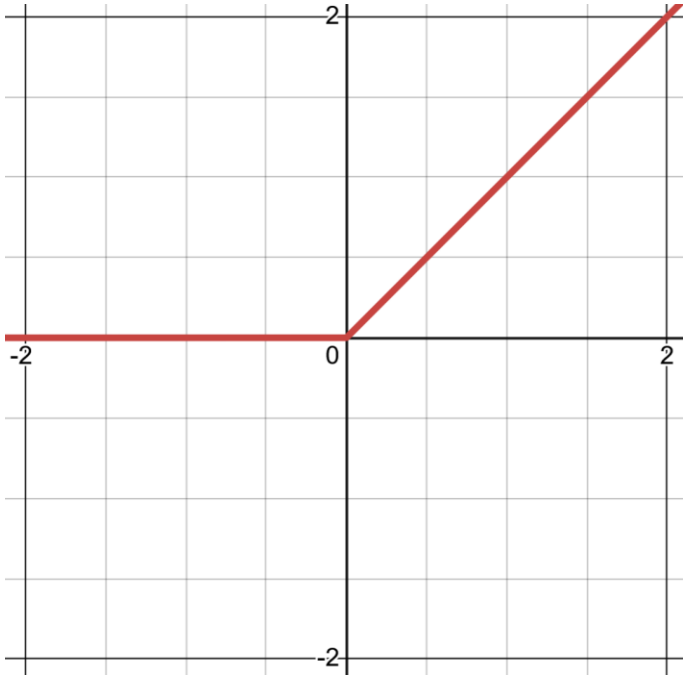
Characterizing the Discrete Geometry of ReLU Networks

Blake Gaines, Jinbo Bi

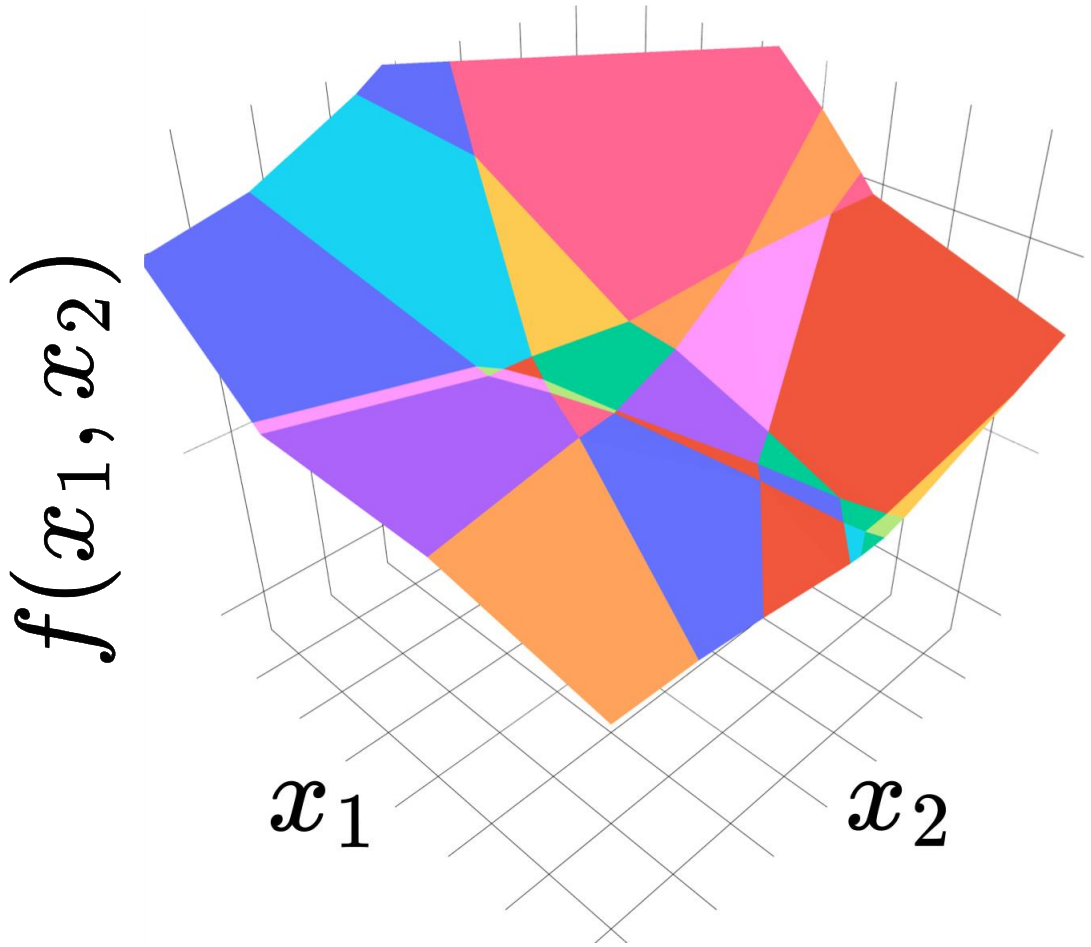
University of Connecticut, United States

$$\mathrm{ReLU}(x) = \quad = \max(x, 0)$$



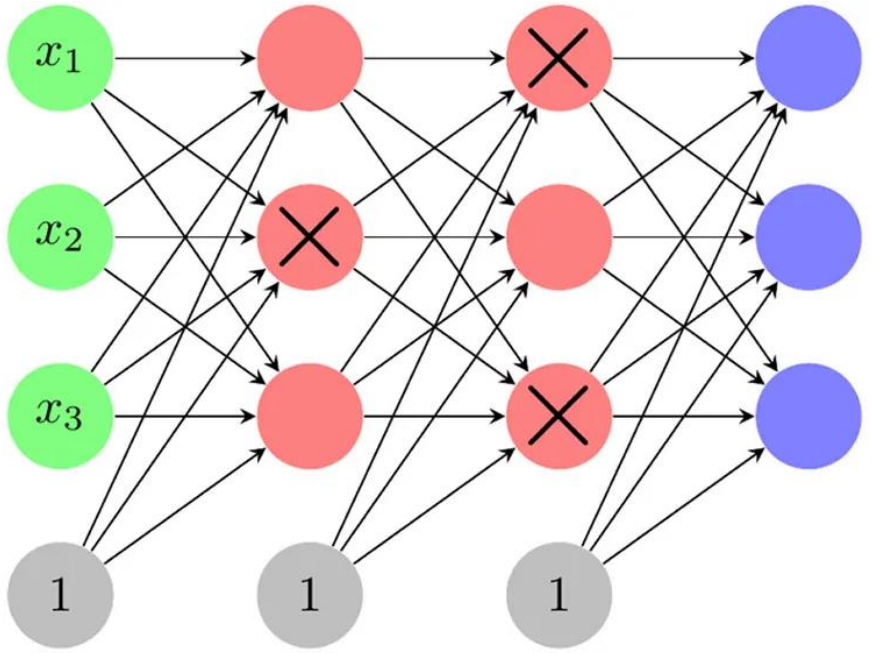
ReLU Networks are Continuous Piecewise Linear

- ReLU function is continuous piecewise-linear (CPWL)
- Affine functions are also CPWL
- Composition is CPWL
- → **ReLU networks are CPWL!**



$$f(x) = W_3\,\mathrm{ReLU}(W_2\,\mathrm{ReLU}(W_1 x + b_1) + b_2) + b_3$$

Some Terminology:



At a point in the input space:

- ReLU Neurons can be
  - “active” (non-negative output)
  - “inactive” (zero output)
- Neuron Boundary:
  - ∂{x: neuron is active} = ∂{x: neuron is inactive}
- Net’s Activation State: a vector of neuron activations
- **Activation Region → Affine Map**

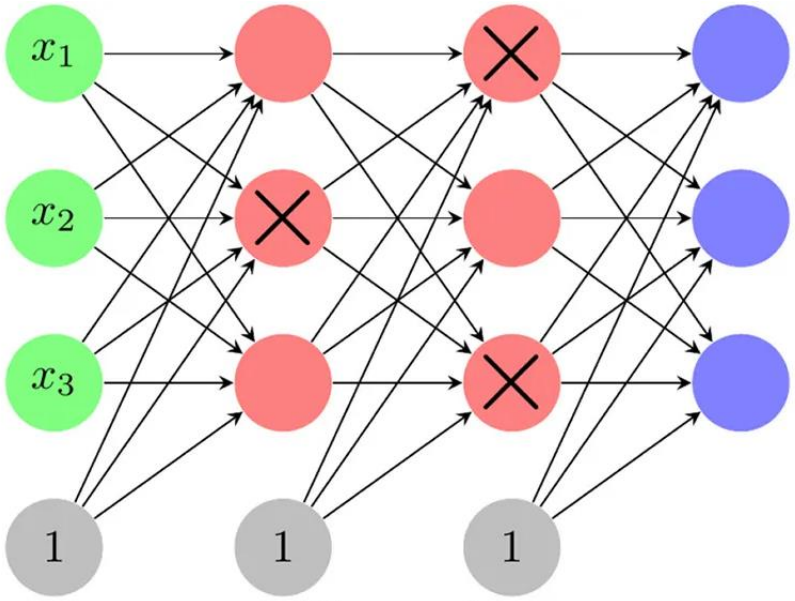
Network:

$$f(x_1,x_2,x_3) = \begin{bmatrix} 7 & 9 & 6 \\ 4 & 0 & 9 \\ 7 & 3 & 7 \end{bmatrix} \text{ReLU}\left(\begin{bmatrix} 2 & 9 & 5 \\ 1 & 3 & 3 \\ 8 & 7 & 0 \end{bmatrix} \text{ReLU}\left(\begin{bmatrix} 3 & 0 & 4 \\ 5 & 9 & 1 \\ 6 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 7 \\ 2 \\ 9 \end{bmatrix}\right) + \begin{bmatrix} 7 \\ 7 \\ 8 \end{bmatrix}\right) + \begin{bmatrix} 5 \\ 5 \\ 2 \end{bmatrix}$$

Figure: Sattelberg et al. (2023) Locally Linear Attributes of ReLU Neural Networks

Neuron Boundaries Divide Up the Linear Regions

Pick an Activation State:
(on, off, on, off, on, off)



Post ReLU:

$$f(x_1,x_2,x_3) = \begin{bmatrix}7&9&6\\4&0&9\\7&3&7\end{bmatrix}\left(\begin{bmatrix}2&9&5\\1&3&3\\8&7&0\end{bmatrix}\left(\begin{bmatrix}3&0&4\\5&9&1\\6&1&2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}+\begin{bmatrix}7\\2\\9\end{bmatrix}\right)+\begin{bmatrix}7\\7\\8\end{bmatrix}\right)+\begin{bmatrix}5\\5\\2\end{bmatrix}$$

$$= \begin{bmatrix}7&9&6\\4&0&9\\7&3&7\end{bmatrix}\left(\begin{bmatrix}0&0&0\\1&3&3\\0&0&0\end{bmatrix}\left(\begin{bmatrix}3&0&4\\0&0&0\\6&1&2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}+\begin{bmatrix}7\\0\\9\end{bmatrix}\right)+\begin{bmatrix}0\\7\\0\end{bmatrix}\right)+\begin{bmatrix}5\\5\\2\end{bmatrix}$$

$$= \begin{bmatrix}42&0&120\\20&0&27\\336&21&0\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}+\begin{bmatrix}695\\376\\562\end{bmatrix}$$

Affine Map

(Layer 1 Neurons) Region of validity: **(Layer 2 Neurons)**

$$\begin{bmatrix}3&0&4\\6&1&2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}+\begin{bmatrix}7\\9\end{bmatrix}\geq\begin{bmatrix}0\\0\end{bmatrix}$$

$$\begin{bmatrix}5&9&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}+2<0$$

- **System of linear inequalities**
- **Intersection of Half-Spaces**
- **A Polyhedron!**

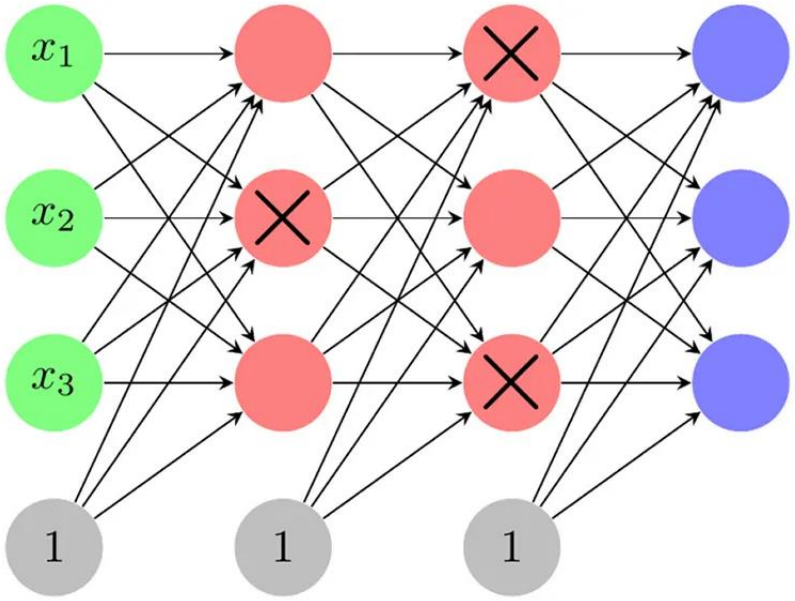
Network:

$$f(x_1,x_2,x_3) = \begin{bmatrix}7&9&6\\4&0&9\\7&3&7\end{bmatrix}\text{ReLU}\left(\begin{bmatrix}2&9&5\\1&3&3\\8&7&0\end{bmatrix}\text{ReLU}\left(\begin{bmatrix}3&0&4\\5&9&1\\6&1&2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}+\begin{bmatrix}7\\2\\9\end{bmatrix}\right)+\begin{bmatrix}7\\7\\8\end{bmatrix}\right)+\begin{bmatrix}5\\5\\2\end{bmatrix}$$

Figure: Sattelberg et al. (2023) Locally Linear Attributes of ReLU Neural Networks

Neuron Boundaries Divide Up the Linear Regions

Pick an Activation State:
(on, off, on, off, on, off)



Post ReLU:

$$f(x_1,x_2,x_3) = \begin{bmatrix}7&9&6\\4&0&9\\7&3&7\end{bmatrix}\left(\begin{bmatrix}2&9&5\\1&3&3\\8&7&0\end{bmatrix}\left(\begin{bmatrix}3&0&4\\5&9&1\\6&1&2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}+\begin{bmatrix}7\\2\\9\end{bmatrix}\right)+\begin{bmatrix}7\\7\\8\end{bmatrix}\right)+\begin{bmatrix}5\\5\\2\end{bmatrix}$$

$$= \begin{bmatrix}7&9&6\\4&0&9\\7&3&7\end{bmatrix}\left(\begin{bmatrix}0&0&0\\1&3&3\\0&0&0\end{bmatrix}\left(\begin{bmatrix}3&0&4\\0&0&0\\6&1&2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}+\begin{bmatrix}7\\0\\9\end{bmatrix}\right)+\begin{bmatrix}0\\7\\0\end{bmatrix}\right)+\begin{bmatrix}5\\5\\2\end{bmatrix}$$

$$= \begin{bmatrix}42&0&120\\20&0&27\\336&21&0\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}+\begin{bmatrix}695\\376\\562\end{bmatrix}$$

Affine Map

(Layer 1 Neurons)

$$\begin{bmatrix}3&0&4\\6&1&2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}+\begin{bmatrix}7\\9\end{bmatrix}\geq\begin{bmatrix}0\\0\end{bmatrix}$$

$$\begin{bmatrix}5&9&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}+2<0$$

Region of validity:

(Layer 2 Neurons)

$$\begin{bmatrix}1&3&3\end{bmatrix}\left(\begin{bmatrix}3&0&4\\5&9&1\\6&1&2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}+\begin{bmatrix}7\\2\\9\end{bmatrix}\right)+7\geq 0$$

$$\begin{bmatrix}2&9&5\\8&7&0\end{bmatrix}\left(\begin{bmatrix}3&0&4\\5&9&1\\6&1&2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}+\begin{bmatrix}7\\2\\9\end{bmatrix}\right)+\begin{bmatrix}7\\8\end{bmatrix}<\begin{bmatrix}0\\0\end{bmatrix}$$

If the activation of a neuron in an earlier layer changes, it will affect both the the affine map and **the boundary of neurons in later layers**

Network:

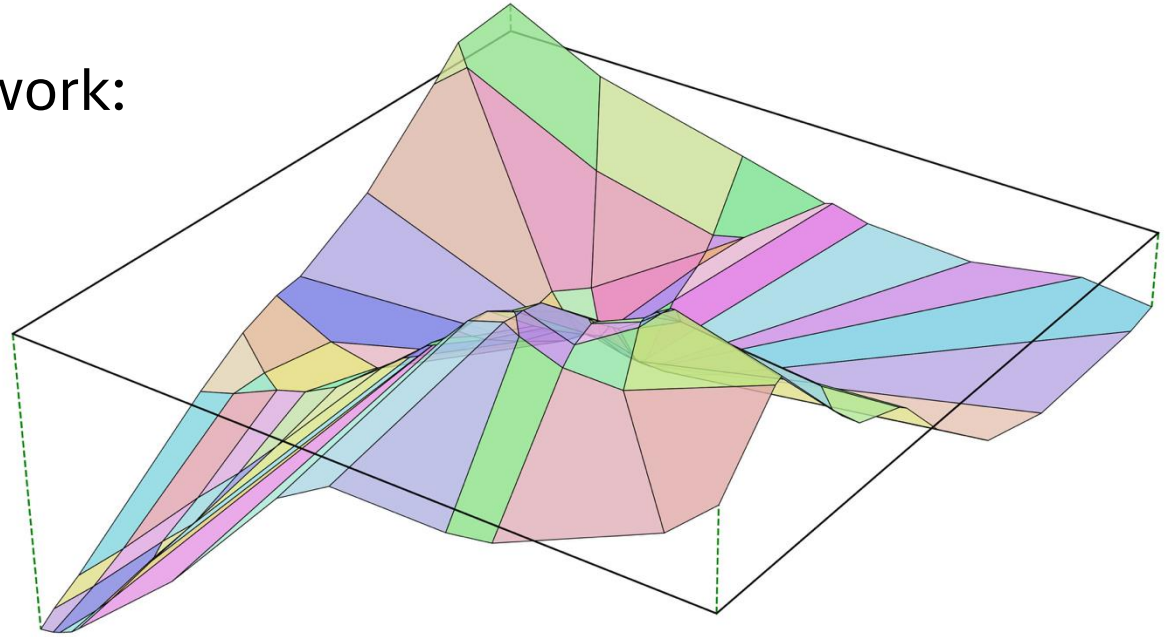
$$f(x_1,x_2,x_3) = \begin{bmatrix}7&9&6\\4&0&9\\7&3&7\end{bmatrix}\text{ReLU}\left(\begin{bmatrix}2&9&5\\1&3&3\\8&7&0\end{bmatrix}\text{ReLU}\left(\begin{bmatrix}3&0&4\\5&9&1\\6&1&2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}+\begin{bmatrix}7\\2\\9\end{bmatrix}\right)+\begin{bmatrix}7\\7\\8\end{bmatrix}\right)+\begin{bmatrix}5\\5\\2\end{bmatrix}$$

Figure: Sattelberg et al. (2023) Locally Linear Attributes of ReLU Neural Networks

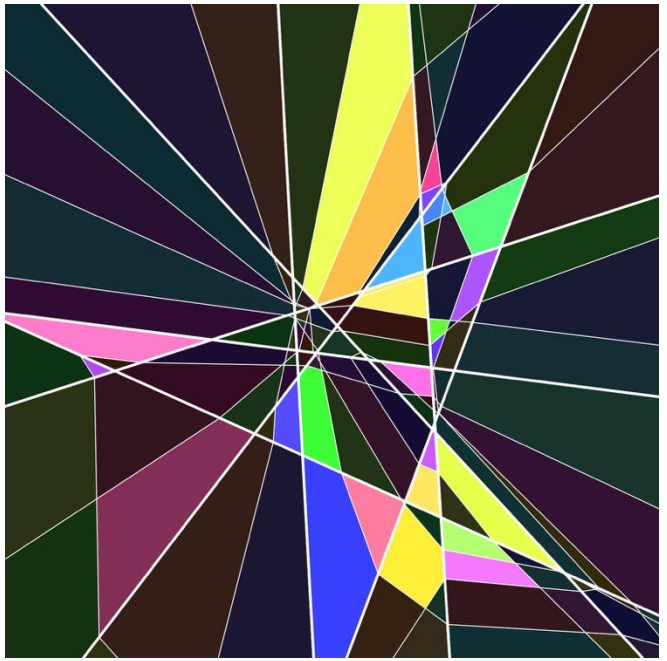
Our Goal

- The number and arrangement of activation regions defines the **expressivity** of these networks
- How do they fit together?

Network:



Activation Regions:

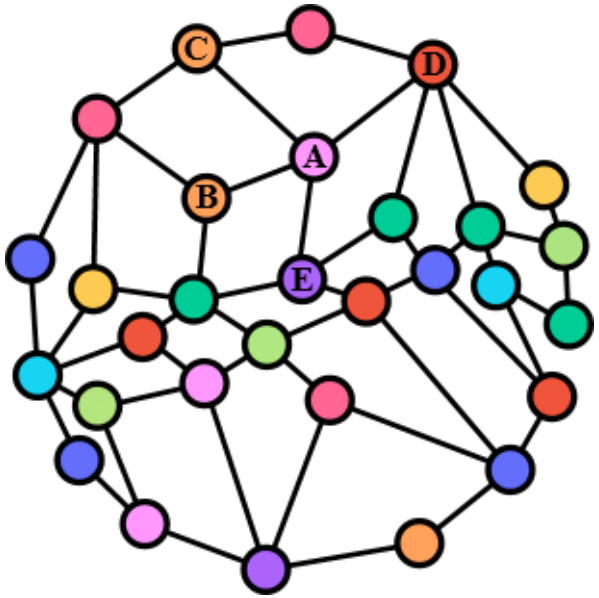


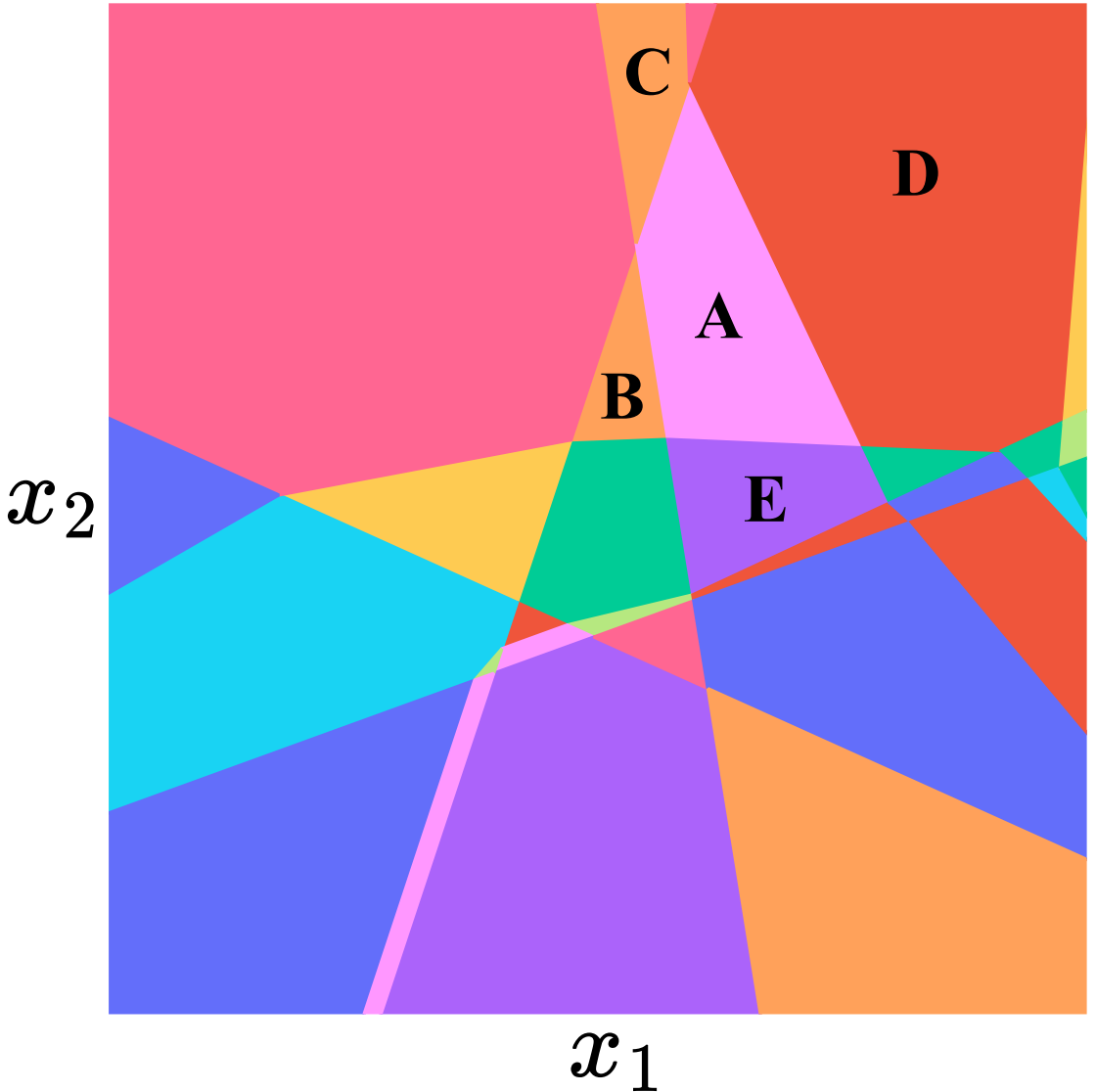
Figures: Ricson Cheng

Distance Between Activation Regions

Theorem: The maximum number of faces you would have to cross to get between a pair of activation regions is

O(width$^{\text{depth}}$)





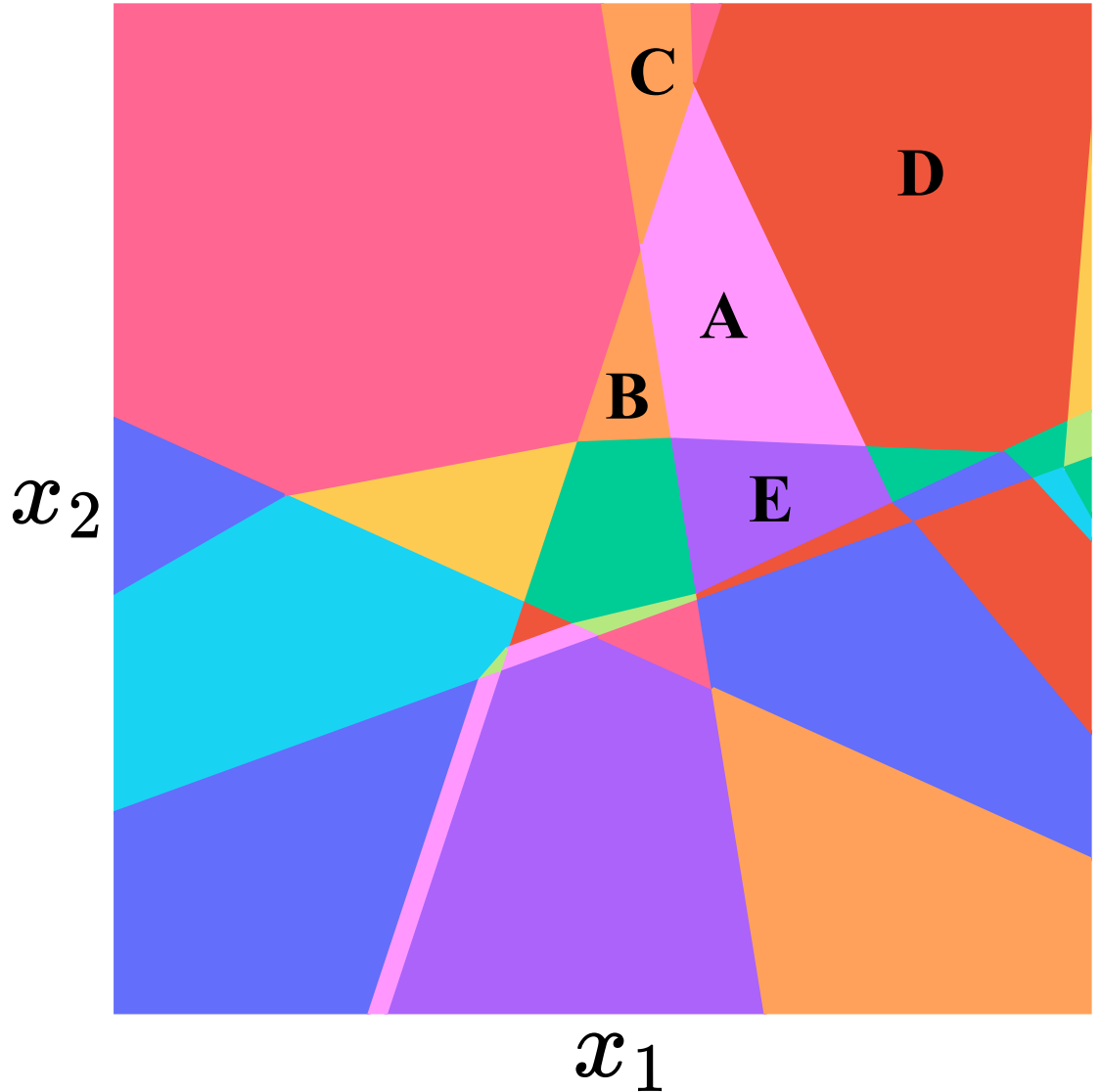
Distance Between Activation Regions

Theorem: The maximum number of faces you would have to cross to get between a pair of activation regions is

$$\mathbf{O(width^{depth})}$$

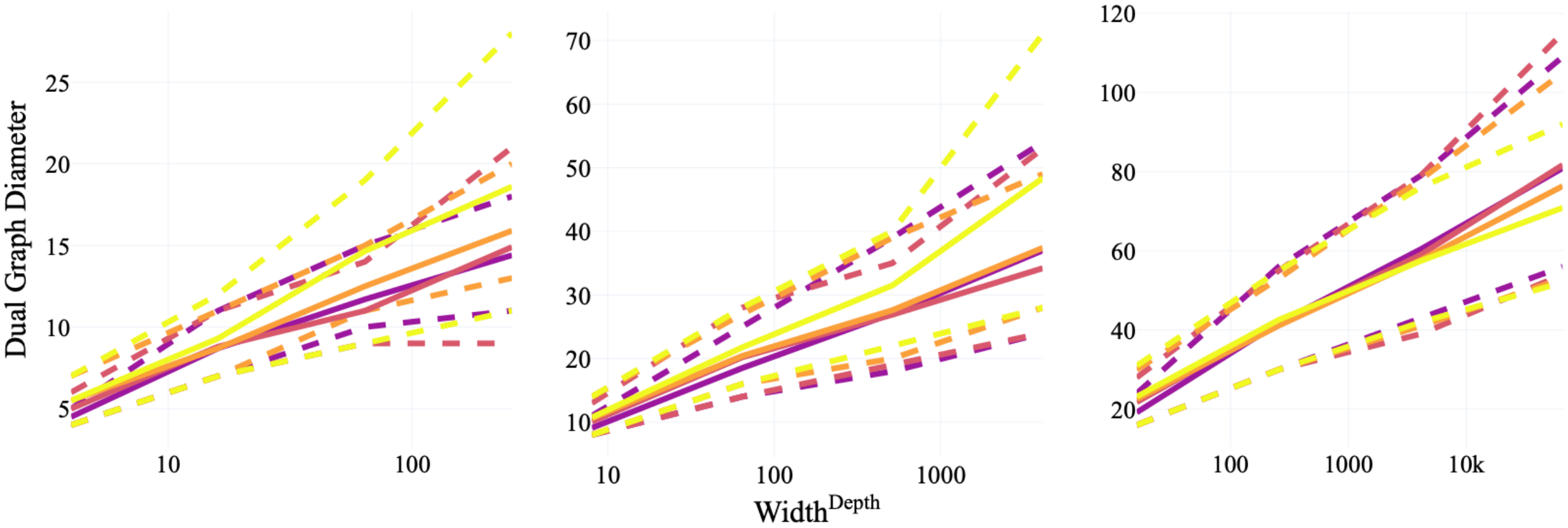
Reminder: The number of activation regions is

$$\mathbf{O(width^{depth*d})}$$



Diameter: Experimental Values vs. Upper Bound

(Different Colors are Dimensions 2,3,4,5)



Bounding the number of k-faces in arrangements of hyperplanes

Komei Fukuda

Graduate School of Systems Management, The University of Tsukuba, 3-29-1 Otsuka, Bunkyo-ku, Tokyo 112, Japan

Shigemasa Saito, Akihisa Tamura

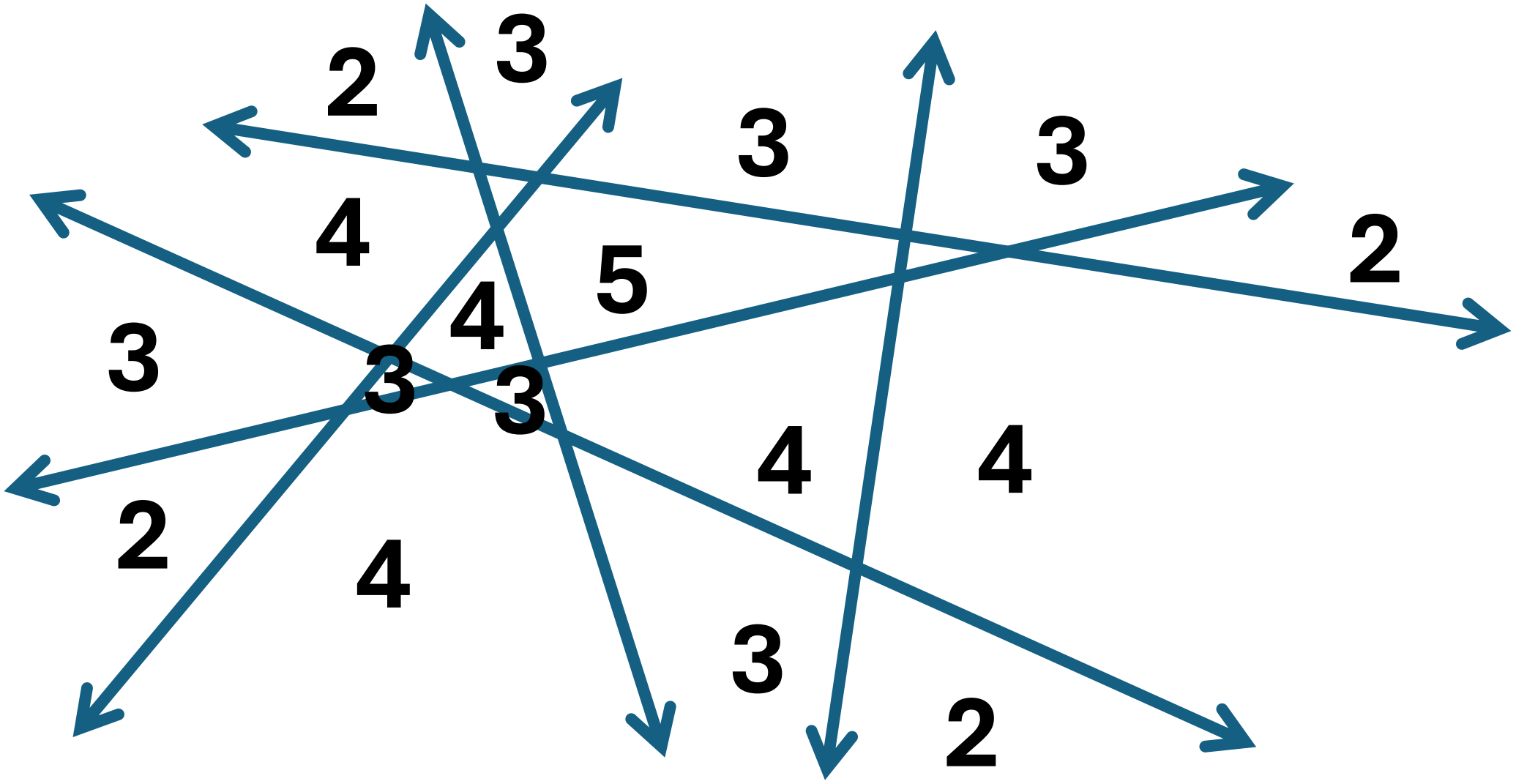
Department of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan

Takeshi Tokuyama

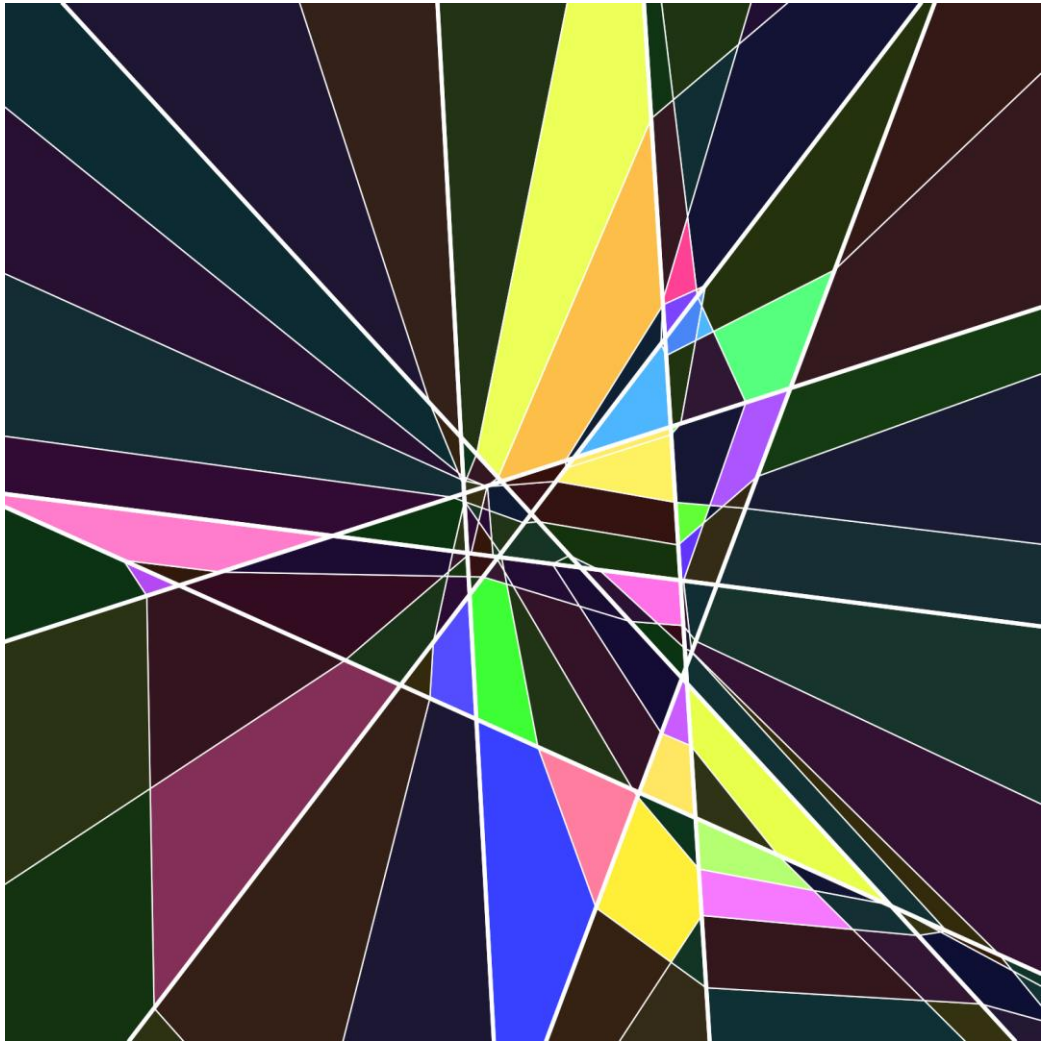
IBM Research, Tokyo Research Laboratory, 5-19 Sanbancho, Chiyoda-ku, Tokyo 102, Japan

Received 21 August 1989
Revised 30 April 1990

Theorem: The average number of faces of the regions of any hyperplane arrangement in d dimensions is at most $2d$



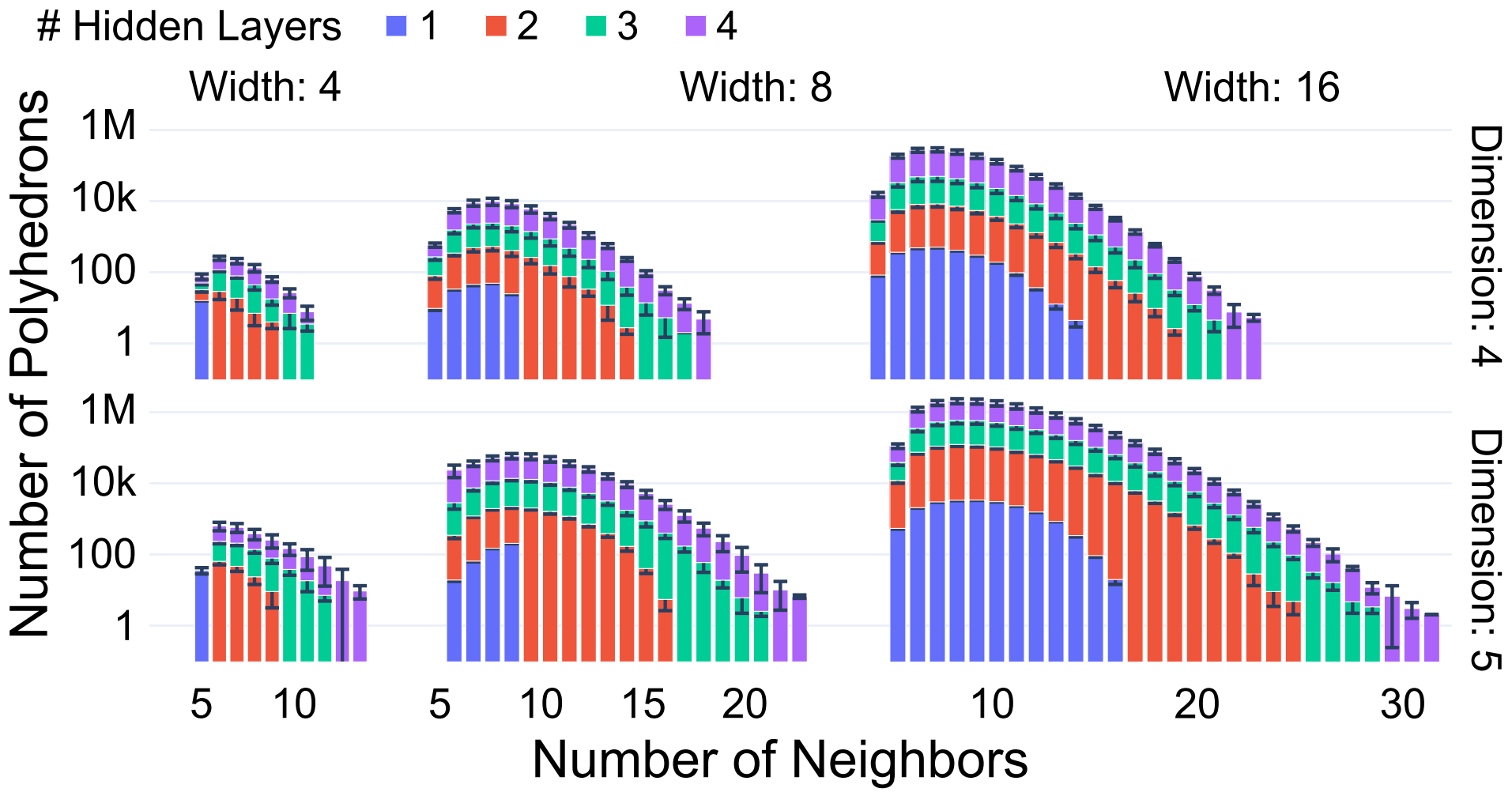
Theorem: The average number of faces of the regions of any hyperplane arrangement in *d* dimensions is at most *2d*



***regardless of network architecture**

Our Theorem: The average number of faces of the regions of **almost** any **bent** hyperplane arrangement in *d* dimensions is at most *2d*

Face Count Distributions of Maximal Regions

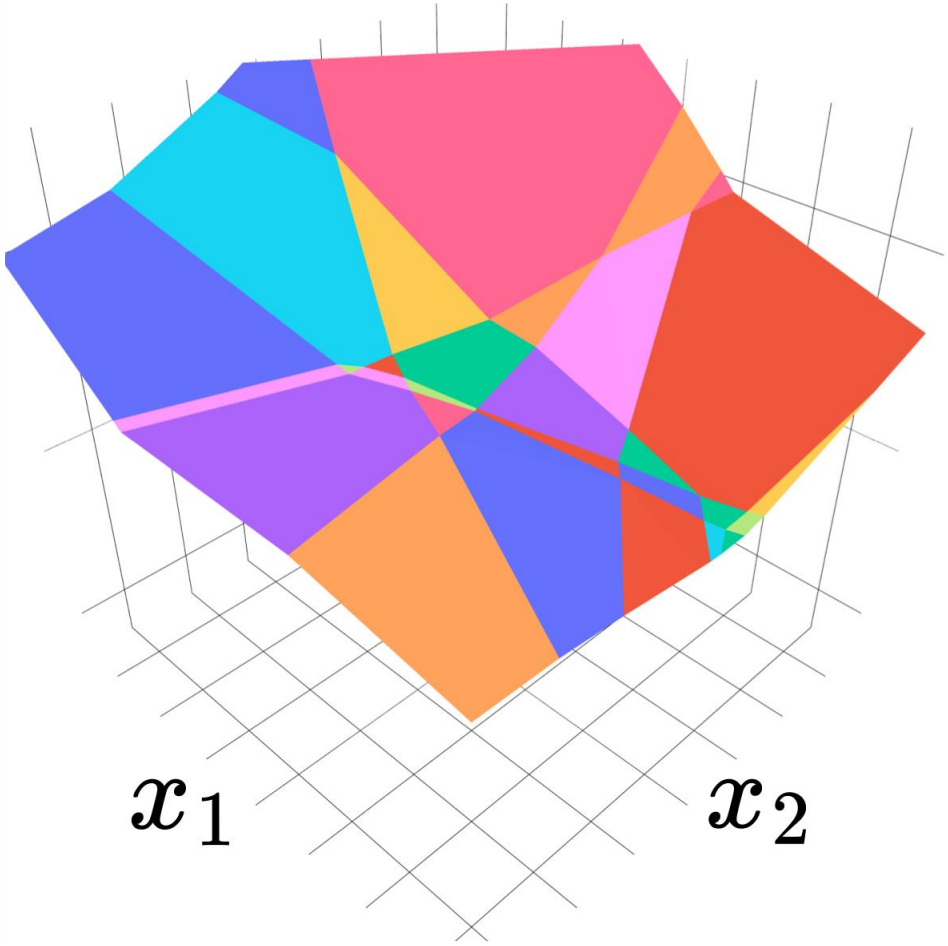


pip install relucent

General-purpose library for calculating the geometry and topology of ReLU networks

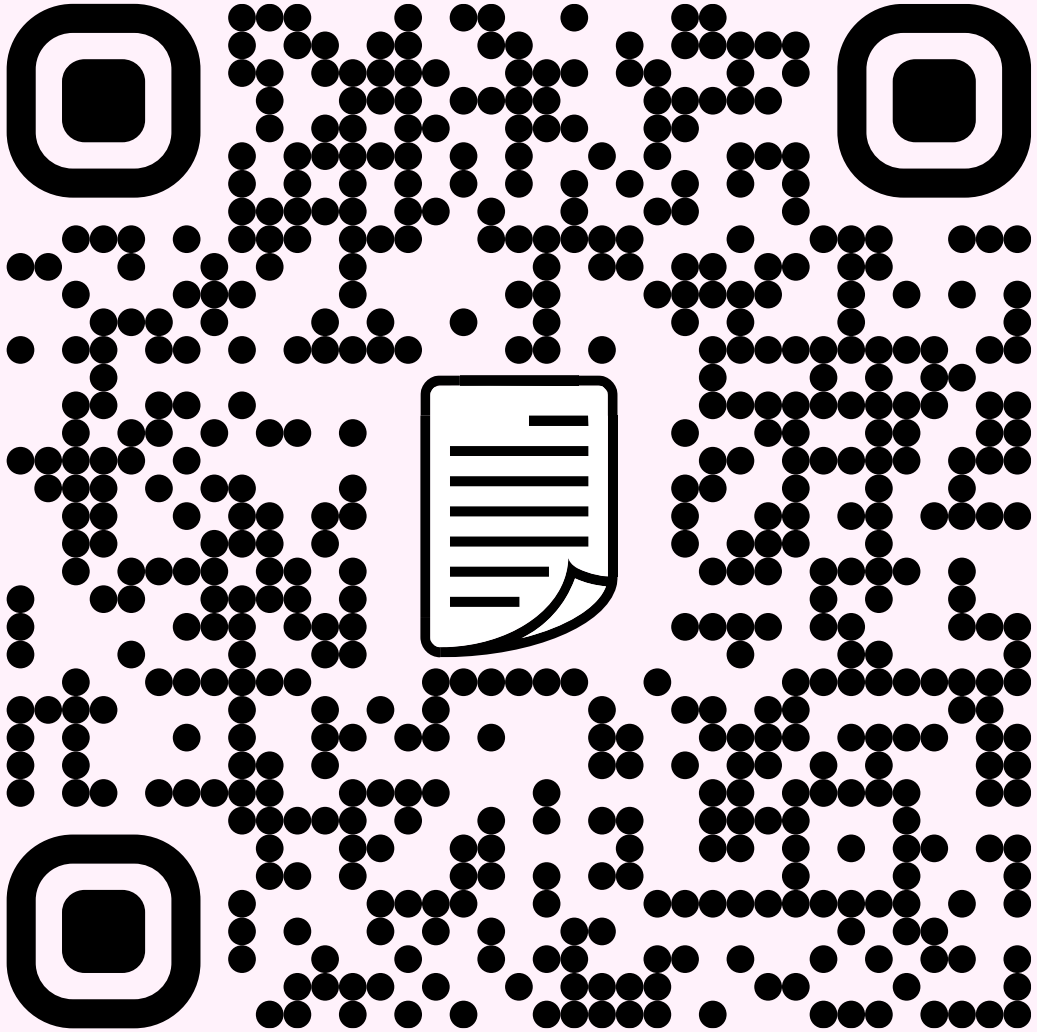
- **Open Source**
- **Available via pip**
- **Natively PyTorch Compatible**
- **Works in High Dimensions**
- **Cross-Platform Support**
- **Has Documentation**

3 lines to generate this figure from a PyTorch network →

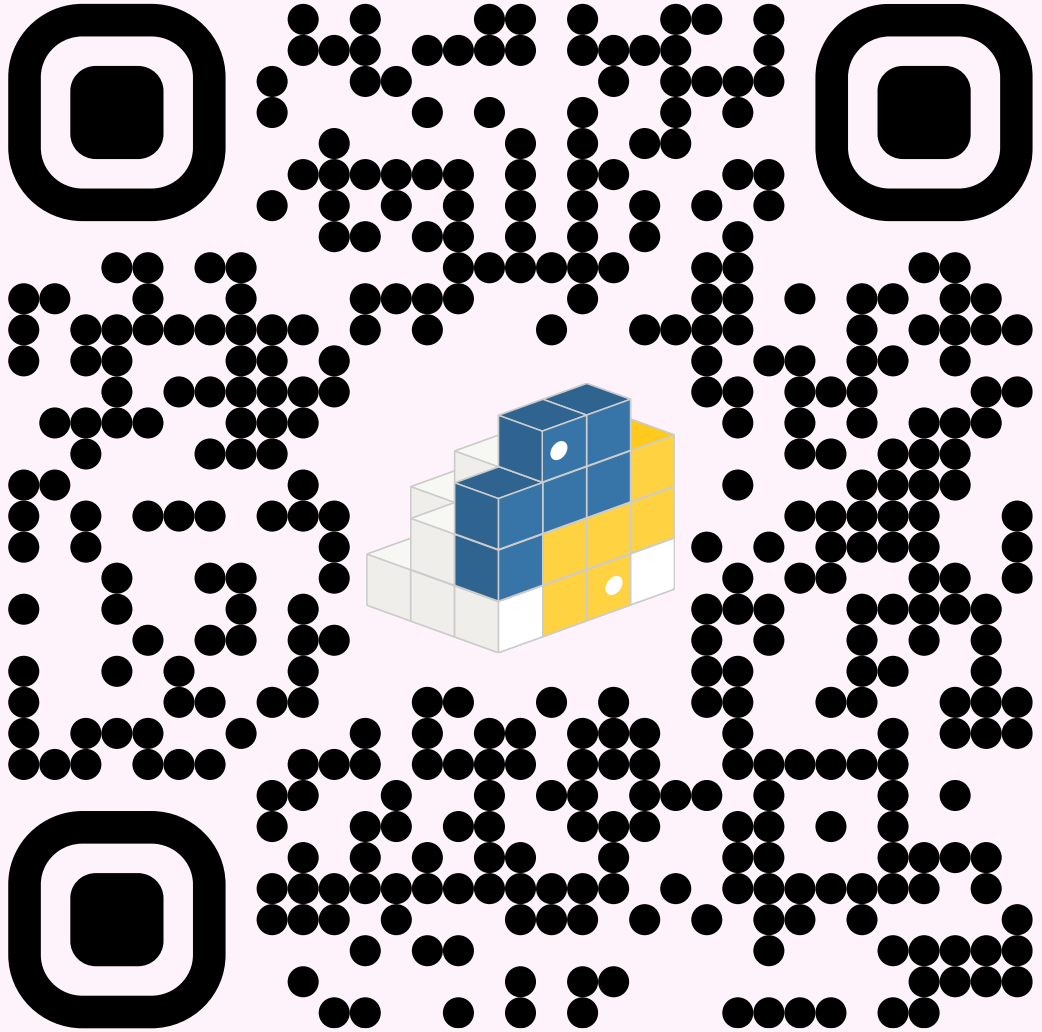


Interactive too!!

Thanks so much for listening!



Paper + Code



Relucent