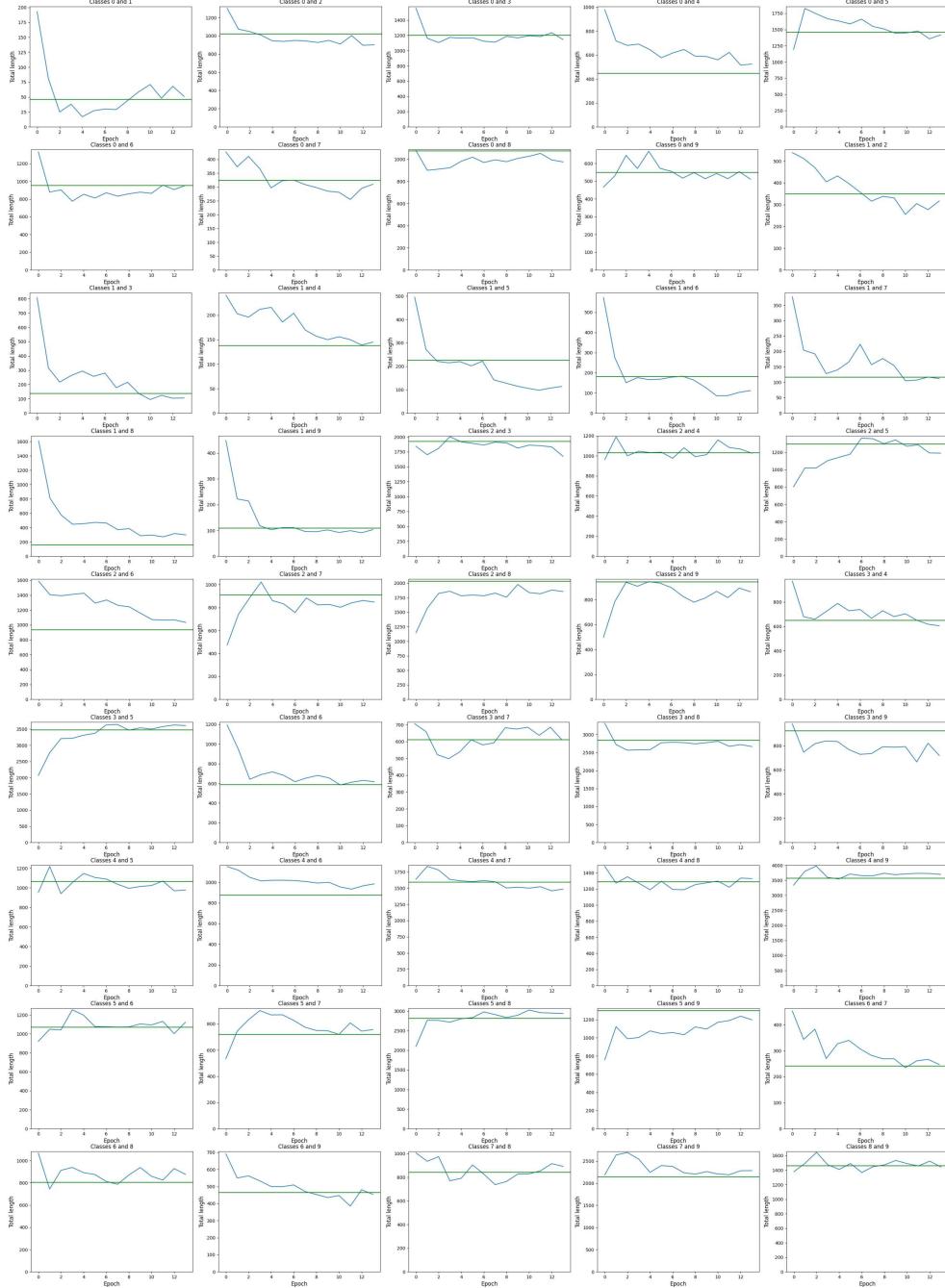


Figure C.3: Evolution of total bar lengths during training for all binary classification pairs.

Appendix D

Multiclass classification topological metrics

D. MULTICLASS CLASSIFICATION TOPOLOGICAL METRICS

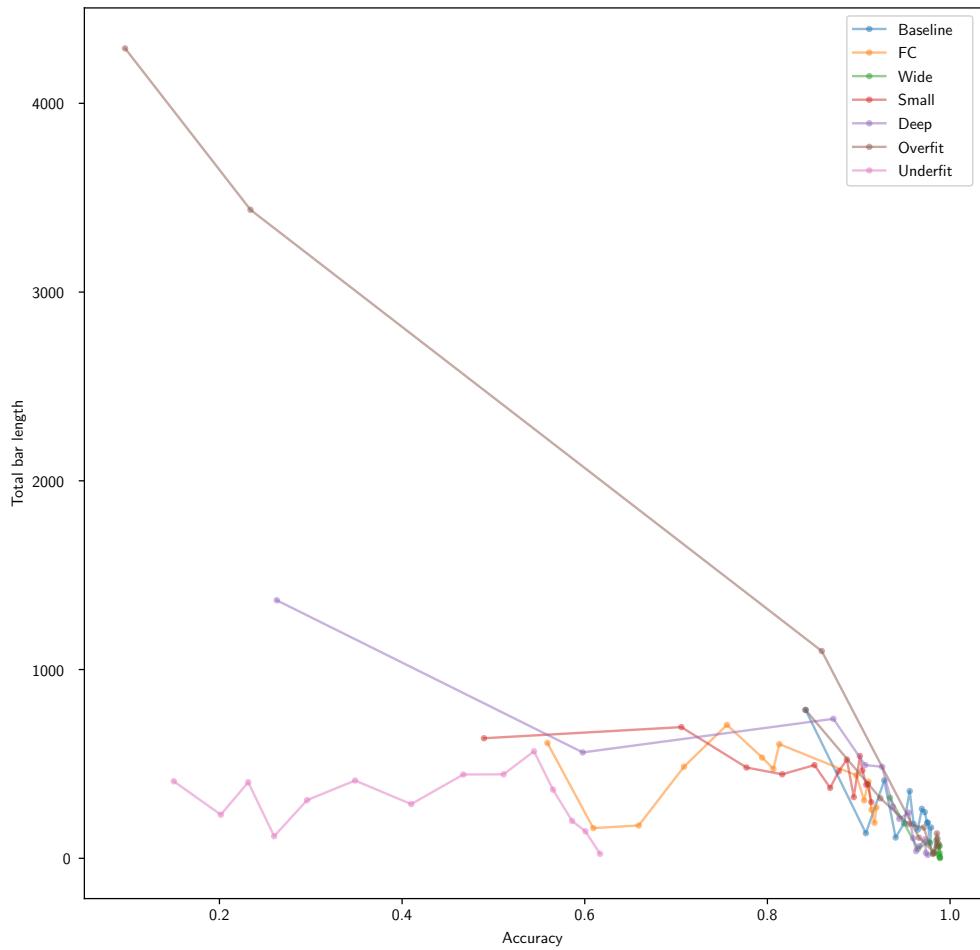


Figure D.1: Relationship between accuracy and total bar length difference in different models.

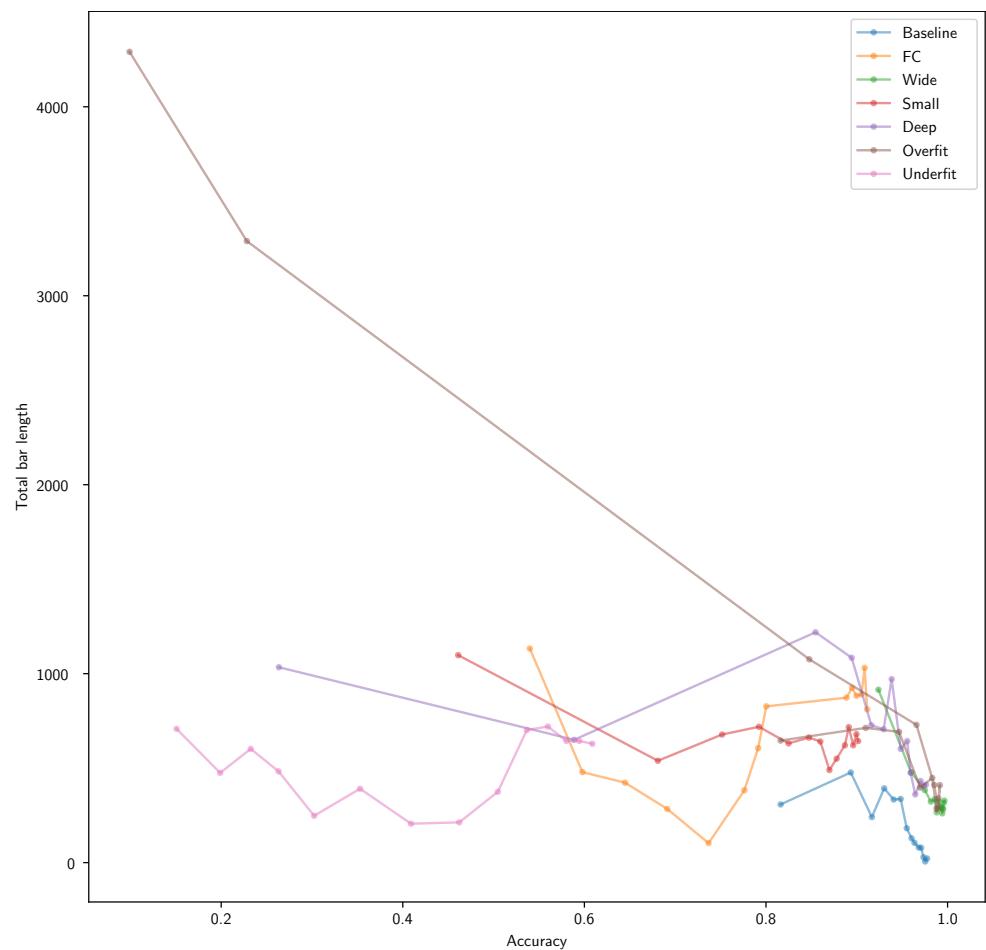


Figure D.2: Relationship between test accuracy and total bar length difference computed on training data.

D. MULTICLASS CLASSIFICATION TOPOLOGICAL METRICS

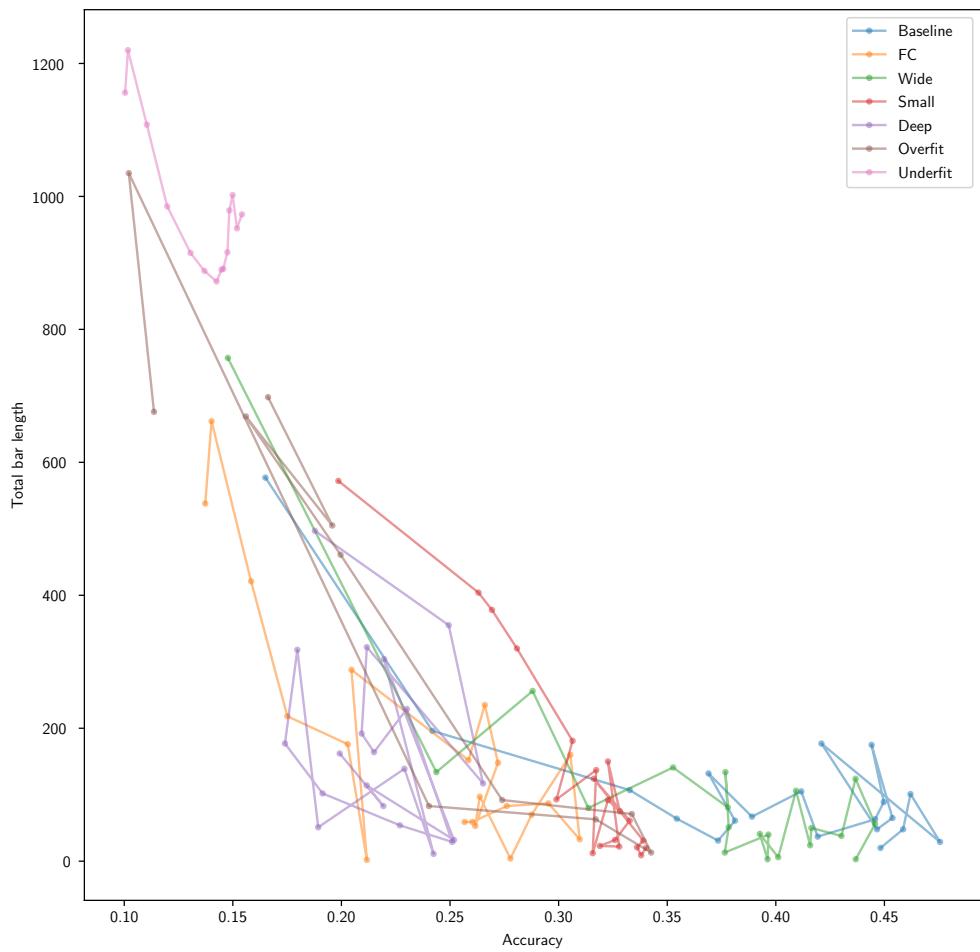


Figure D.3: Relationship between total bar length difference and accuracy across different model architectures in multiclass classification on the CIFAR-10 dataset with 2000 points. Connected points show consecutive epochs for each model.

Bibliography

- [1] Rubén Ballester, Carles Casacuberta, and Sergio Escalera. Topological data analysis for neural network analysis: A comprehensive survey, 2024.
- [2] Ulrich Bauer. Ripser: efficient computation of Vietoris-Rips persistence barcodes. *J. Appl. Comput. Topol.*, 5(3):391–423, 2021.
- [3] Tolga Birdal, Aaron Lou, Leonidas Guibas, and Umut Şimşekli. Intrinsic dimension, persistent homology and generalization in neural networks, 2021.
- [4] Rickard Brüel Gabrielsson and Gunnar Carlsson. Exposition and interpretation of the topology of neural networks. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 1069–1076, 2019.
- [5] Ciprian A. Corneanu, Meysam Madadi, Sergio Escalera, and Aleix M. Martinez. What does it mean to learn in deep networks? and, how does one detect adversarial attacks? In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4752–4761, 2019.
- [6] Li Deng. The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012.
- [7] C. H. Dowker. Homology groups of relations. *Annals of Mathematics*, 56(1):84–95, 1952.
- [8] H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. Applied Mathematics. American Mathematical Society, 2010.
- [9] Gamaleldin F. Elsayed, Dilip Krishnan, Hossein Mobahi, Kevin Regan, and Samy Bengio. Large margin deep networks for classification, 2018.

BIBLIOGRAPHY

- [10] Maxime Gabella. Topology of learning in feedforward neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(8):3588–3592, 2021.
- [11] Thomas Gebhart, Paul Schrater, and Alan Hylton. Characterizing the shape of activation space in deep neural networks, 2019.
- [12] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.
- [13] Weizhi Li, Gautam Dasarathy, Karthikeyan Natesan Ramamurthy, and Visar Berisha. Finding the homology of decision boundaries with active learning. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 8355–8365. Curran Associates, Inc., 2020.
- [14] Weizhi Li, Gautam Dasarathy, Karthikeyan Natesan Ramamurthy, and Visar Berisha. Finding the homology of decision boundaries with active learning. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 8355–8365. Curran Associates, Inc., 2020.
- [15] Meng Liu, Tamal K. Dey, and David F. Gleich. Topological structure of complex predictions. *Nature Machine Intelligence*, 5(12):1382–1389, Dec 2023.
- [16] Ephy Love, Benjamin Filippenko, Vasileios Maroulas, and Gunnar E. Carlsson. Topological convolutional neural networks. In *TDA & Beyond*, 2020.
- [17] Quynh Nguyen. On connected sublevel sets in deep learning, 2019.
- [18] Julián Burella Pérez, Sydney Hauke, Umberto Lupo, Matteo Caorsi, and Alberto Dassatti. giotto-ph: A python library for high-performance computation of persistent homology of vietoris–rips filtrations, 2021.
- [19] Karthikeyan Natesan Ramamurthy, Kush R. Varshney, and Krishnan Mody. Topological data analysis of decision boundaries with application to model selection, 2018.
- [20] Rieck, Bastian Alexander, Togninalli, Matteo, Bock, Christian, Moor, Michael, Horn, Max, Gumbisch, Thomas, and Borgwardt, Karsten. Neural persistence: A complexity measure for deep neural networks using algebraic topology. 2023.

Bibliography

- [21] Gurjeet Singh, Facundo Memoli, and Gunnar Carlsson. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition . In M. Botsch, R. Pajarola, B. Chen, and M. Zwicker, editors, *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007.
- [22] Bernadette J. Stolz, Jagdeep Dhesi, Joshua A. Bull, Heather A. Harrington, Helen M. Byrne, and Iris H. R. Yoon. Relational persistent homology for multispecies data with application to the tumor microenvironment, 2023.
- [23] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms, 2017.
- [24] Simon Zhang, Mengbai Xiao, and Hao Wang. Gpu-accelerated computation of vietoris-rips persistence barcodes. In *36th International Symposium on Computational Geometry (SoCG 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.
- [25] Umut Şimşekli, Ozan Sener, George Deligiannidis, and Murat A Erdogan. Hausdorff dimension, heavy tails, and generalization in neural networks*. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(12):124014, December 2021.

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):

First name(s):

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the '[Citation etiquette](#)' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date

Signature(s)

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.