Activation degree thresholds and expressiveness of polynomial neural networks

Bella Finkel (UW–Madison)
with Jose Israel Rodriguez, Chenxi Wu, Thomas Yahl

Spring Southeastern Sectional Meeting
Clemson University, March 8, 2025

arXiv:2408.04569

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 2137424.

Polynomial neural networks

Definition

A **polynomial neural network** $p_{\mathbf{w}} : \mathbb{R}^{d_0} \to \mathbb{R}^{d_L}$ with fixed activation degree r and architecture $\mathbf{d} = (d_0, d_1, \ldots, d_L)$ is a feedforward neural network

$$p_{\mathbf{w}}(x) = (W_L \circ \sigma_{L-1} \circ W_{L-1} \circ \sigma_{L-2} \circ \cdots \circ \sigma_1 \circ W_1)(x)$$

where the **weights** $W_i \in \mathbb{R}^{d_i \times d_{i-1}}$

and the **activation maps** $\sigma_i(x) : \mathbb{R}^{d_i} \to \mathbb{R}^{d_i}$ are given by coordinate-wise exponentiation to the r-th power,

$$\sigma_i(x) := (x_1^r, \ldots, x_{d_i}^r).$$

$$\mathbf{w} = (W_1, W_2, \ldots, W_L)$$

Neurovarieties

An L-layer polynomial neural network $p_{\mathbf{w}}$ with architecture $\mathbf{d} = (d_0, d_1, \dots, d_L)$ and activation degree r is represented by a tuple of homogeneous polynomials of degree r^{L-1}

$$\Psi_{\mathbf{d},r} : \mathbb{R}^{d_1 \times d_0} \times \cdots \times \mathbb{R}^{d_L \times d_{L-1}} \to (\mathrm{Sym}_{r^{L-1}}(\mathbb{R}^{d_0}))^{d_L} \subset \mathrm{Fun}(\mathbb{R}^{d_0}, \mathbb{R}^{d_L}),$$
$$\mathbf{w} \mapsto p_{\mathbf{w}} = \begin{pmatrix} p_{\mathbf{w}_1} \\ \vdots \\ p_{\mathbf{w}_{d_L}} \end{pmatrix}$$

The Zariski closure of the image of $\Psi_{\mathbf{d},r}$ in $\mathrm{Sym}_{r^{L-1}}(\mathbb{R}^{d_0}))^{d_L}$ is a **neurovariety** $\mathcal{V}_{\mathbf{d},r}$

Example

If $d_L = 1$, the codomain of the parameter map is the space of degree r^{L-1} homogeneous polynomials.

Why study polynomial neural networks?

Somewhat surprising slogan

Somebody cares!

- Recognizing tensor rank for quantum entanglement
  `arXiv:1908.10247`

And the characterization of neurovarieties speaks to broader questions. . .

`arXiv:2410.00722` On the Geometry and Optimization of Polynomial Convolutional Networks (Vahid Shahverdi, Giovanni Luca Marchetti, Kathlén Kohn)

`arXiv:2408.17221` Geometry of Lightning Self-Attention: Identifiability and Dimension (Nathan W. Henry, Giovanni Luca Marchetti, Kathlén Kohn)

Neurovarieties and their invariants

How do we characterize this function space?

One way: use the dimension of $\mathcal{V}_{\mathbf{d},r}$ to characterize the network's *expressiveness*.

Mulit-homogeneity (Kileel-Trager-Bruna)

For all invertible diagonal matrices $D_i \in \mathbb{R}^{d_i \times d_i}$ and permutation matrices $P_i \in \mathbb{R}^{d_i \times d_i}$, $\Psi_{\mathbf{d},r}$ gives the same neural network under the replacement

$$\begin{aligned} W_1 &\leftarrow P_1 D_1 W_1 \\ W_2 &\leftarrow P_2 D_2 W_2 D_1^{-r} P_1^T \\ &\vdots \\ W_L &\leftarrow W_L D_{L-1}^{-r} P_{L-1}^T. \end{aligned}$$

Thus the dimension of a generic pre-image of $\Psi_{\mathbf{d},r}$ is *at least* $\sum_{i=1}^{L-1} d_i$.

Neurovarieties and their invariants

The **expected dimension** of the neurovariety $\mathcal{V}_{\mathbf{d},r}$ is

$$\mathrm{edim}\mathcal{V}_{\mathbf{d},r} := \min\left\{ d_L + \sum_{i=0}^{L-1}(d_i d_{i+1} - d_{i+1}), \quad d_L \binom{d_0 + r^{L-1} - 1}{r^{L-1}} \right\}.$$

Example ($L = 1$)

In the case when $L = 1$, the neurovariety $\mathcal{V}_{\mathbf{d},r}$ has the expected dimension $d_0 d_1$, as the map $\Psi_{\mathbf{d},r}$ is an isomorphism.
The function space is parameterized by the $d_1 \times d_0$ matrix W_1.

Neurovarieties and their invariants

If $\dim \mathcal{V}_{\mathbf{d},r} \neq \mathrm{edim} \mathcal{V}_{\mathbf{d},r}$, then the neurovariety is said to be **defective**

$$\mathrm{defect}(\mathcal{V}_{\mathbf{d},r}) := \mathrm{edim}(\mathcal{V}_{\mathbf{d},r}) - \dim(\mathcal{V}_{\mathbf{d},r}).$$

Theorem (Alexander-Hirschowitz)

Let X be a general collection of d_1 double points in $\mathbb{P}^{d_0-1}$ and let $Y(d) \subseteq \mathrm{Sym}_r(\mathbb{C}^{d_0-1})$ be the subspace of polynomials through X. Then $Y(d)$ has the expected dimension $\min\{d_0 d_1, \binom{d_0+r-1}{r}\}$.
So, if $\mathbf{d} = (d_0, d_1, 1)$, then $\mathcal{V}_{\mathbf{d},r}$ is defective exactly when

$$\begin{aligned} r = 2, 2 \leq d_1 < d_0 \quad & r = 3, d_0 = 5, d_1 = 7, \\ r = 4, d_0 = 3, d_1 = 5 \quad & r = 4, d_0 = 4, d_1 = 9 \\ r = 4, d_0 = 5, d_1 = 14 \quad & \end{aligned}$$

Activation degree thresholds

Problem

Given an architecture $\mathbf{d}$, does there exist a non-negative integer $\tilde{r}$ such that the following holds:

$$\dim \mathcal{V}_{\mathbf{d},r} = \mathrm{edim} \mathcal{V}_{\mathbf{d},r} \quad \text{for all} \quad r > \tilde{r}\,?$$

We'd like to make a definition.

Definition

The activation degree threshold of an architecture $\mathbf{d} = (d_0, d_1, \ldots, d_L)$ with $L > 1$, $d_i > 1$ is

$$\mathrm{ActThr}(\mathbf{d}) := \min\{\tilde{r} \in \mathbb{N}_{\geq 0} : \dim \mathcal{V}_{\mathbf{d},r} = \mathrm{edim} \mathcal{V}_{\mathbf{d},r} \text{ for all } r > \tilde{r}\}.$$

Activation degree thresholds exist!

Theorem (Finkel-Rodriguez-Wu-Yahl)

For fixed $\mathbf{d} = (d_0, d_1, \ldots, d_L)$ satisfying $d_i > 1$ $(i = 0, \ldots, L-1)$, the activation degree threshold $\mathrm{ActThr}(\mathbf{d})$ exists.

In other words, there exists $\tilde{r}$ such that whenever $r > \tilde{r}$, the neurovariety $\mathcal{V}_{\mathbf{d},r}$ has the expected dimension,

$$\dim \mathcal{V}_{\mathbf{d},r} = \mathrm{edim} \mathcal{V}_{\mathbf{d},r} = d_L + \sum_{i=0}^{L-1} (d_i - 1) d_{i+1}.$$

Moreover, for $L > 1$, we have

$$\mathrm{ActThr}(\mathbf{d}) \leq 20m^2 - 32m + 10, \quad m = \max\{d_1, \ldots, d_{L-1}\}.$$

Powers of non-proportional multivariate polynomials

Proposition (Newman-Slater, Finkel-Rodriguez-Wu-Yahl)

Let

$$R_1^n(x) + R_2^n(x) + \cdots + R_k^n(x) = 1$$

where the R_i's are linearly independent non-constant rational functions in the variable x. Then $n \leq 5k^2 - 6k + 3$.

Proposition (Finkel-Rodriguez-Wu-Yahl)

Let $\mathbb{K}$ be a subfield of $\mathbb{C}$.
Given integers d, k, there exists an integer $\tilde{r} = \tilde{r}(k)$ with the following property.
If $r > \tilde{r}(k)$ and $p_1, \ldots, p_k \in \mathbb{K}[x_1, \ldots, x_d]$ are pairwise non-proportional, then $p_1^r, \ldots, p_k^r$ are linearly independent (over $\mathbb{K}$).
Moreover, $\tilde{r}(k) = 5k^2 - 16k + 10$ has the desired property.

Equi-width architectures

In special cases, we can do a lot better...

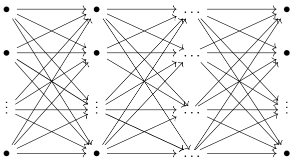
Let $\mathbf{d} = (d_0, d_1, \ldots, d_L)$ be an equi-width architecture, i.e., $d = d_0 = d_1 \cdots = d_L$, $L > 1$, and $d > 1$.

Theorem (Finkel-Rodriguez-Wu-Yahl)

The activation degree threshold associated to $\mathbf{d}$ is $\mathrm{ActThr}(\mathbf{d}) = 1$.

That is, if $\mathbf{d}$ is equi-width with $L > 1$ and $d > 1$, then for all $r > 1$ the neurovariety $\mathcal{V}_{\mathbf{d},r}$ has the expected dimension

$$\dim \mathcal{V}_{\mathbf{d},r} = Ld^2 - (L-1)d.$$



Better bounding activation degree thresholds

Except in a small number of cases, it is an open question to describe the fibers of the parameter map $\Psi_{\mathbf{d},r}$ when the neurovariety dimension does not equal the expected dimension.

This is a difficult but interesting problem! For example...

Patterns of linear dependence among polynomial powers

The **ticket** $T(F)$ of a finite set of polynomials $F = \{f_i\}$ is

$$T(F) = \{m \in \mathbb{N} | \{f_j^m\} \text{ is linearly dependent}\}$$

If

$$f_1 = x^2 + \sqrt{2}xy - y^2, \quad f_2 = ix^2 - \sqrt{2}xy + iy^2,$$
$$f_3 = -x^2 + \sqrt{2}xy + y^2, \quad f_4 = -ix^2 - \sqrt{2}xy - iy^2,$$

then $T(\{f_1, f_2, f_3, f_4\}) = \{2, 5\}$

Patterns among polynomial powers

Theorem (Reznick)

Let $F = \{f_1, \ldots, f_{d_L}\}$ be a set of pairwise non-proportional polynomials over $\mathbb{C}$. Then $|T(F)| \leq \binom{r-1}{2}$.

Conjecture (Reznick)

Every finite subset of $\mathbb{N}$ appears as a ticket for some family of homogeneous polynomials over $\mathbb{C}$.

Thank you!

AWM Research Symposium

⋆ Special Session on Advances in Applied Algebra and Algebraic Statistics ⋆
May 16-18, 2025 in Madison, WI

Joe Kileel, Matthew Trager, Joan Bruna `arXiv:1905.12207`

Kaie Kubjas, Jiayi Li, Maximilian Wiesmann `arXiv:2402.00949`

Giovanni Luca Marchetti, Vahid Shahverdi, Stefano Mereta, Matthew Trager, Kathlén Kohn `arXiv:2501.18915`