Topological Data Analysis on Well-Trained, Deep Neural Networks for Binary Classification.
Mathematics Capstone
Winter 2022

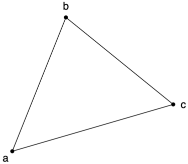
Adam Goeddeke

Table of Contents

Motivation

Example

How do we topologically describe an empty triangle?



$$H_p(K_t) \cong \begin{cases} \mathbb{Z}_2 & \text{if} \quad p = 0 \\ \mathbb{Z}_2 & \text{if} \quad p = 1 \\ \{0\} & \text{if} \quad p \geq 2 \end{cases}.$$

Where

$$K_t = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}\}.$$

Motivation, Continued(1)

Definition

Let K be a finite abstract simplicial complex. For integers $q \geq -1$, let $C_q(K) :=$ The free $\mathbb{Z}_2$-vector space on the set of q-simplicies of K. For integers $q \geq 0$, let

$$\partial_q : C_q(K) \to C_{q-1}(K)$$

be the $\mathbb{Z}_2$-linear map defined by linearly extending this assignment on the basis elements

$$\partial_q(\{v_0, ..., v_q\}) = \sum_{i=0}^{q} \{v_0, ..., \hat{v}_i, ..., v_q\}.$$

Motivation, Continued (2)

Definition

Where $\hat{v}_i$ means to leave out v_i. Then, consider the linear maps

$$C_{p+1}(K) \xrightarrow{\partial_{p+1}} C_p(K) \xrightarrow{\partial_p} C_{p-1}(K).$$

We say that the $\underline{p^{\text{th}} \text{ homology group of } K}$ is

$$H_p(K) := \frac{\ker(\partial_p)}{\operatorname{im}(\partial_{p+1})}.$$

Motivation, Continued (3)

Theorem

Let S^n denote the closed sphere in $\mathbb{R}^{n+1}$. That is,

$$S^n = \{(x_1, x_2, ..., x_{n+1}) \in \mathbb{R}^{n+1} \mid x_1^2 + x_2^2 + \cdots + x_{n+1}^2 = r\}$$

for some radius $r > 0$. Then,

$$H_p(S^n) \cong \begin{cases} \mathbb{Z}_2 & \text{if} \quad p = n \\ \{0\} & \text{if} \quad p \neq n \end{cases}.$$

Note: $H_p(S^n) = H_p(K_t)$ when $n = 1$..

Motivation, Continued (4)

Example

Let S^2 be the empty, closed sphere in $\mathbb{R}^3$. That is, for some radius $r > 0$,

$$S^2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = r\}.$$

Since $S^2 \cong K_T$, where K_T is the empty tetrahedron, we can see that

$$ker(\partial_2) = \mathbb{Z}_2$$

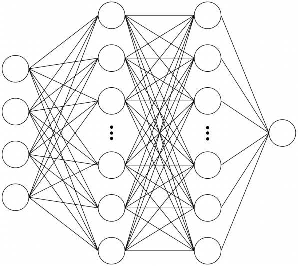
$$im(\partial_3) = \{0\}.$$

Thus, we have the quotient

$$H_2(S^2) = \frac{ker(\partial_2)}{im(\partial_3)} = \frac{\mathbb{Z}_2}{\{0\}} = \mathbb{Z}_2.$$

Its dimension, $dim(H_2(S^2)) = 1$, is the number of 3-dimensional voids in S^2.

How do Neural Networks *actually* work?



$$z_i^{[n]} = (\overrightarrow{w_i^{[n]}})^T \cdot \overrightarrow{a} + b_i^{[n]}, \quad \forall i \in \{1, ..., L_n\}.$$

$$\text{Sigmoid: } \sigma(z) = \frac{1}{1 + e^{-z}}$$

$$\text{ReLU: } R(z) = max(0, z)$$

$$\text{tanh: } tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

Why do we need TDA?

Questions:

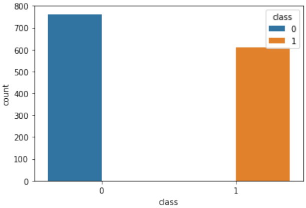
- How can we investigate what exactly is happening to the data inside a neural network?
- How can we quantify the efficacy of different activation functions?

Answer: We use persistent homology as shown by Gregory Naitzat, Andrey Zhitnikov, and Lek-Heng Lim in the twenty-first issue of the *Journal of Machine Learning Research* in their article "Topology of Deep Neural Networks."

This data is from the The University of California-Irvine's (UCI) online repository of data sets. It originally contains 1,372 rows of classified data. Each row represents one grey-scale image (with dimensions ranging from 96x96 pixels to 128x128 pixels) that has undergone wavelet transformation in order to better quantify the data.

Attributes	Data Type
Variance of Wavelet Transformed Image	float
Skewness of Wavelet Transformed Image	float
Curtosis of Wavelet Transformed Image	float
Entropy of Image	float
Class	int

Exploratory Data Analysis, Continued (1)



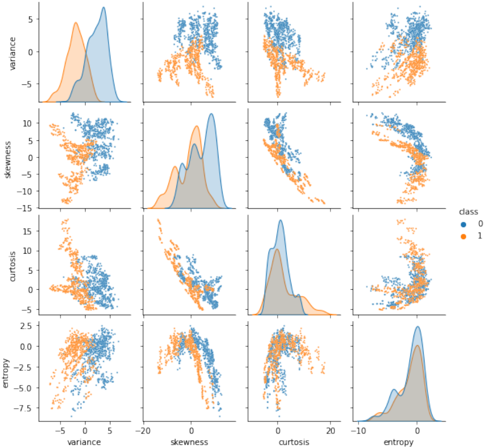


Figure: Count of Genuine vs Counterfeit Bank Notes (Left) and Pair Plot of Data (Right).

Exploratory Data Analysis Continued (2)

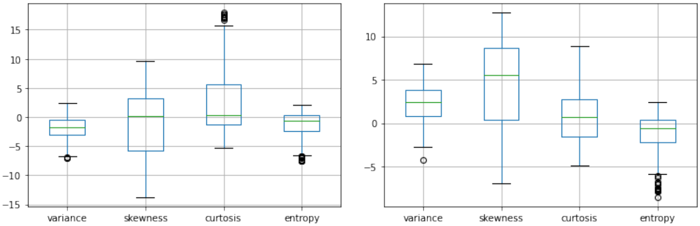
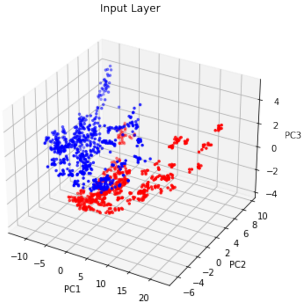
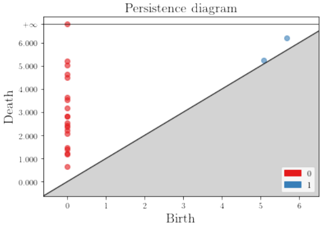
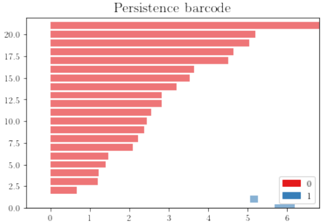


Figure: Box Plot of Genuine Bank Notes (Left) and Counterfeit Bank Notes (Right)

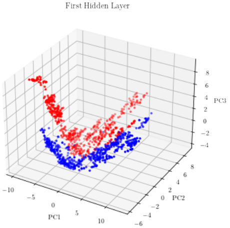
Persistence Diagrams/Barcodes - Input Layer

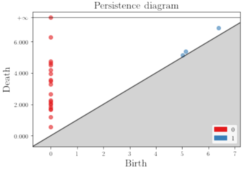


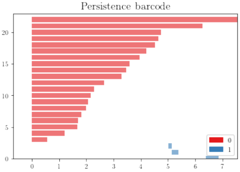


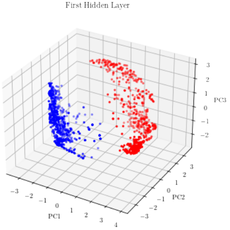


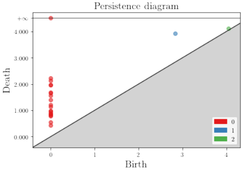
Persistence Diagrams/Barcodes - First Hidden Layer

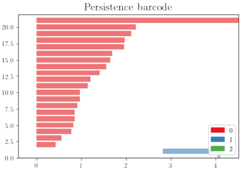




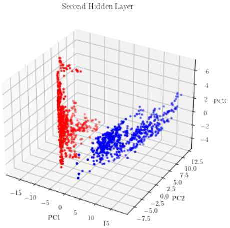


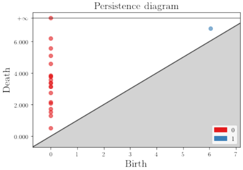


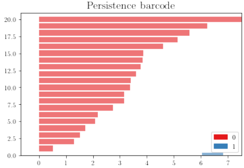


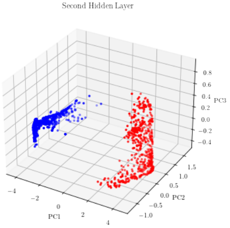


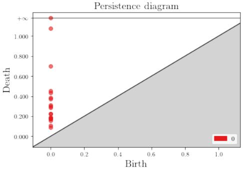
Persistence Diagrams/Barcodes - Second Hidden Layer

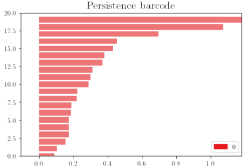




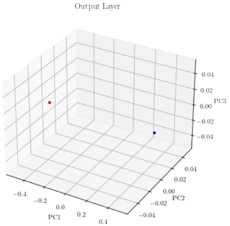


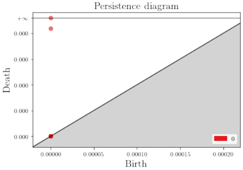


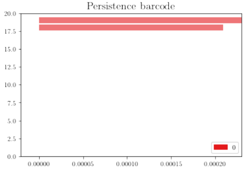


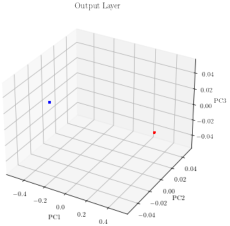


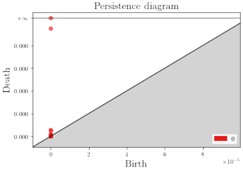
Persistence Diagrams/Barcodes - Output Layer

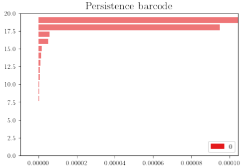












Conclusions

- ReLU and tanh are both able to accurately predict the class of each row in a trained network.
- Tanh was better at reducing the Betti numbers than ReLU in this case.
- After each layer in the neural network, we were able to use persistent homology to show the data becoming more clustered by their class.