

# Topological Analysis of Decision Boundaries

Maxim Mikhaylov with supervision by Dr. Patrick Schnider

ETH Zürich

# Introduction

- ▶ Machine learning classifiers partition their input space into regions corresponding to different class labels  $1 \dots n$
- ▶ The boundaries between these regions, *decision boundaries*, are important for generalization
- ▶ Let's use topological data analysis to study how decision boundaries evolve during training to detect overfitting

Illustration of decision boundaries in 2D here.

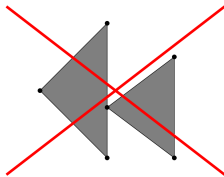
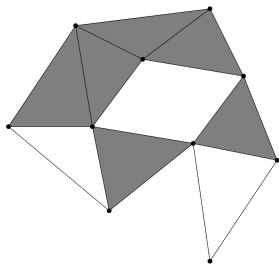
# Theoretical background: Overfitting

- ▶ Model fits training data too closely, fails to generalize to unseen data
- ▶ Performance improves on training data, but degrades on validation data
- ▶ Maybe this is reflected in the topology of decision boundaries?

Illustration of overfitting here; plot of training acc improving, validation acc dropping.

# Theoretical background: Persistent homology

- ▶ Simplicial complex: a collection of simplices glued together “nicely”



## Theoretical background: Persistent homology

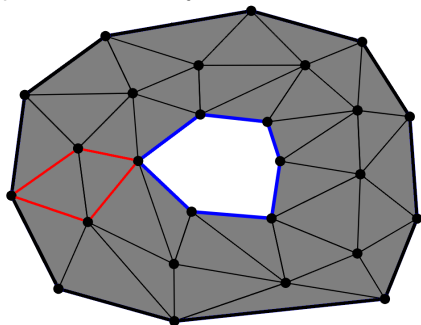
- ▶ Simplicial complex: a collection of simplices glued together “nicely”
- ▶ Homology groups of dimension  $p$ : measure the number of  $p$ -dimensional holes in a simplicial complex

# Theoretical background: Persistent homology

- ▶ Simplicial complex: a collection of simplices glued together “nicely”
- ▶ Homology groups of dimension  $p$ : measure the number of  $p$ -dimensional holes in a simplicial complex
  - ▶  $p = 0$ : connected components

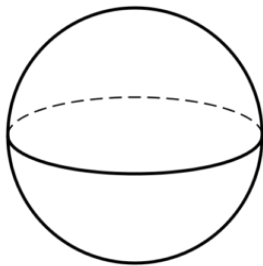
# Theoretical background: Persistent homology

- ▶ Simplicial complex: a collection of simplices glued together “nicely”
- ▶ Homology groups of dimension  $p$ : measure the number of  $p$ -dimensional holes in a simplicial complex
  - ▶  $p = 0$ : connected components
  - ▶  $p = 1$ : unfilled cycles



# Theoretical background: Persistent homology

- ▶ Simplicial complex: a collection of simplices glued together “nicely”
- ▶ Homology groups of dimension  $p$ : measure the number of  $p$ -dimensional holes in a simplicial complex
  - ▶  $p = 0$ : connected components
  - ▶  $p = 1$ : unfilled cycles
  - ▶  $p = 2$ : unfilled voids



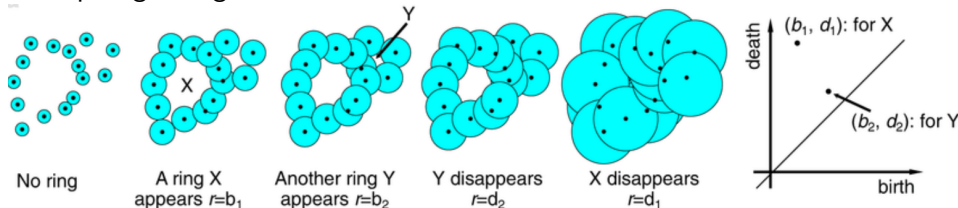


# Theoretical background: Persistent homology

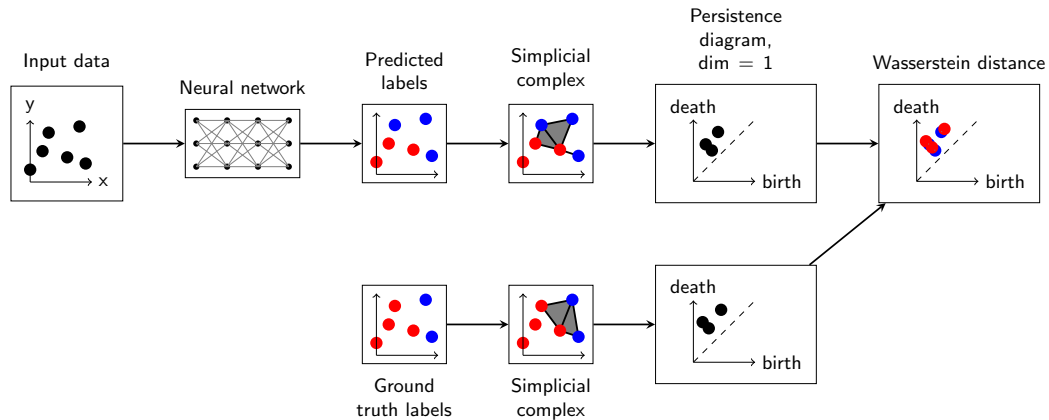
- ▶ Simplicial complex: a collection of simplices glued together “nicely”
- ▶ Homology groups of dimension  $p$ : measure the number of  $p$ -dimensional holes in a simplicial complex
  - ▶  $p = 0$ : connected components
  - ▶  $p = 1$ : unfilled cycles
  - ▶  $p = 2$ : unfilled voids
- ▶ *Persistent* homology: homology groups of a simplicial complex as it evolves

# Theoretical background: Persistent homology

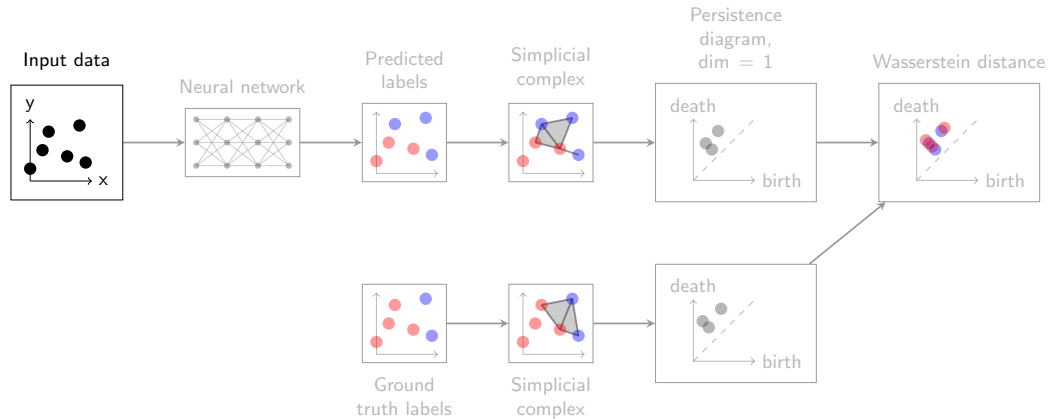
- ▶ Simplicial complex: a collection of simplices glued together “nicely”
- ▶ Homology groups of dimension  $p$ : measure the number of  $p$ -dimensional holes in a simplicial complex
  - ▶  $p = 0$ : connected components
  - ▶  $p = 1$ : unfilled cycles
  - ▶  $p = 2$ : unfilled voids
- ▶ *Persistent* homology: homology groups of a simplicial complex as it evolves
- ▶ Example: growing balls



# Pipeline

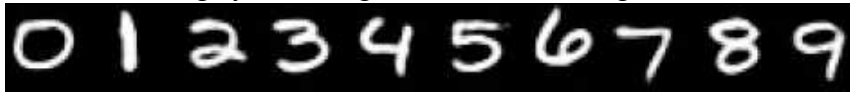


# Pipeline



## Input data

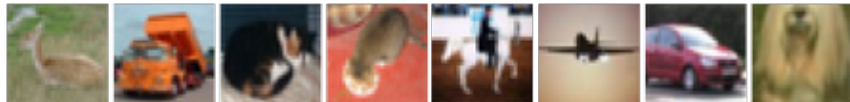
- ▶ MNIST: 28x28 grayscale images of handwritten digits, dim = 784



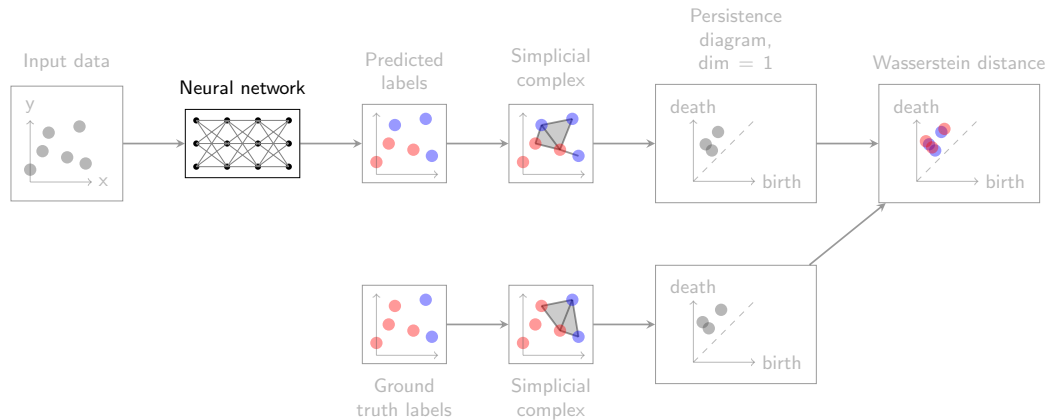
- ▶ FashionMNIST: 28x28 grayscale images of fashion articles, dim = 784



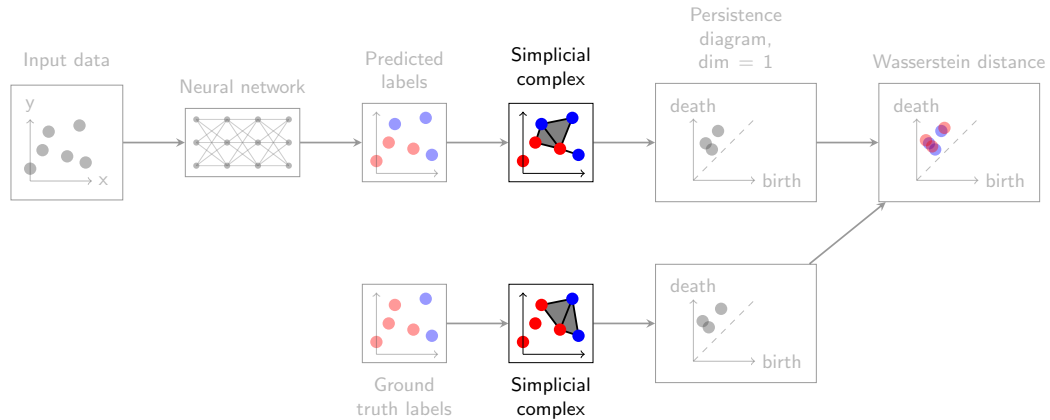
- ▶ CIFAR-10: 32x32 color images of 10 classes, dim = 3072



# Pipeline



# Pipeline

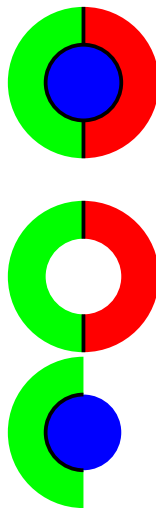


# Binarization

Previous work only considered binary classification by splitting a dataset with  $n$  classes into  $\binom{n}{2}$  binary datasets. This changes the homology groups:

- ▶ Original has  $H_1 \not\cong 0$
- ▶ All binary decompositions have  $H_1 \cong 0$

It is better to avoid binarization.

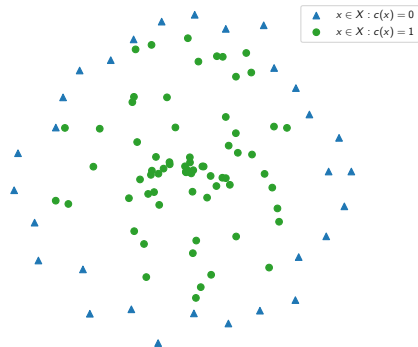




# Labeled Vietoris-Rips complex

- ▶ Set of points  $X$
- ▶ Labels  $c : X \rightarrow \mathbb{Z}_k$
- ▶ Parameter  $\varepsilon$

Constructed in three steps:

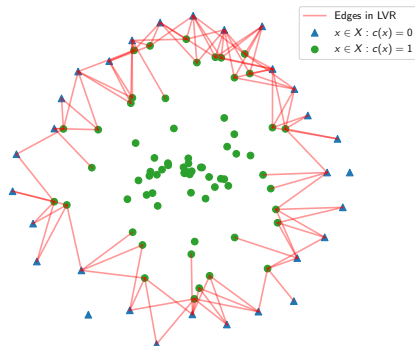


# Labeled Vietoris-Rips complex

- ▶ Set of points  $X$
- ▶ Labels  $c : X \rightarrow \mathbb{Z}_k$
- ▶ Parameter  $\varepsilon$

Constructed in three steps:

1. Create a graph  $G_\varepsilon$  with vertex set  $X$  by adding an edge between points  $x_i, x_j \in X$  iff:
  - ▶  $\|x_i - x_j\| \leq \varepsilon$  (points are close enough)
  - ▶  $c(x_i) \neq c(x_j)$  (different classes)

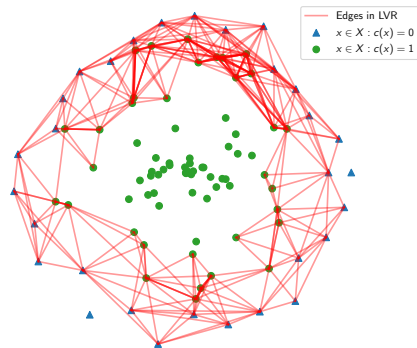


# Labeled Vietoris-Rips complex

- ▶ Set of points  $X$
- ▶ Labels  $c : X \rightarrow \mathbb{Z}_k$
- ▶ Parameter  $\varepsilon$

Constructed in three steps:

1. Create a graph  $G_\varepsilon$  with vertex set  $X$  by adding an edge between points  $x_i, x_j \in X$  iff:
  - ▶  $\|x_i - x_j\| \leq \varepsilon$  (points are close enough)
  - ▶  $c(x_i) \neq c(x_j)$  (different classes)
2. Add edges between all 2-hop neighbors in  $G_\varepsilon$

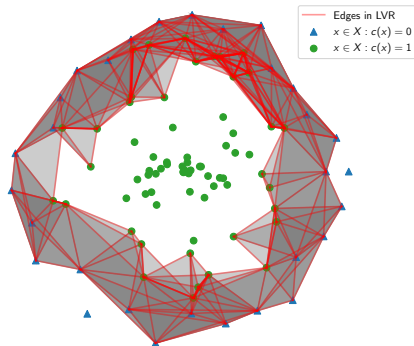


# Labeled Vietoris-Rips complex

- ▶ Set of points  $X$
- ▶ Labels  $c : X \rightarrow \mathbb{Z}_k$
- ▶ Parameter  $\varepsilon$

Constructed in three steps:

1. Create a graph  $G_\varepsilon$  with vertex set  $X$  by adding an edge between points  $x_i, x_j \in X$  iff:
  - ▶  $\|x_i - x_j\| \leq \varepsilon$  (points are close enough)
  - ▶  $c(x_i) \neq c(x_j)$  (different classes)
2. Add edges between all 2-hop neighbors in  $G_\varepsilon$
3. Standard Vietoris-Rips construction: include simplex if all faces are included

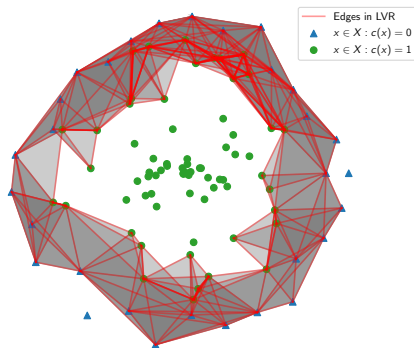


# Labeled Vietoris-Rips complex

- ▶ Set of points  $X$
- ▶ Labels  $c : X \rightarrow \mathbb{Z}_k$
- ▶ Parameter  $\varepsilon$

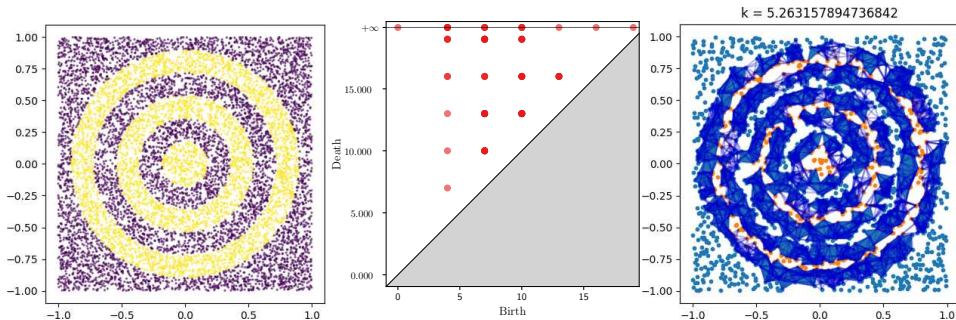
Constructed in three steps:

1. Create a graph  $G_\varepsilon$  with vertex set  $X$  by adding an edge between points  $x_i, x_j \in X$  iff:
    - ▶  $\|x_i - x_j\| \leq \varepsilon$  (points are close enough)
    - ▶  $c(x_i) \neq c(x_j)$  (different classes)
  2. Add edges between all 2-hop neighbors in  $G_\varepsilon$
  3. Standard Vietoris-Rips construction: include simplex if all faces are included
- ▶ Simplices cross the boundary
  - ▶ Applicable to multiple classes



## Synthetic 2D experiemnts

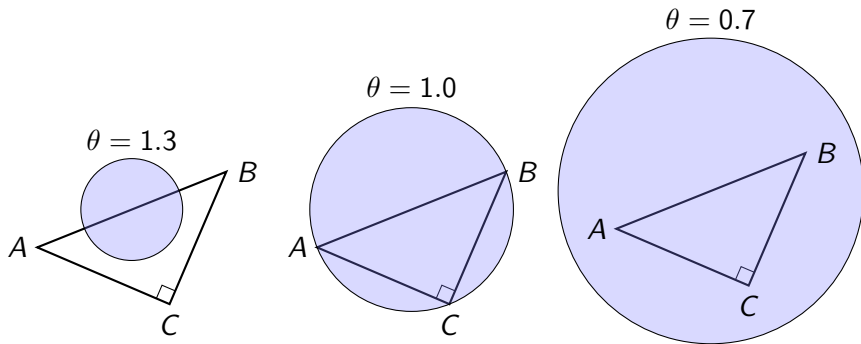
To see if the LVR complex recovers the homology of the decision boundary, we use synthetic 2D data with nested annuli.



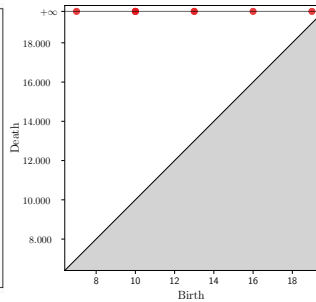
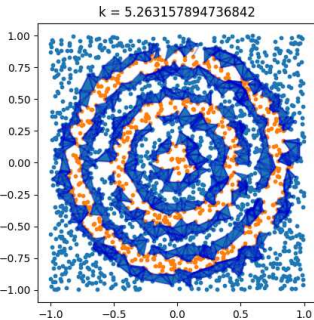
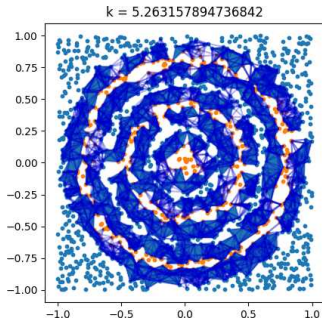
Edges cross the decision boundary multiple times  $\implies$  spurious holes appear.

## Circumcircle filtering

- ▶ Remove an edge  $AB$  if exists vertex  $C$  :  $|AB|^2 > (|AC|^2 + |BC|^2)\theta$ .
- ▶  $\theta$  is a parameter to relax the condition.



# Circumcircle filtering impact

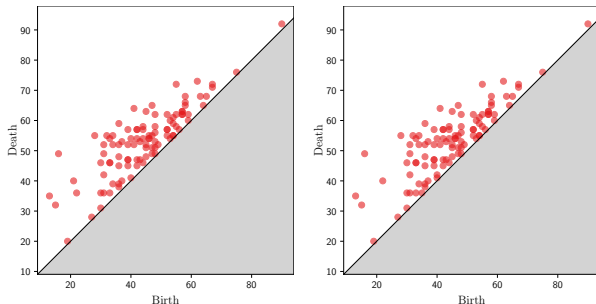




# Circumcircle filtering impact on high-dimensional data

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

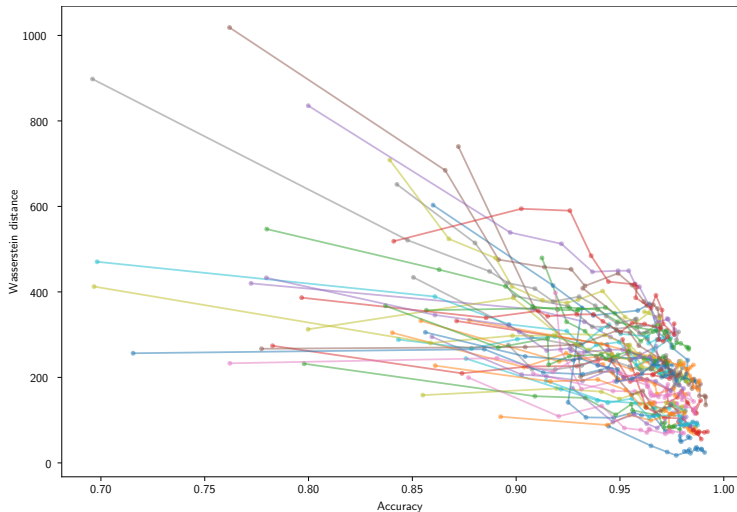
Maximum differences in Wasserstein distance between persistence diagrams with and without CC.



The two persistence diagrams that differ the most when CC is applied.

The difference is negligible  $\implies$  CC is not used

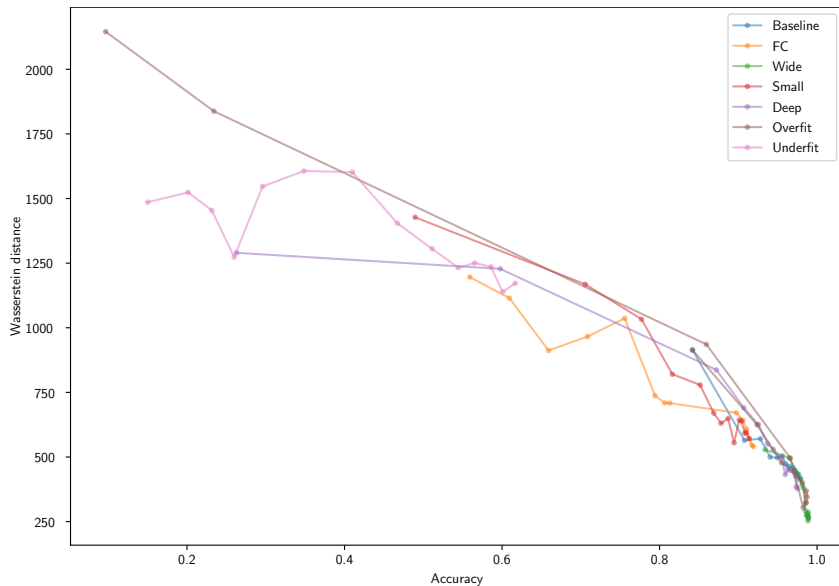
## Results: binary classification



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

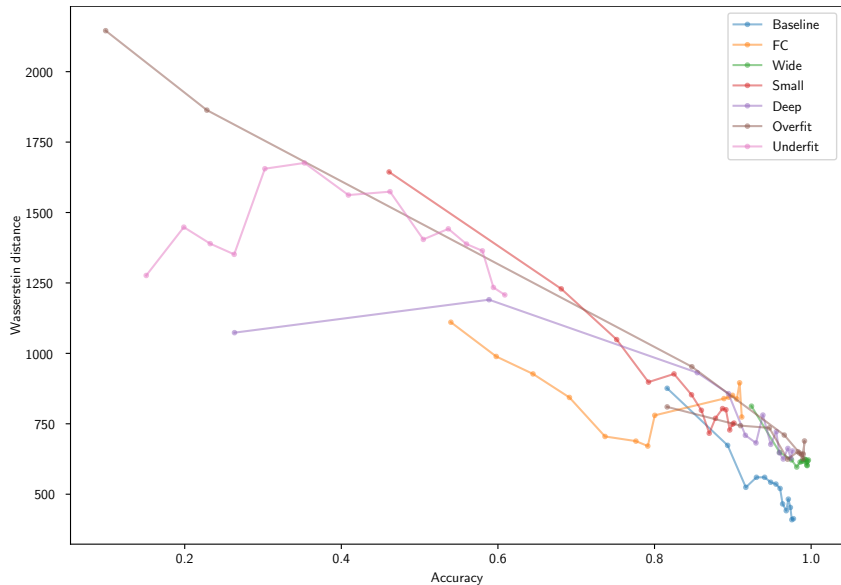
# Results: multiclass classification

Clear indication of overfit in the overfit model's trajectory



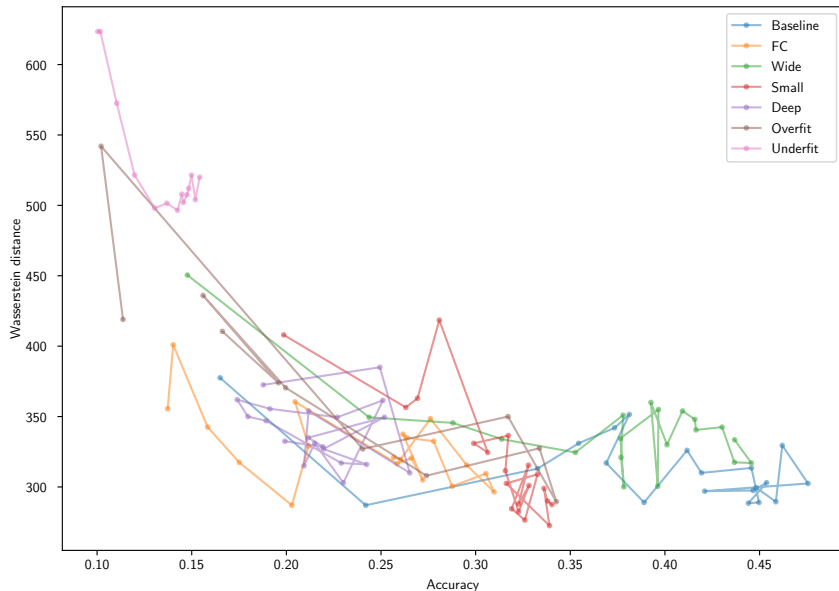
# Results: multiclass classification, train data

TDA on training  
data, accuracy  
on test data



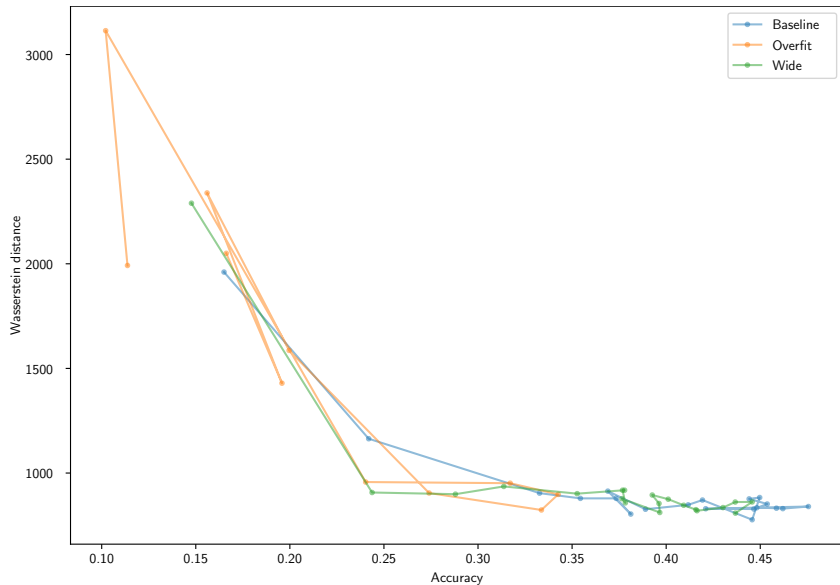
## Results: multiclass classification, CIFAR10, 2000 points

Weak correlation  
as dimensionality  
of data increases,  
number of  
datapoints is  
unchanged



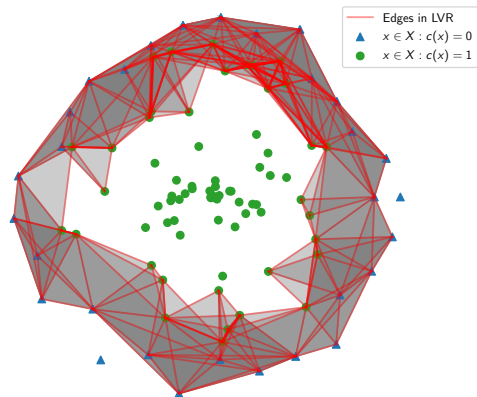
## Results: multiclass classification, CIFAR10, 8000 points

Increasing  
number of points  
improves  
correlation



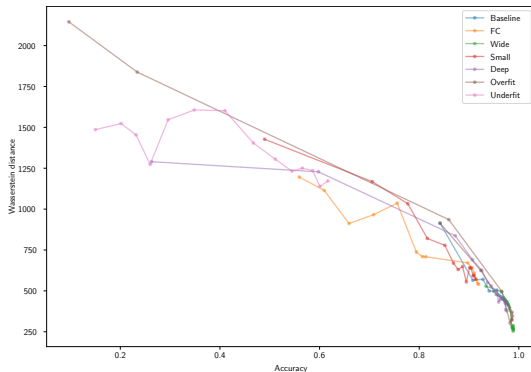
# Conclusion

- ▶ Multiclass LVR complex is better
  - ▶ Preserves more topological information
  - ▶ Stronger correlation
  - ▶ More computationally efficient



# Conclusion

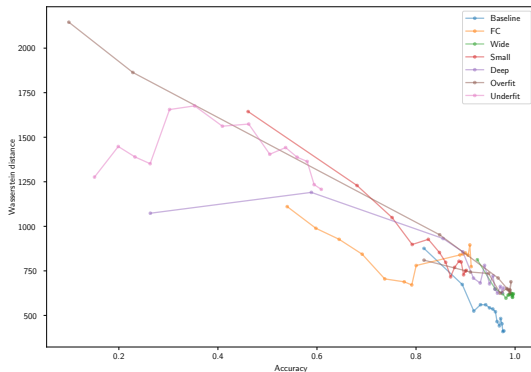
- ▶ Multiclass LVR complex is better
- ▶ TDA provides insights into classifier behavior
  - ▶ Strong correlation between Wasserstein distance and model accuracy
  - ▶ Wasserstein distance increases during overfitting





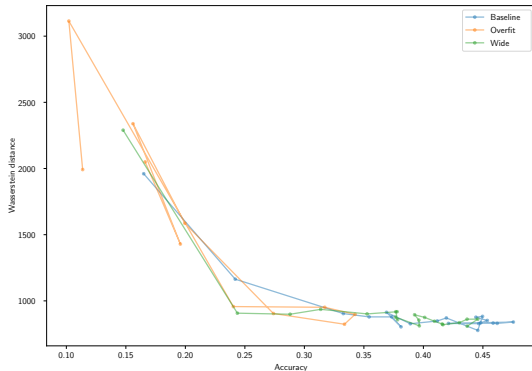
# Conclusion

- ▶ Multiclass LVR complex is better
- ▶ TDA provides insights into classifier behavior
- ▶ Training data decision boundary topology correlates with test performance
  - ▶ Enables model evaluation without separate test sets
  - ▶ Could serve as early overfitting detection



# Conclusion

- ▶ Multiclass LVR complex is better
- ▶ TDA provides insights into classifier behavior
- ▶ Training data decision boundary topology correlates with test performance
- ▶ Dimensionality scaling from CIFAR-10 experiments
  - ▶ Initial effectiveness decrease in higher dimensions
  - ▶ Performance restored by increasing sample size
  - ▶ Limitation: too many points required for very high-dimensional  $X$



# Conclusion

- ▶ Multiclass LVR complex is better
- ▶ TDA provides insights into classifier behavior
- ▶ Training data decision boundary topology correlates with test performance
- ▶ Dimensionality scaling from CIFAR-10 experiments

Thank you for your attention!