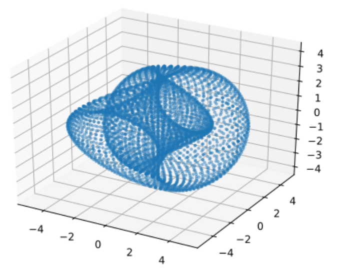
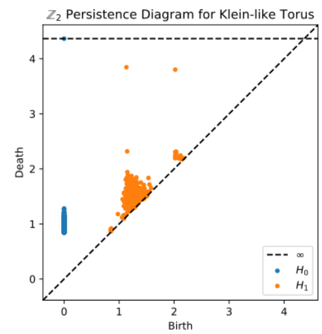


(a) The three-dimensional projection of the point cloud.



(b) The persistence diagram of the point-cloud.

Figure 1: A torus with a Klein bottle-like twist along with its persistence diagram.

3 Persistent Homology

Persistent homology is a technique to compute the topological features of a point-cloud dataset and returns these results in a persistence diagram.

3.1 Method

Given a point-cloud, persistent homology grows n-dimensional balls around each point and marks when these balls intersect with the balls of another point. By connecting these balls together, one can represent the data as simplicial complexes by chaining together simplices (generalizations of points, lines, and triangles in higher-dimensional space). The validity of this is a result of the abelian group structure imposed on the homology of our dataset. These simplicial complexes can then represent topological features such as loops. At some point, all points will intersect as the balls will grow to encompass the whole dataset. We record the birth and death times of these features and plot them in a persistence diagram.

3.2 Persistence Diagrams

We generated a variation of a torus with a Klein bottle-like twist, X, according to the following parametrization,

$$x = (R + 2P\cos\theta)\cos\phi \ ,$$
$$y = (R + 2P\cos\theta)\sin\phi \ ,$$
$$z = 2P\sin\theta\cos\frac{\phi}{2} \ ,$$
$$w = 2P\sin\theta\sin\frac{\phi}{2}$$

We display the point-cloud (only the first three dimensions) and its persistence diagram in figure 1. For $H_0(X)$, we can see that all features but one die soon after they are birthed, which is indicated by points residing close to the diagonal of the diagram. The single point that persists at infinity corresponds to the single connected component obtained when all the balls intersect in persistent homology. For $H_1(X)$, we similarly see most features die close to birth. However, there are two features that appear to persist for a significant time before ultimately dying during their merger. These correspond to loops found by chaining together simplices generated from datapoints to form a simplicial complex by the algorithm. In general, the Betti numbers are difficult to extract for complicated persistence diagrams, however we can make a reasonable guess that were this manifold to be taken in the continuum limit, we would get Betti numbers $b_0 = 1$ and $b_1 = 2$.

Table 1: Architecture of Neural Network

Name	Hidden Units	Activation Function
Input	4	ReLU
Layer 1	10	ReLU
Layer 2	30	ReLU
Layer 3	10	ReLU
Output	1	Sigmoid

Figure 2: First two clusters of the layer 1 representations of the ReLU network.

4 Experiment and Results

This section outlines the experiment conducted, the layer representations obtained, their persistence diagrams, and a PCA projection of them.

4.1 Experiment Design

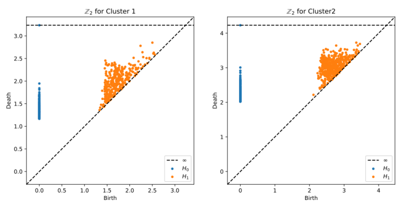
The experiment consisted of the modified torus parametrized in the previous section along with noise sampled from the uniform distribution. The total dataset length was 9800. The network, whose architecture is given in table 1, was trained on the binary classification task of distinguishing whether or not a point lay on the modified torus. The model was built using TensorFlow.

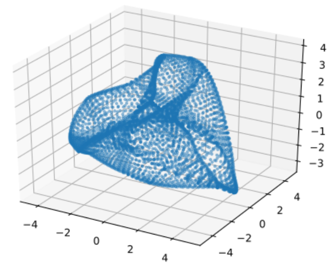
Using the Adam optimizer, the network was trained for 300 epochs when an accuracy of approximately 98% was obtained. Using the HDBSCAN clustering algorithm, the output of each layer was clustered and projected back onto the data space. Persistence diagrams on the clusters were calculated using Ripser. Lastly, principal component analysis (PCA) was applied on the layer representations corresponding to the torus to project them onto three dimensions.

The experiment was then repeated using the Tanh activation function on the network architecture described in table 1.

4.2 Results

The results from the experiments are summarized in figures 2 to 13. We notice that the clustering becomes progressively less defined throughout the layers as the network's representation becomes more consolidated. Notably, despite the visual resemblance to the data, persistent homology appeared unable to calculate topological features with the same degree of accuracy as on the raw data. The PCA projections obtained appear to resemble the original data less as the layers become deeper. It is noted that we were unable to compute the persistence diagrams of the clusterings of the third layer of the ReLU network due to the these clusterings being extremely noisy.

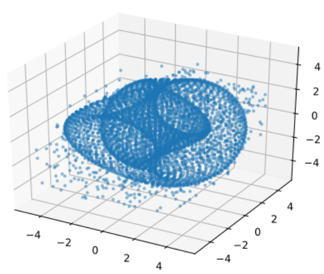




(b) PCA projection of torus.

(a) Persistence diagrams of clusters.

Figure 3: Persistence diagrams of layer 1 clusters for the ReLU network and PCA projection.



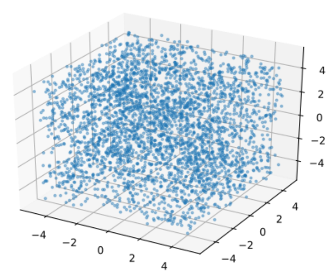
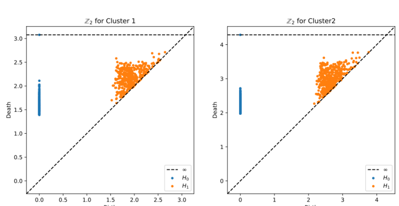
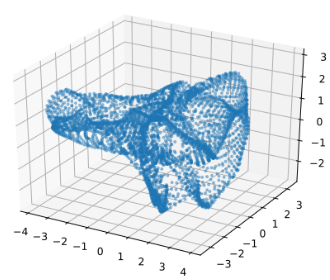


Figure 4: First two clusters of the layer 2 representations of the ReLU network.





(b) PCA projection of torus.

(a) Persistence diagrams of clusters.

Figure 5: Persistence diagrams of layer 2 clusters for the ReLU network and PCA projection.