Endomorphisms of Integer Valued Neural Networks with $ReLU_t$

Eric Dolores-Cuenca

Industrial Mathematics Center
Pusan National University

Outline

Introduction to operad theory

Introduction to Computer Vision/ ML

New pooling filters

Joint work with

Aldo Guzman-Saenz, Sangil Kim, Susana Lopez-Moreno, Jose Mendoza-Cortes.

Posets

Denote by $< 3 >$ the space of three points $\{x, y, z\}$ with the Alexandrov topology:
$\tau = \{\{\}, \{x\}, \{x, y\}, \{x, y, z\}\}$.

Posets

Denote by $< 3 >$ the space of three points $\{x, y, z\}$ with the Alexandrov topology:
$\tau = \{\{\}, \{x\}, \{x, y\}, \{x, y, z\}\}$.
What are continuous functions from $< 3 >$ to $< 3 >$?

Posets

The identity map is continuous on $(\{x, y, z\}, \{\{\}, \{x\}, \{x, y\}, \{x, y, z\}\})$, but

$$x \to y, y \to z, z \to x$$

is not continuous, as the inverse image of $\{x\}$ is not open.

Posets

We define the order

$$x < y < z.$$

Then continuous functions on $(\{x, y, z\}, \{\{\}, \{x\}, \{x, y\}, \{x, y, z\}\})$ are order preserving functions: $a \leq b$ then $f(a) \leq f(b)$.

Posets

We work with finite partially ordered sets (posets). Denote by $< n >= 1 < 2 < \cdots < n$.
The order to run programs can be described with posets.

Hasse diagram

A Hasse diagram of a poset P, is a graph whose points are the points of the poset, with a vertical line between an element and it's successors.

$$\{w < x < y, x < z\} =$$

Lexicographic sum

Given $P = \{x_1, \cdots, x_n, \leq_P\}$ and n finite posets $P_1, \cdots, P_n$, we define the lexicographic sum $P(P_1, \cdots, P_n)$ [Tro02, Sch16, BHK18] as the set $\sqcup P_i$ with the order

$$a \leq_{P(P_1,\cdots,P_n)} b \text{ if } \begin{cases} a \leq_{P_i} b \text{ for some } i, \\ a \in P_r, b \in P_t, \text{ and } x_r \leq_P x_t. \end{cases}$$

Lexicographic sum

Given $P = \{x_1, \cdots, x_n, \leq_P\}$ and n finite posets $P_1, \cdots, P_n$, we define the lexicographic sum $P(P_1, \cdots, P_n)$ [Tro02, Sch16, BHK18] as the set $\sqcup P_i$ with the order

$$a \leq_{P(P_1,\cdots,P_n)} b \text{ if } \begin{cases} a \leq_{P_i} b \text{ for some } i, \\ a \in P_r, b \in P_t, \text{ and } x_r \leq_P x_t. \end{cases}$$

$$\{x < y\}(\{a,b\},\{c < d\}) =$$

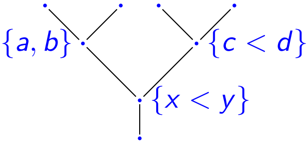
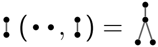


Figure: 

Operads

We represent addition using trees

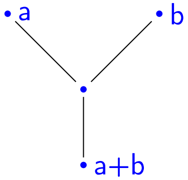


Figure: Representation of addition.

Operads

In this notation the notion of associativity is equivalent to request that the following two trees coincide for all possible inputs:

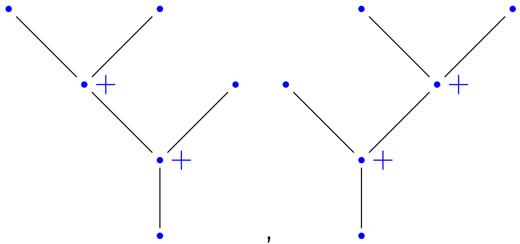


Figure: If we label the leaves with $2, 3, 7$ then we want $(2 + 3) + 7 = 12 = 2 + (3 + 7)$.

This way we reduce the "abstract study of operations" to study of diagrams of trees.

Operads

An important ingredient for this idea to work is that composition comes from grafting of trees.

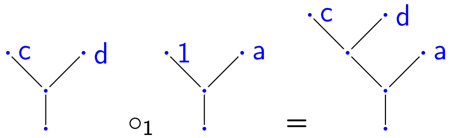


Figure: Grafting: We take the root of a tree and we identify it as the leave of another tree to form a bigger tree.

Operads

An operad consist of sets $\{O(n)\}_{n\in\mathbb{N}}$ (in the example, binary trees with n leaves/posets with n points), together with associative composition rules compatible with certain actions of S^n.

Boardman, Vogt [BV73], and May [May72], introduced the concept of operads in algebraic topology.
Kontsevich [Kon99] and Tamarkin [Tam03] popularized it in mathematical physics.
Giraudo [Gir16], Fauvet, Foissy, and Manchon [FFM18], studied the relationship between posets and operads.

Operads

For every set X we have the operad End_X with $End_X(n) = \{f : X^n \to X\}$

Finite poset operad

The operad FP has $FP(n)$ as the set of trees with n leaves, where a vertex with k children is labeled with posets with k points.

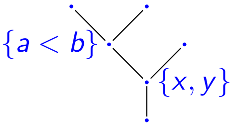


Figure: A ternary operation in $FP(3)$.

Algebra over an operad

A set Z is an algebra over an operad O if there are operadic maps

$$O \mapsto End_Z$$

(where $End_Z(n) = \{f : Z^n \to Z\}$).

Application to AI

Consider the operad of finite posets.

- We constructed a family of neural networks S (IVNN with $ReLU_t$) indexed by posets,
- and map $FP \to End_S$,

Application to AI

Consider the operad of finite posets.

- We constructed a family of neural networks S (IVNN with $ReLU_t$) indexed by posets,
- and map $FP \to End_S$,
- then, we proved that for posets with four points, the corresponding neural networks can be used as a pooling filter on any convolutional neural network (not only IVNN).

Images

In computer graphics, an image is represented by a matrix of dimensions width$\times$ height$\times$ 3.

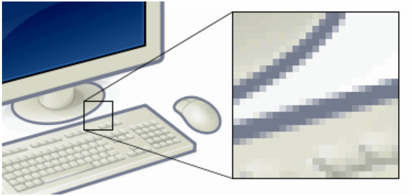


Figure: Images are matrices of colors. Image by By ed g2s

The last 3 coordinates are integers from 0 to 255, they stand for the amount of red, green, and blue.

A classifier algorithm

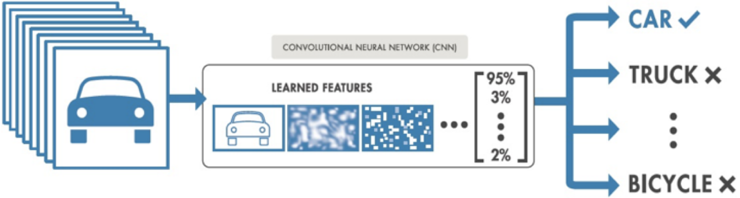


Figure: A classifier takes images and returns vectors of probabilities.

Images in this section are from "Neural Networks and Deep Learning" by Michael Nielsen and Adit Deshpande's blog.

Inside the black box

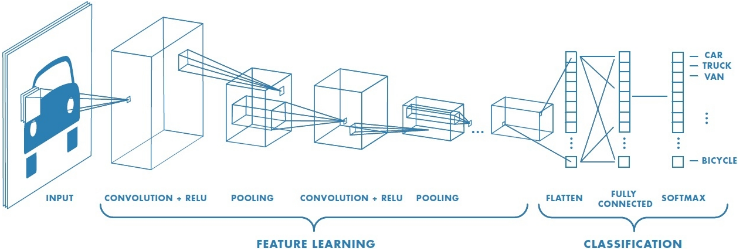


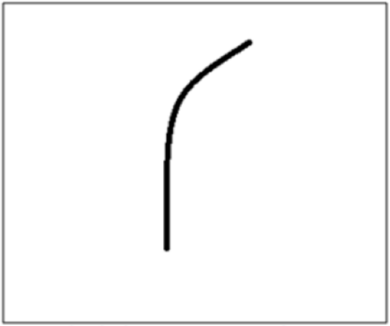
Figure: Architecture of a convolutional neural network.

`https://adamharley.com/nn_vis/cnn/2d.html` made by Adam Harley.

Convolutional filters

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

Pixel representation of filter



Visualization of a curve detector filter

Convolutional filters

Original image

Visualization of the filter on the image

Visualization of the receptive field

0	0	0	0	0	0	30
0	0	0	0	50	50	50
0	0	0	20	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0

Pixel representation of the receptive field

*

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

Pixel representation of filter

Multiplication and Summation = (50*30)+(50*30)+(50*30)+(20*30)+(50*30) = (**6600** r!)

Convolutional filters

Visualization of the filter on the image

0	0	0	0	0	0	0
0	40	0	0	0	0	0
40	0	40	0	0	0	0
40	20	0	0	0	0	0
0	50	0	0	0	0	0
0	0	50	0	0	0	0
25	25	0	50	0	0	0

Pixel representation of receptive field

*

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

Pixel representation of filter

Multiplication and Summation = 0

Pooling layers

1	2	2	0
1	2	3	2
3	1	3	2
0	2	0	2

Max pooling

2	3
3	3

*How about average pooling?

New pooling filters

Let P be a poset with 4 points. The order polytope [Sta86]:

$$Poly : Poset \to Polytopes$$

$Poly(P) \subset [0,1]^4$ is the subspace of points satifying the inequalities of the poset. (Every point of P is assigned to a basis element).

New pooling filters

Let P be a poset with 4 points. The order polytope [Sta86]:

$$Poly : Poset \to Polytopes$$

$Poly(P) \subset [0,1]^4$ is the subspace of points satifying the inequalities of the poset. (Every point of P is assigned to a basis element). For example,

$$\begin{aligned} Poly(\{w,x,y,z|x<y\}) &= \{(w,x,y,z) \in [0,1]^4 \text{ such that} \\ &\quad 0 \le x \le y \le 1, \\ &\quad 0 \le w \le 1, \\ &\quad 0 \le z \le 1\} \end{aligned}$$

New pooling filters

Then for every P poset with four points, $P \to Poly(P) \to Filter$. If $A = \begin{matrix} a_{0,0} & a_{0,1} \\ a_{1,0} & a_{1,1} \end{matrix}$, $f(A) = (a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1})$, then the filter is

$$\max_{\{v_j\}} \{f(A) \cdot v_j\},$$

where max is indexed by the vertices of $Poly(P)$.

Explicitly.

The order polytope of • • • • $= [0,1]^4$.

The order polytope of  $= \Delta[4]$.

The order polytope of  is union of 5 copies of $\Delta[4]$, glued along their boundaries using 5 $\Delta[3]$, with four $\Delta[3]$ sharing a $\Delta[2]$ face.

Explicitly.

The formula for filters $\max_{\{v_j \in Poly(P)\}}\{f(A)\cdot v_j\}$,

$$\begin{bmatrix} a_{0,0} & a_{0,1} \\ a_{1,0} & a_{1,1} \end{bmatrix} = \max\left\{ \begin{array}{c} 0,\ a_{1,1},\ a_{1,0},\ a_{1,0}+a_{1,1},\ a_{0,0}+a_{1,0},\ a_{1,0}+a_{0,1}+a_{1,1}, \\ a_{0,0}+a_{1,0}+a_{1,1},\ a_{0,0}+a_{0,1}+a_{1,0}+a_{1,1} \end{array} \right\}$$

$$\begin{bmatrix} a_{0,0} & a_{0,1} \\ a_{1,0} & a_{1,1} \end{bmatrix} = \max\left\{ \begin{array}{l} 0, \\ \max\limits_{i,j}\{a_{i,j}\}, \\ \max\limits_{\substack{i,j,k,l \\ (i,j)\neq(k,l)}}\{a_{i,j}+a_{k,l}\}, \\ \max\limits_{\substack{i,j,k,l,m,n \\ (i,j)\neq(k,l) \\ (i,j)\neq(m,n) \\ (k,l)\neq(m,n)}}\{a_{i,j}+a_{k,l}+a_{m,n}\}, \\ a_{0,0}+a_{1,0}+a_{0,1}+a_{1,1} \end{array} \right\}.$$

$$\begin{bmatrix} a_{0,0} & a_{0,1} \\ a_{1,0} & a_{1,1} \end{bmatrix} = \max\{0,\ a_{1,1},\ a_{1,1}+a_{1,0},\ a_{1,1}+a_{1,0}+a_{0,1},\ a_{1,1}+a_{1,0}+a_{0,1}+a_{0,0}\}$$

Repository:

`https://github.com/mendozacortesgroup/Poset-filters`

Performance

Model QCNN	Training runs	# params	Time	Test acc	Std
Original	1	2,032,650	–	77.78%	–
Weeda impl. wihtout seeds	3	2,032,650	39' 34"	78.14%	–
Weeda impl.	14	2,032,650	25' 40"	77.83%	0.730
Weeda impl. + 	14	656,394	34' 20"	**78.92%**	0.514
Weeda impl. + max pooling (instead of )	14	656,394	27' 15"	76.75%	0.397
Weeda impl. + avg. pooling (instead of )	14	656,394	28' 15"	76.097%	1.104
Weeda impl. + mixed pooling (instead of )	14	656,394	27' 10"	76.41%	0.670

Table 5: Experiment results on CIFAR10 classification task for quaternion convolutional neural networks with 80 epochs

What if we use random vectors?

Model QCNN+	Training runs	# params	Time	Test acc	Std
rndm vt # 1	10	656,394	50' 25"	78.11%	0.758
rndm vt # 2	10	656,394	53' 35"	78.05%	0.736
rndm vt # 3	10	656,394	47' 15"	78.03%	0.805
rndm vt # 4	10	656,394	42' 10"	78.09%	0.705
rndm vt # 5	10	656,394	39' 55"	77.91%	0.642
rndm vt # 6	10	656,394	38' 20"	78.22%	0.629
rndm vt # 7	10	656,394	36' 15"	78.16%	0.569
rndm vt # 8	10	656,394	35' 25"	78.09%	0.539
rndm vt # 9	10	656,394	34' 55"	78.18%	0.769
rndm vt # 10	10	656,394	34' 40"	**78.38%**	0.734
rndm vt # 11	10	656,394	32' 50"	77.73%	0.725
rndm vt # 12	10	656,394	32' 25"	77.73%	0.746

Table 10: Experiment results for random vectors on CIFAR10 classification task for quaternion convolutional neural networks with 80 epochs

What does order have to do with machine learning?

The filter of the simplex, adds variables one at a time:
$0, a_{00}, a_{00} + a_{01}, a_{00} + a_{01} + a_{11}, a_{00} + a_{01} + a_{10} + a_{11}$.
We restrict our attention to n-simplices whose vertices have 0,1 coordinates in R^n.

Lemma ([DCGSK+24])

Let C be a convex n-dimensional subset of the unit n-cube. If C is the union of n-simplices, all sharing the line from the zero vector to the one vector, then the convex set C is an order polytope.

Gracias I

Thomas Browning, Max Hopkins, and Zander Kelley.
Doppelgangers: the Ur-operation and posets of bounded height, 2018.

J. M. Boardman and R. M. Vogt.
Homotopy invariant algebraic structures on topological spaces, volume 347 of *Lect. Notes Math.*
Springer, Cham, 1973.

Eric Dolores-Cuenca, Aldo Guzman-Saenz, Sangil Kim, Susana Lopez-Moreno, and Jose Mendoza-Cortes.
Order theory in the context of machine learning: an application, 2024.

Frédéric Fauvet, Loïc Foissy, and Dominique Manchon.
Operads of finite posets.
Electron. J. Comb., 25(1):research paper p1.44, 23, 2018.

Gracias II

Samuele Giraudo.
Operads from posets and Koszul duality.
Eur. J. Comb., 56:1–32, 2016.

Maxim Kontsevich.
Operads and motives in deformation quantization.
Letters in Mathematical Physics, 48, 1999.

J. P. May.
The geometry of iterated loop spaces, volume 271 of *Lect. Notes Math.*
Springer, Cham, 1972.

Bernd Schröder.
Ordered sets. An introduction with connections from combinatorics to topology.
Basel: Birkhäuser/Springer, 2nd edition edition, 2016.

Gracias III

R. P. Stanley.
Two poset polytopes.
Discrete & Computational Geometry, 1:9–23, 1986.

Dmitry E. Tamarkin.
Formality of chain operad of little discs.
Lett. Math. Phys., 66(1-2):65–72, 2003.

W.T. Trotter.
Combinatorics and Partially Ordered Sets: Dimension Theory.
Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2002.

Liwen Zhang, Gregory Naitzat, and Lek-Heng Lim.
Tropical Geometry of Deep Neural Networks.
Preprint, arXiv:1805.07091 [cs.LG] (2018), 2018.

NN

Tropical algebra: $(\mathbb{R} \cup -\infty, \oplus, \otimes)$, $\oplus = \max, \otimes = +$.
To every poset P, we associate a tropical polynomial $\oplus x^v$ where v are corners of $Poly(P)$.

Example

The tropical polynomial of the linear order with n points $1 < 2 < \cdots < n$ is of the form:

$$0 \oplus x_n \otimes \Big(0 \oplus x_{n-1} \otimes \big(\cdots \otimes (0 \oplus x_1)\big)\Big). \qquad (1)$$

NN

Let $I_{1,n-1}$ be the $(n-1)$-dimensional identity matrix with the first row repeated. Given a vector of dimension n, consider $ReLU_{0,-\infty,\ldots,-\infty}$, the ReLU activation function with a parameter vector containing $n-1$ coordinates with value $-\infty$, and the first coordinate with value 0.

Lemma

The neural network associated to the tropical polynomial of a chain is an iteration of transformations of the form:

$$X_i \mapsto Y_i = ReLU_{0,-\infty,\ldots,-\infty}(X_i)$$

and

$$Y_i \mapsto X_{i-1} = X_i I_{1,i-1},$$

where each X_i is a $(1 \times i)$ vector and where X_n is the input.

Proof.

It follows from direct computation using [ZNL18] and Equation (1). □

NN

Table: Transpose of Posets, their tropical polynomials, their polytopes, and their IVNN.

Poset	$\{x < y\}$	$\{x < y < z\}$
Tropical Polynomial	$0 \oplus y \oplus xy$	$(0 \oplus z \oplus yz \oplus xyz)$
Polytope	$\{0 \leq x \leq y \leq 1\}$	$\{0 \leq x \leq y \leq z \leq 1\}$

The IVNN of $\{x < y\}$ is

$$ReLU(\begin{bmatrix} ReLU(x,0) & ReLU(y,-\infty) \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, 0)$$