Path homologies of deep feedforward networks

Samir Chowdhury[1], **Thomas Gebhart**[2], Steve Huntsman[3], Matvey Yutin[3,4]

1. *Stanford University*

2. *University of Minnesota*

3. *BAE Systems FAST Labs*

4. *University of California, Berkeley*

Importance of Parameter Topology

- Lottery ticket hypothesis [1]
  - Networks contain a sparse, functional subgraph.
- Weight agnostic neural networks [2]
  - Parameter topologies can be constructed to be robust to randomly-assigned weights.
- Random networks [3]
  - Random architectures solve computer vision tasks better than human-designed architectures.
- Neural architecture search [4]
  - Path information is beneficial when searching for optimal architectures.

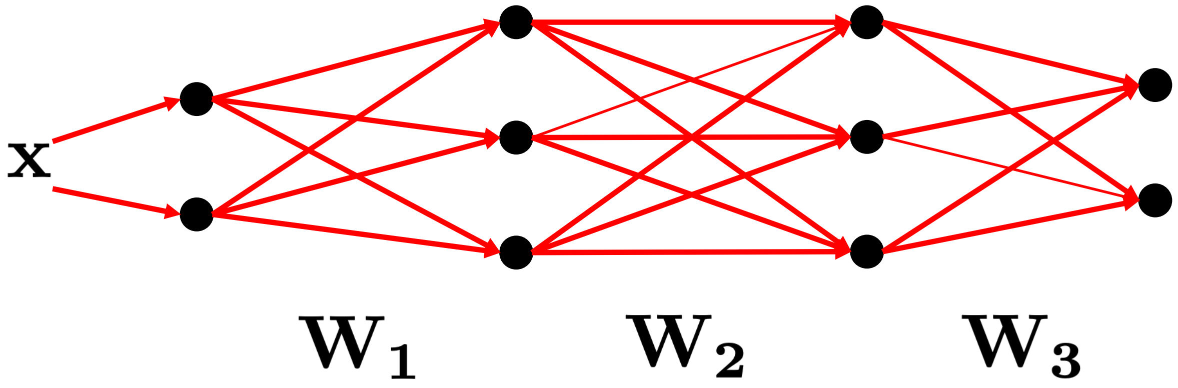
Topological priors matter to the performance of neural networks.

[1] Frankle, J., & Carbin, M. (2018). The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv preprint arXiv:1803.03635.*
[2] Gaier, A., & Ha, D. (2019). Weight Agnostic Neural Networks. *arXiv preprint arXiv:1906.04358.*
[3] Xie, S., Kirillov, A., Girshick, R., & He, K. (2019). Exploring randomly wired neural networks for image recognition. *arXiv preprint arXiv:1904.01569.*
[4]. White, C., Neiswanger, W., & Savani, Y. (2019). BANANAS: Bayesian Optimization with Neural Architectures for Neural Architecture Search. *arXiv preprint arXiv:1910.11858.*

Feedforward Networks as DAGs

- Feedforward neural networks are parameterized by a set of weight matrices $\{\mathbf{W}_i\}_{i=1}^{L}$.
- Ignoring nonlinearities, for a (linear) network f with $L = 3$:

$$f(\mathbf{x}) = \mathbf{W}_3 \mathbf{W}_2 \mathbf{W}_1 \mathbf{x}$$

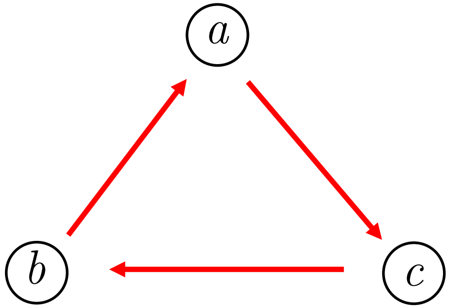


- Convolutional and pooling layers may be represented similarly.

Path Homology

Let $G = (X, E)$ be a digraph. For each $p \in \mathbb{Z}_+$ define $(x_0, \ldots, x_p)$ to be an *allowed p-path on X* if $(x_i, x_{i+1}) \in E$ for each $0 \leq i \leq p-1$.

Denote the free vector space on the collection of allowed p-paths by $\mathcal{A}_p = \mathcal{A}_p(G) = \mathcal{A}_p(X, E, \mathbb{K})$ where $\mathcal{A}_{-1} = \mathbb{K}$ and $\mathcal{A}_p = \{0\}$ for $p \leq 2$.



$$\mathcal{A}_0(G) = \mathbb{K}[\{a, b, c\}]$$
$$\mathcal{A}_1(G) = \mathbb{K}[\{ab, bc, ca\}]$$
$$\mathcal{A}_2(G) = \mathbb{K}[\{abc, bca, cab\}]$$
$$\mathcal{A}_2(G) = \mathbb{K}[\{abca, bcab, cabc\}$$

[6] Alexander Grigor'yan, Yong Lin, Yuri Muranov, and Shing-Tung Yau. Homologies of path complexes and digraphs. arXiv preprint arXiv:1207.2834, 2012.
[7] Chowdhury, S., & Mémoli, F. (2018). Persistent path homology of directed networks. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms* (pp. 1152-1169). Society for Industrial and Applied Mathematics.

Path Homology

The *space of ∂-invariant p-paths on G* is defined to be the following subspace of $\mathcal{A}_p(G)$:

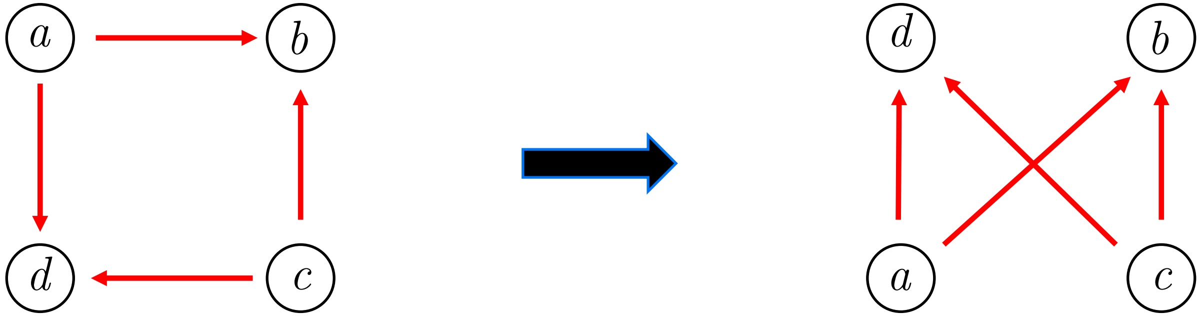
$$\Omega_p = \Omega_p(G) = \Omega_p(X, E, \mathbb{K}) = \{c \in \mathcal{A}_p \mid \partial_p(c) \in \mathcal{A}_{p-1}\}$$

which defines a chain complex $\ldots \xrightarrow{\partial_3} \Omega_2 \xrightarrow{\partial_2} \Omega_1 \xrightarrow{\partial_1} \Omega_0 \xrightarrow{\partial_0} \mathbb{K} \xrightarrow{\partial_{-1}} 0$

From this, the *p-dimensional (reduced) path homology group of* $G = (X, E)$ is defined as:

$$H_p^{\Xi}(G) = H_p^{\Xi}(X, E, \mathbb{K}) = \ker(\partial_p) / \operatorname{im}(\partial_{p+1})$$

Canonical Example



$$\Omega_0(G) = \mathbb{K}[\{a, b, c, d\}]$$
$$\Omega_1(G) = \mathbb{K}[\{ab, cb, cd, ad\}]$$
$$\Omega_2(G) = \{0\}$$

Note that $\partial_1^G(ab - cb + cd - ad) = b - a - b + c + d - c - d + a = 0$ so $\ker(\partial_1^G) \neq \{0\} = \mathrm{im}(\partial_2^G)$ implying that $\dim(H_1^\Xi(G)) = 1$.

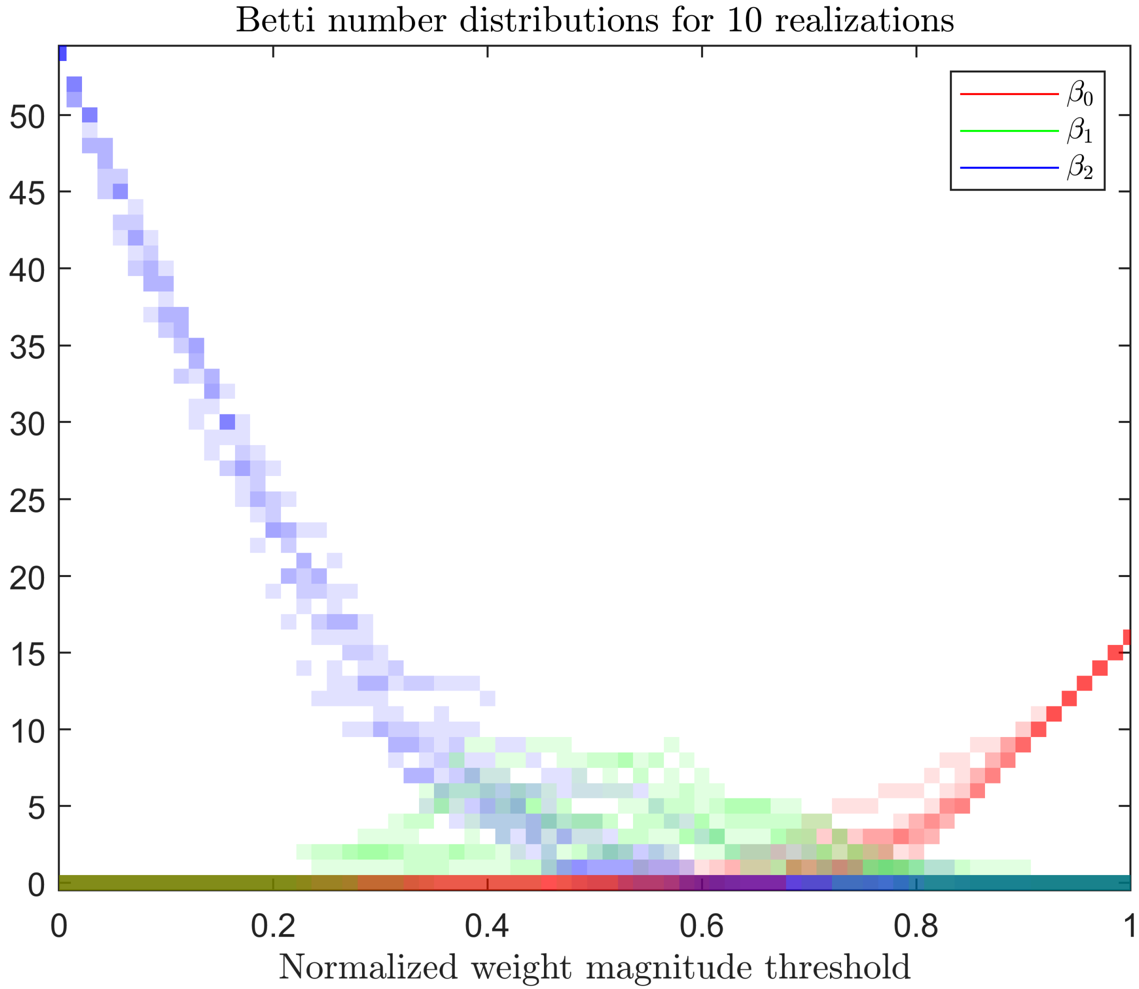
Main Result

Let $K^{\rightarrow}_{n_1,\ldots,n_L}$ denote the DAG corresponding to the architecture of a feedforward network with L layers with widths $\{n_1, n_2, \ldots, n_L\}$.

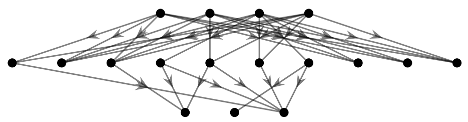
$K^{\rightarrow}_{n_1,\ldots,n_L}$ has nontrivial reduced path homology precisely in degree $(L-1)$:

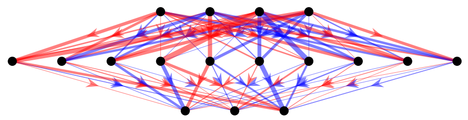
$$\dim\left(H_p^{\Xi}(K^{\rightarrow}_{n_1,\ldots,n_L})\right) = \delta_p^{L-1}\prod_{i=1}^{L}(n_i - 1)$$

Applications - Weight Thresholding



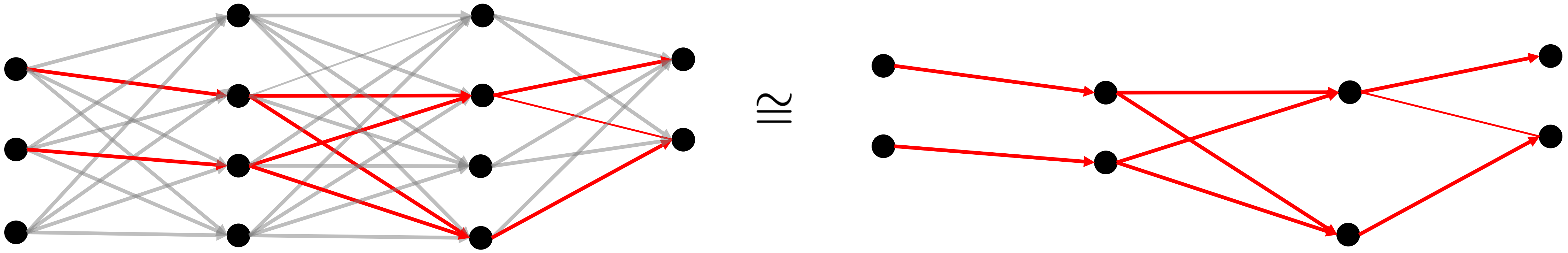
- Monotonically decreasing β_2
- Curious bump in β_1 in range 0.2-0.8
- Network disconnects around weight value 0.75





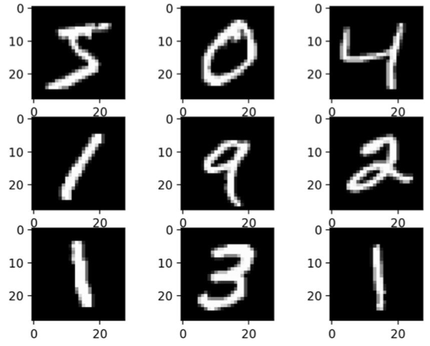
Applications - Lottery Ticket Topology

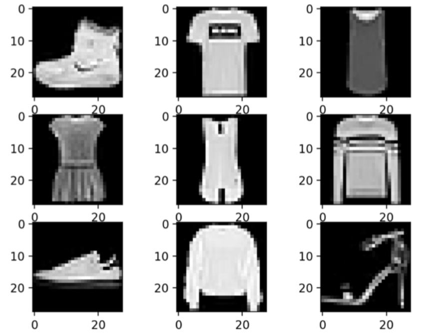
The Lottery Ticket Hypothesis: *Given a neural network with sufficient parameterization for a given task, there exists with high probability a subnetwork that, when trained starting with its original parameter initialization, achieves task performance reaching or exceeding that of the original network.*



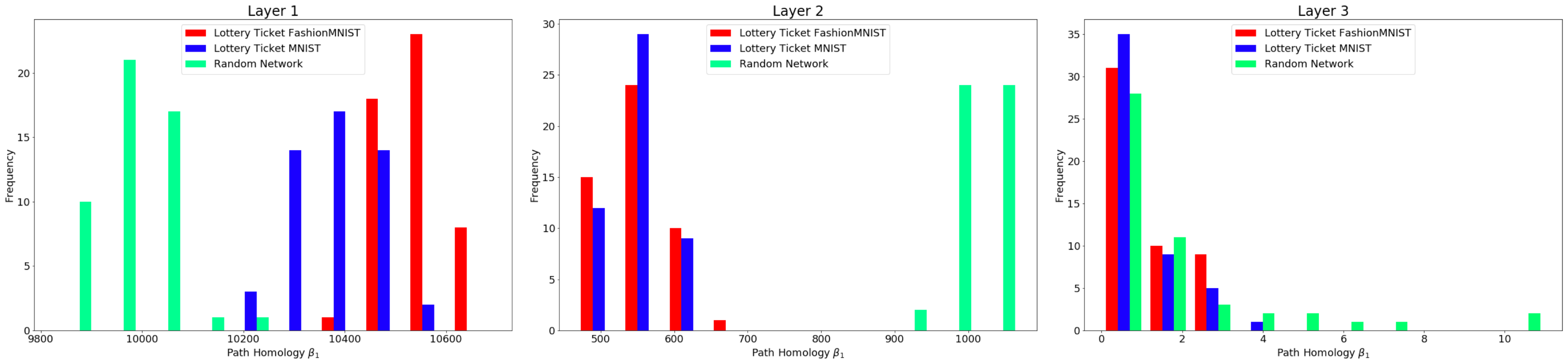
Applications - Lottery Ticket Topology

- Train a fully-connected ReLU network on MNIST and FashionMNIST.
- Network size $(784 \times 300) \rightarrow (300 \times 100) \rightarrow (100 \times 10)$
- Construct 50 lottery ticket networks for each dataset.
- Compute 1-dimensional path homology of each layer individually.
- Compare to network with same number of edges uniformly sampled from parent network.





Applications - Lottery Ticket Topology

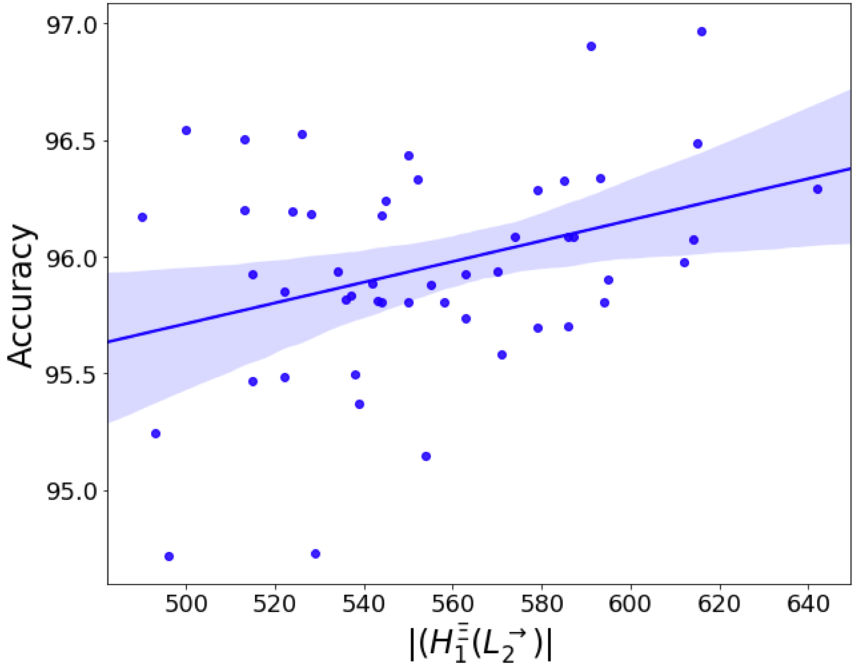


Expected MNIST:

Expected FashionMNIST:

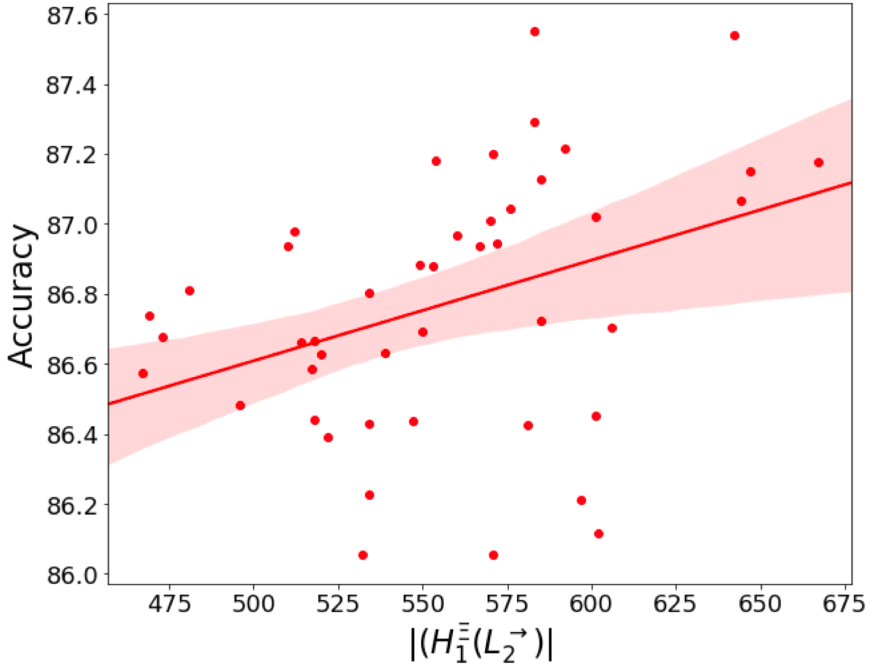
$$\text{Accuracy} = \alpha_0 + \alpha_1|(H_1^{\Xi}(L_1^{\rightarrow})| + \alpha_2|(H_1^{\Xi}(L_2^{\rightarrow})| + \alpha_3|(H_1^{\Xi}(L_3^{\rightarrow})|$$

MNIST



	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	83.6597	8.034	10.413	0.000	67.488	99.831
$H_1^{\Xi}(L_1^{\rightarrow})$	0.0009	0.001	1.190	0.240	-0.001	0.002
$H_1^{\Xi}(L_2^{\rightarrow})$	0.0051	0.002	2.402	0.020	0.001	0.009
$H_1^{\Xi}(L_3^{\rightarrow})$	0.0168	0.071	0.238	0.813	-0.126	0.159

Fashion MNIST



	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	81.9135	9.537	8.589	0.000	62.653	101.174
$H_1^{\Xi}(L_1^{\rightarrow})$	0.0003	0.001	0.338	0.737	-0.001	0.002
$H_1^{\Xi}(L_2^{\rightarrow})$	0.0030	0.001	2.459	0.018	0.001	0.006
$H_1^{\Xi}(L_3^{\rightarrow})$	0.0138	0.056	0.245	0.808	-0.100	0.128

References

1. Frankle, J., & Carbin, M. (2018). The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv preprint arXiv:1803.03635*.
2. Gaier, A., & Ha, D. (2019). Weight Agnostic Neural Networks. *arXiv preprint arXiv:1906.04358*.
3. Xie, S., Kirillov, A., Girshick, R., & He, K. (2019). Exploring randomly wired neural networks for image recognition. *arXiv preprint arXiv:1904.01569*.
4. White, C., Neiswanger, W., & Savani, Y. (2019). BANANAS: Bayesian Optimization with Neural Architectures for Neural Architecture Search. *arXiv preprint arXiv:1910.11858*.
5. Luetgehetmann, D., Govc, D., Smith, J., & Levi, R. (2019). Computing persistent homology of directed flag complexes. *arXiv preprint arXiv:1906.10458*.
6. Alexander Grigor'yan, Yong Lin, Yuri Muranov, and Shing-Tung Yau. Homologies of path complexes and digraphs. arXiv preprint arXiv:1207.2834, 2012.
7. Chowdhury, S., & Mémoli, F. (2018). Persistent path homology of directed networks. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms* (pp. 1152-1169). Society for Industrial and Applied Mathematics.