Tropical Geometry of Deep Neural Networks

June 21, 2020

Contents

0 Introduction

1 Tropical Algebra

Our basic object of study is the following object $(\mathbb{R} \cup \{-\infty\}, \oplus, \odot)$. As a set this is the real numbers $\mathbb{R}$, together with an extra element $-\infty$ whitch represents minus infinity. In this (semiring) the tropical sum of real numbers is ther maximum and the tropical product of real numbers is their usual sum

$$x \oplus y := \max(x,\ y) \quad \& \quad x \odot y := x + y$$

——————————————————————

In Tropical Geometry often infinity instead of minus infinity and min in stead of max are used. This does not change any of the underlying theories of tropical algebra, as the the two semirings are tropically Isomorph. Meaning

$$(\mathbb{R} \cup \{-\infty\}, \oplus := \max, \odot) \to (\mathbb{R} \cup \{\infty\}, \oplus := \min, \odot), x \mapsto \begin{cases} -x & x \in \mathbb{R} \\ \infty & x = -\infty \end{cases}$$

is a tropical Isomorpism.

——————————————————————

First we will introduce the the tropical semiring and with it a usual semiring formally.

Definition 1.1. A semiring is a set R equipped with two binary operations $+$ and $\cdot$, called addition and multiplication, such that (Wikipedia.en Semiring)

- $(R,\ +)$ is a cummutative monoid with identity element 0:
  - $(a + b) + c = a + (b + c)$
  - $0 + a = a + 0 = 0$
  - $a + b = b + a$
- $(R,\ \cdot)$ is a monoid with identity element 1:
  - $(a \cdot b) \cdot c = a \cdot (b \cdot c)$
  - $1 \cdot a = a \cdot 1 = a$
- Multiplication left and right distributes over addition
  - $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$
  - $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$
- Multiplication by 0 annihilates
  - $0 \cdot a = a \cdot 0 = 0$

Proposition 1.2. $\mathbb{T} := (\mathbb{R} \cup \{-\infty\}, \oplus, \odot)$ is a semiring called the tropical semiring. (Maclagan & Sturmfels, 2015, p. 10)

Proof.

(1): The neutral element for the tropical sum is $-\infty$ since for $x \in \mathbb{R} \cup \{-\infty\}$ the following stands $x \oplus \infty = \max(x,\ -\infty) = x$ and with $x \odot 0 = x + 0 = x$ for $x \in \mathbb{R}$, 0 is the neutral elemet of tropical multiplication.

(2): Both addition and multiplication are commutative. To prove this we take $x, y \in \mathbb{R} \cup \{-\infty\}$ and do a case distinct. Because $\mathbb{R}$ is a field w.l.o.g. we set $x = -\infty, y \in \mathbb{R}$

$$-\infty \oplus y = \max(-\infty,\ y) = y = \max(y,\ -\infty) = y \oplus \infty$$
$$\infty \odot y = \infty + y = \infty = y + \infty = y \odot \infty$$

(3): Tropical multiplication distributes over addition. Take $x, y, z \in \mathbb{R}$ then

$$x \odot (y \oplus z) = x + \max(y,\ z) = \max(x + y,\ x + z) = (x \odot y) \oplus (x \odot z)$$
$$(y \oplus z) \odot x = \max(y,\ z) + x = \max(y + x,\ z + x) = (y \odot x) \oplus (z \odot x)$$

(4): Multiplication by $-\infty$ annihilates $-\infty \odot x = -\infty\, \forall x \in \mathbb{R} \cup \{-\infty\}$.

□

An essential feature of tropical arithmetics is that there is no subtraction. Take $a, b \in \mathbb{R} \cup \{-\infty\}$ with $a < b$ then the equation $a \oplus x = b$ has no solution x at all. (Maclagan & Sturmfels, 2015, p. 11)

To get more familiar with the tropical arithmetics we will have a look at an arithemtical example, before we jump into tropical polynomials.

Example 1.3. The tropical Pascal's triangle, whose rows are the coefficients appearing in a binomial expansion, is very simple to remember. All its coefficients are Zero. For example, the fourth row in the triangle is represented be the following equation

$$\begin{aligned}(a \oplus b)^3 &= 3\max(a, b) \\ &= \max(3a, 3b) = (a^3 + b^3) \\ &= \max(0 \odot a^3, 0 \odot a^2 \odot b, 0 \odot a \odot b^2, 0 \odot b^3) with a, b \in \mathbb{T}\end{aligned}$$

You may say the Pascal's coefficients are four zeroes. The same applies to all cases

$$\begin{aligned}(a \oplus b)^n &= a^n \oplus b^n \\ &= 0a^n \oplus 0a^{n-1}b \oplus \cdots \oplus 0ab^{n-1} \oplus b^n.\end{aligned}$$

And Pascal's triangle has the following form.

$$
\begin{array}{llllllllllllll}
n=0 & & & & & & & 0 & & & & & & \\
n=1 & & & & & & 0 & & 0 & & & & & \\
n=2 & & & & & 0 & & 0 & & 0 & & & & \\
n=3 & & & & 0 & & 0 & & 0 & & 0 & & & \\
n=4 & & & 0 & & 0 & & 0 & & 0 & & 0 & & \\
n=5 & & 0 & & 0 & & 0 & & 0 & & 0 & & 0 & \\
n=6 & 0 & & 0 & & 0 & & 0 & & 0 & & 0 & & 0.
\end{array}
$$

The goal of this first section and the following is to get to understand tropical rational functions and $<$ maps, as the underlying objects are key to later relations between neural networks and Tropical Geometry. Linking these two field is core to this thesis. First, so that we can introduce multidimensional tropical polynomials properly a notion of monomials is needed.

Definition 1.4. (Zhang, Naitzat, & Lim, 2018, p. 2) A tropical monomial in d variables $x_1, \ldots, x_d$ is an expression or function $\mathbb{T}^d \to \mathbb{T}$ of the from

$$c \odot x_1^{a_1} \odot x_2^{a_2} \odot \cdots \odot x_{d-1}^{a_{d-1}} \odot x_d^{a_d}$$

where $c \in \mathbb{R} \cup \{-\infty\}$ and $a_1, \ldots, a_d \in \mathbb{N}$. As a convenient shorthand, we will also write a tropical monomial in multiindex notation as cx_α where $\alpha = (a_1, \ldots, a_d) \in \mathbb{N}_d$ and $x = (x_1, \ldots, x_d)$. Note that $x^\alpha = 0 \odot x^\alpha$.

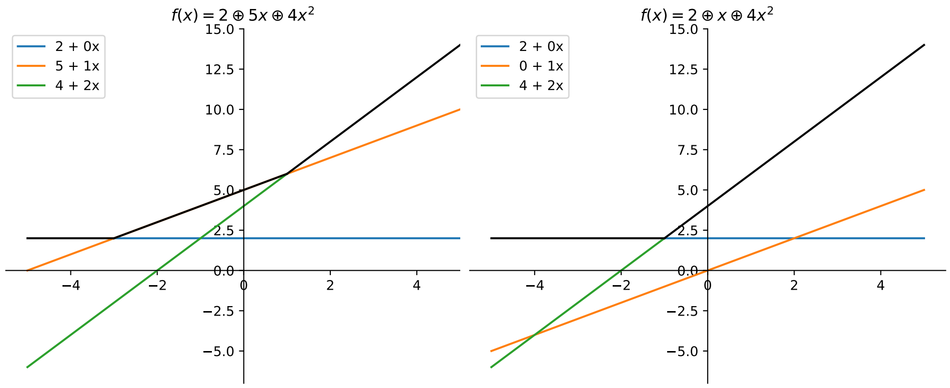
Definition 1.5. (Zhang et al., 2018, p. 2) Following notations above, a tropical polynomial $f(x) = f(x_1, \ldots, x_d), f : \mathbb{T}^d \to \mathbb{T}$ is a finite tropical sum of tropical monomials

$$f(x) = c_1 x^{\alpha_1} \oplus \cdots \oplus c_r x^{\alpha_r}$$

where $\alpha_i = (\alpha_{i1}, \ldots, \alpha_{id}) \in \mathbb{N}^d$ and $c_i \in \mathbb{R} \cup \{-\infty\}$, $i = 1, \ldots, r$. We will assume that a monomial of a given multiindex appears at most once in the sum, i.e. $\alpha_i \neq \alpha_j$ for any $i \neq j$.

Example 1.6. We are going to examine tropical polynomials in one variable. Monomials in one variable are of the form $c \odot x^a, \mathbb{T} \to \mathbb{T}$ where $c \in \mathbb{T}$ and $a \in \mathbb{N}$. Meaning a tropical polynomial in one variable is of the form $f(x) = c_1 x^{\alpha_1} \oplus \cdots \oplus c_r x^{\alpha_r}\ with\ c_i \in \mathbb{T}, a_i \in \mathbb{N}\ for\ i = 1, \ldots, r$.
For a quadratic tropical polynomial $f(x) = a \oplus bx \oplus cx^2 witha, b, c \in \mathbb{T}$ linearity breaks at two points $b - c$ and $a - b$ if $b - c > a$, otherwise it only breaks at one point $(a - c) : 2$. Visualising two quadratic tropical functions gives a good intuition of the piecewise-linearity.
The black colored parts of the followig figures one and two indicate the graph of the tropical polynomial.

We can see the degree of a tropical polynomial, defined the same as a degree of a usual polynomial, gives an upper bound for the number of non linear edges of the tropical polynomial, but not the exact value. With higher degree polynomials more non linear edges are possible:



(a) Take $f(x) = 2 \oplus 5x \oplus 4x^2$ a quadratic polynomial. The graph equals $max(2, 5 + x, 4 + 2x)$. Until $5 - 2$, 2 dominates, from there $5 + x$ dominates to the point $4 + 2x$.

(b) Take $f(x) = 2 \oplus x \oplus 4x^x$, we have changed the scalar in the second part, then this second part plays no part in the tropical polynomial.

Figure 1: Quadratic tropical polynomials.

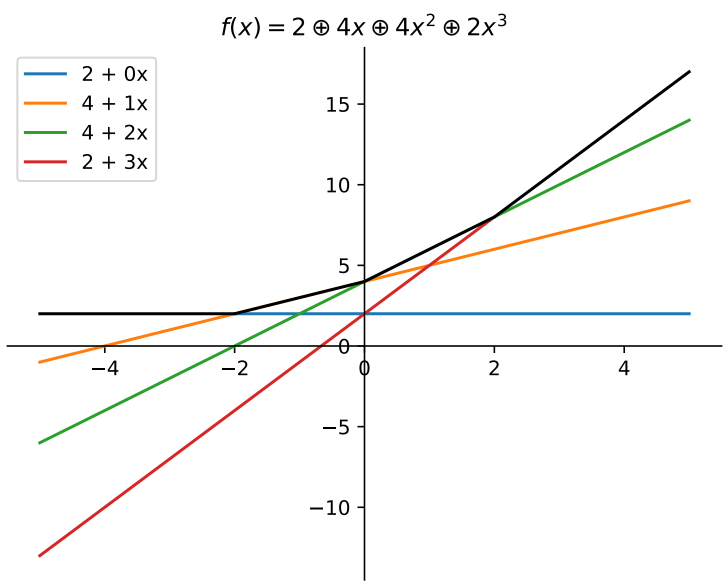


Figure 2: Take $f(x) = 2 \oplus x \oplus 4x^x \oplus 4x^2 \oplus 3x^3$. For a degree three tropical polynomial in one variable this polynomial has reached the maximum number of non linear edges.

In multiple variables we can characterise tropical polynomials as functions from $\mathbb{R}^n \to \mathbb{R}$ that satisfy the following three properties.

Lemma 1.7. Let f be a tropical polynomial

$$f(x) = c_1^{\alpha_1} \oplus \cdots \oplus c_r x^{\alpha_r}$$

as in 1.5. Then f has three important properties:

(1) f is continuous.

(2) f is piecewise-linear, where the number of pieces is finite.

(3) f is convex, i.e. $p(\frac{x+y}{2}) \leq \frac{1}{2}(p(x)+p(y)) \forall x, y \in \mathbb{R}$

Proof. (1) The minimum of continuous functions is still continuous.
(2) Every monomial $c_i x^{\alpha_i} = c_i + xy\alpha_{i1} + \cdots + x_r \alpha_{ir}$ is per definition linear. Because of (1), linearity can only be broken where $c_i x^{\alpha_i} = c_j x^{\alpha_j}$ for $i \leq j$ and $i, j = 1, \ldots, r$.
A piece is a single piece of $c_i x^{\alpha_i}$ where linearity is not broken. If we introduce $c_i x^{\alpha_i}$ one after another, then in the i-th step not more than i^2 new pieces can be created, so there can only be $\sum_{i=1}^{r} i^2$ or less pieces.
(3) On one piece f is convex. Moving from one piece to another the slope can only increase, with means f is still convex. ☐

Proposition 1.8. Every function $\mathbb{R}^n \to \mathbb{R}$ which satisfies the three properties (1), (2) and (3) has a representation as the minimum of a finite set of linear functions. Thus, the tropical polynomial in n variables $x_1, \ldots x_n$ represent the class of piecewise-linear convex functions on $\mathbb{R}^n$ with integer coefficients.

Proof. This follows directly from lemma 1.7. ☐

Now we are ready to introduce tropical rational functions. These are important to understand the core section, section 4, in particular the actual connection build between neural networks and tropical geometry in this thesis, as are the semifields $\mathbb{T}[X_1, \ldots, X_d]$ and $\mathbb{T}(X_1, \ldots, X_d)$.

Definition 1.9. (Zhang et al., 2018, p. 3) Following notations above, a tropical rational function is a standard difference, or, equivalently, a tropical quotient of two tropical polynomials $f(x)$ and $g(x)$:

$$(f-g)(x) = f(x) - g(x) = f(x) \oslash g(x) = (f \oslash g)(x)$$

Proposition 1.10. $\mathbb{T}[X_1, \ldots, X_d] := \{f : \mathbb{T}^d \to \mathbb{T}; f \textit{ is tropical polynomial}\}$ and $\mathbb{T}(X_1, \ldots, X_d) := \{f : \mathbb{T}^d \to \mathbb{T}; f \textit{ is tropical rational function}\}$ are both semifields. (Zhang et al., 2018, p. 3)

Proof. Let $g, f, h \in \mathbb{T}(X_1, \ldots, X_d)$ with

$$f(x) = f_1(x) \oslash f_2(x) = (\oplus_{i=0}^r c_{1i} x^{\alpha_{1i}}) \oslash (\oplus_{i=0}^r c_{2i} x^{\alpha_{2i}})$$
$$g(x) = g_1(x) \oslash g_2(x) = (\oplus_{i=0}^r d_{1i} x^{\beta_{1i}}) \oslash (\oplus_{i=0}^r d_{2i} x^{\beta_{2i}})$$
$$h(x) = h_1(x) \oslash h_2(x) = (\oplus_{i=0}^r e_{1i} x^{\gamma_{1i}}) \oslash (\oplus_{i=0}^r e_{2i} x^{\gamma_{2i}})$$

We begin the proof by showing, that for two tropical polynomials $a(x) = \oplus_{i=0}^{r} z_i x^{\zeta_i}), b(x) = \oplus_{i=0}^{r} o_i x^{\omega_i}) \in \mathbb{T}[X_1, \ldots, X_r]$ the normal sum is a tropical polynomial $(a+b)(x) \in \mathbb{T}[X_1, \ldots, X_r]$.

$$\begin{aligned}
(a+b)(x) &= a(x) + b(x) \\
&= (\oplus_{i=0}^{r} z_i x^{\zeta_i}) + (\oplus_{i=0}^{r} o_i x^{\omega_i}) \\
&= \oplus_{i,j=0,\ldots,r}(z_i + \zeta_i * x + o_j + \omega_i * x) \\
&= \oplus_{i,j=0,\ldots,r}((z_i + o_j) + x^{\zeta_i + \omega_i}) \in \mathbb{T}[X_1, \ldots X_r]
\end{aligned}$$

Other than with the proof of the Tropical semiring we will show that topical tropical addition and tropical multiplication of tropical polynomials as tropical rational functions stay tropical polynomials respectively tropical rational functions. All other axioms stay pointwise the same.

(1): The Tropical sum of a Tropical rational functions is a tropical tropical rations function

$$\begin{aligned}
(f \oplus g)(x) &= f(x) \oplus g(x) \\
&= (f_1(x) \oslash f_2(x)) \oplus (g_1(x) \oslash g_2(x)) \\
&= \min\{f_1(x) - f_2(x), g_1(x) - g_2(x)\} \\
&= \min\{f_1(x) + g_2(x), g_1(x) + f_2(x)\} - f_2(x) - g_2(x) \\
&= (f_1(x) + g_2(x) \oplus g_1(x) + f_2(x)) \oslash (f_2(x) + g_2(x)) \in \mathbb{T}(X_1, \ldots, X_r).
\end{aligned}$$

Since Addition as tropical addition of tropical polynomials is a tropical polynomial.

(2): The Tropical product of a Tropical rational functions is a tropical tropical rational function

$$\begin{aligned}
(f \odot g)(x) &= f(x) \odot g(x) \\
&= (f_1(x) \oslash f_2(x)) \odot (g_1(x) \oslash g_2(x)) \\
&= f_1(x) - f_2(x) + g_1(x) - g_2(x) \\
&= (f_1(x) + g_1(x)) - (f_2(x) + g_2(x)) \in \mathbb{T}(X_1, \ldots, X_r)
\end{aligned}$$

(3): The neutral element for the tropical sum is $-\infty = -\infty \oslash x = x \oslash -\infty \; x \in \mathbb{T}$ since for $f(x) \in \mathbb{T}(X_1, \ldots, X_r)$ as above, the following stands $f(x) \oplus \infty = \max(f(x),\ -\infty) = f(x)$ and with $f(x) \odot 0 = f(x) + 0 = f(x) \forall f(x) \in \mathbb{T}$ 0 is the neutral element of tropical multiplication.

□

Comment 1.11. We regard a tropical polynomial $f = f \oslash 0$ as a special case of a tropical rational function and thus $\mathbb{T}[X_1, \ldots, X_r] \subseteq \mathbb{T}(X_1, \ldots, X_r)$ (Zhang et al., 2018, p. 3).

Comment 1.12.

- A d-variate tropical polynomial $f(x)$ defines a function $f : \mathbb{T}^d \to \mathbb{T}$ that is a convex function in the usual sense as taking max and $\sum$ of convex functions preserve convexity (Boyd, Boyd, & Vandenberghe, 2004).

- As such, a tropical rational function $f \oslash g : \mathbb{T}^d \to \mathbb{T}$ is a DC function or differenceconvex function (Hartman et al., 1959).

Example 1.13. The only thing missing to this section are some examples for tropical rational functions and tropical rational maps.
!!! Expand by adding some examples !!!

Definition 1.14. $R : \mathbb{R}^d \to \mathbb{R}^p, x = (x_1, \ldots, x_d) \mapsto (f_1(x), \ldots, f_p(x))$, is called a tropical polynomial map if each $f_i : \mathbb{R}^d \to \mathbb{R}$ is a tropical polynomial, $i = 1, \ldots, p$, and a tropical rational map if $f_1, \ldots, f_p$ are tropical rational functions. We will denote the set of tropical polynomial maps by $Pol(d, p)$ and the set of tropical rational maps by $Rat(d, p)$. So $Pol(d, 1) = \mathbb{T}[X_1, \ldots, x_d]$ and $Rat(d, 1) = \mathbb{T}(x_1, \ldots, x_d)$ (Zhang et al., 2018, p. 3).

2 Tropical hypersurfaces

Definition 2.1. The tropical hypersurface of a tropical polynomial $f(x) = c_1x^{\alpha_1} \oplus \cdots \oplus c_rx^{\alpha_r}$ is

$$\Gamma(f) := \{x \in \mathbb{R}^d : c_ix^{\alpha_i} = c_jx^{\alpha_j} = f(x) for some \alpha_i \neq \alpha_j\}$$

i.e., the set of points x at which the value of f at x is attained by two or more monomials in f (Zhang et al., 2018, p. 3).

Comment 2.2. A tropical polynomial f determines a dual subdivision of $\delta(f)$, constructed as follows. First, lift each α_i from $\mathbb{R}^d$ into $\mathbb{R}^{d+1}$ by appending c_i as the last coordinate. Denote the convex hull of the lifted $\alpha_1, \ldots, \alpha_r$ as

$$\mathcal{P}(f) := Conv(\alpha_i, c_i) \in \mathbb{R}^d \times \mathbb{R} : i = 1, \ldots, r.(1)$$

Next let $UF((P)(f))$ denote the collection of upper faces in $\mathcal{P}(f)$ and $\pi : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ be the projection that drops the last coordinate. The dual subdivision determined by f is then

$$\delta(f) := \pi(p) \subset \mathbb{R}^d : p \in UF(\mathcal{P}(f)).$$

. $\delta(f)$ forms a polyhedral complex with support $\delta(f)$. By (Maclagan & Sturmfels, 2015), the tropical hypersurface $\mathcal{T}(f)$ is the $(d-1)$-skeleton of the polyhedral complex dual to $\delta(f)$. This means that each vertex in $\delta(f)$ corresponds to one "cell" in R^d where the function f is linear. Thus, the number of vertices in $\mathcal{P}(f)$ provides an upper bound on the number of linear regions of f.

Definition 2.3. The Newton polygon of a tropical polynomial $f(x) = c_1x^{\alpha_1} \oplus \cdots \oplus c_rx^{\alpha_r}$ is the convex hull of $\alpha_1, \ldots, \alpha_r \in \mathbb{N}^d$, regarded as points in $\mathbb{R}^d$,

$$\Delta(f) := Conv \alpha_i \in \mathbb{R}^d : c_i \neq -\infty, i = 1, \ldots, r$$

(Zhang et al., 2018, p. 3).

Definition 2.4. A linear region of $F \in Rat(d, m)$ is a maximal connected subset of the domain on which F is linear. The number of linear regions of F is denoted $\mathcal{N}(f)$ (Zhang et al., 2018, p. 4).

2.1 Transformations of tropical polynomial

Proposition 2.5. Let f be a tropical polynomial and let $a \in \mathbb{N}$. Then

$$\mathcal{P}(f^a) = a\mathcal{P}(f)$$

. $a\mathcal{P}(f) = \{ax : x \in \mathcal{P}(f)\} \subset \mathbb{R}^{d+1}$ is a scaled version of $\mathcal{P}(f)$ with the same shape but different volume (Zhang et al., 2018, p. 4).

Proof. □

Definition 2.6. The Minkowski sum of two sets P_1 and P_2 in $\mathbb{R}^d$ is the set

$$P_1 + P_2 := \{x_1 + x_2 \in \mathbb{R}^d : x_1 \in P_1, x_2 \in P_2\}$$

; and for $\lambda_1, \lambda_2 \geq 0$, their weighted Minkowski sum is

$$\lambda_1 P_1 + \lambda_2 P_2 := \{\lambda_1 x_1 + \lambda_2 x_2 \in \mathbb{R}^d : x_1 \in P_1, x_2 \in P_2\}$$

(Zhang et al., 2018, p. 4).

Proposition 2.7. (Zhang et al., 2018, p. 4) Let $f, g \in Pol(d, 1) = \mathbb{T}[x_1, \ldots, x_d]$ be tropical polynomials. Then

$$\mathcal{P}(f \odot g) = \mathcal{P}(f) + \mathcal{P}(g), \mathcal{P}(f \oplus g) \qquad = Conv(\mathcal{V}(\mathcal{P}(f)) \cup \mathcal{V}(\mathcal{P}(g))).$$

Theorem 2.8. (Gritzmann-Sturmfels). Let $P_1, \ldots, P_k$ be polytopes in $\mathbb{R}^d$ and let m denote the total number of nonparallel edges of $P_1, \ldots, P_k$. Then the number of vertices of $P_1 + \cdots + P_k$ does not exceed

$$\sum_{j=0}^{d-1} \binom{m-1}{j}.$$

The upper bound is attained if all P_i's are zonotopes and all their generating line segments are in general positions. (Gritzmann & Sturmfels, 1993)

Corollary 2.9. Let $\mathcal{P} \in \mathbb{R}^{d+1}$ be a zonotope generated by m line segments $P_1, \ldots, P_m$. Let $\pi : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ be the projection. Suppose P satisfies:

(i) the generating line segments are in general positions;

(ii) the set of projected vertices $\{\pi(v) : v \in \mathcal{V}(\mathcal{P})\} \subseteq \mathbb{R}^d$ are in generalposition.

Then P has

$$\sum_{j=0}^{d} \binom{m}{j}$$

vertices on its upper faces. If either (i) or (ii) is violated, then this becomes an upper bound. (Zhang et al., 2018, p. 4)

3 Neural networks

Neural networks viewed as a topic, make for a very interesting field of interest in their own right. Historically the term "Neural Network" was introduced in attempts to describe the functionality of biological processes, in particular the nervous system and the brain, in an mathematical sense (McCulloch & Pitts, 1943; Widrow & Hoff, 1960; Rumelhart, Hinton, & Williams, 1986). Simplified the nervous system is a net of neurons, each having a soma and an axon. At any instant a neuron has some threshold, which excitation must exceed to initiate an impulse. This, except for the fact and the time of its occurrence, is determined by the neuron, not by the excitation. From the point of excitation the impulse is propagated to all parts of the neuron (McCulloch & Pitts, 1943). Through synapses the axons are connected to further soma through with the impulse is passed on. Impulses passing through the nervous system partly consist of electrical impulses and chemical reactions (Palay, 1956).

Depiction of a neuron

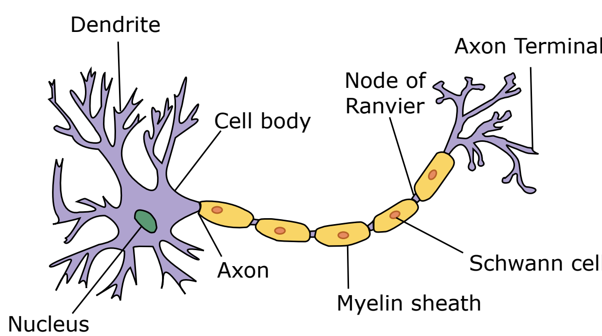


Figure 3: A representation of a neuron. The axon terminal attach to dendrite. This way impulses pass from neuron to neuron.

Definition 3.1. Viewed abstractly, an L-layer feedforward neural network is a map $\nu : \mathbb{R}^d \to \mathbb{R}^p$ given by a composition of functions

$$\nu = \sigma^{(L)} \circ \rho^{(L)} \circ \sigma^{(L-1)} \circ \rho^{(L-1)} \circ \cdots \circ \sigma^{(1)} \circ \rho^{(1)}$$

The preactivation functions $\rho^{(1)}, \ldots, \rho^{(L)}$ are affine transformations to be determined and the activation functions $\sigma^{(1)}, \ldots, \sigma^{(L)}$ are chosen and fixed in advanced.
We denote the width, i.e., the number of nodes, of the lth layer by $n_l, l = 1, \ldots, L-1$. We set $n_0 := d$ and $n_L := p$, respectively the dimensions of the input and output of the network. The output from the lth layer will be denoted by

$$\nu = \sigma^{(l)} \circ \rho^{(l)} \circ \sigma^{(l-1)} \circ \rho^{(l-1)} \circ \cdots \circ \sigma^{(1)} \circ \rho^{(1)}$$

i.e., it is a map $\nu^{(l)} : \mathbb{R}^d \to \mathbb{R}^{n_l}$. For convenience, we assume $\nu^{(0)}(x) := x$.
The affine function $\nu^{(l)} : \mathbb{R}^{n_{l-1}} \to \mathbb{R}^{n_l}$ is given by a weight matrix $A^{(l)} \in$

$\mathbb{Z}^{n_l \times n_{l-1}}$ and a bias vector $b^{(l)} \in \mathbb{R}^{n_l}$:

$$\rho^{(l)}(\nu^{(l-1)}) := A^{(l)}\nu(l-1) + b(l).$$

The (i,j)th coordinate of $A^{(l)}$ will be denoted a^l_{ij} and the ith coordinate of $b^{(l)}$ by $b^{(l)}_i$. Collectively they form the parameters of the lth layer

Comment 3.2. For a vector input $x \in \mathbb{R}^{n_l}, \sigma^{(l)}(x)$ is understood to be in coordinatewise sense; so $\sigma : \mathbb{R}^{n_l} \to \mathbb{R}^{n_l}$. We assume the final output of a neural network $\nu(x)$ is fed into a score function $s : \mathbb{R}^p \to \mathbb{R}^m$ that is application specific.

Comment 3.3. We will make the following mild assumptions on the architecture of our feedforward neural networks:

(a) the weight matrices $Aa^{(1)}, \ldots, A^{(L)}$ are integer-valued;

(b) the bias vectors $b^{(1)}, \ldots, b^{(L)}$ are real-valued;

(c) the activation functions $\sigma^{(1)}, \ldots, \sigma^{(L)}$ take the form

$$\sigma^{(l)}(x) := \max\{c, t^{(L)}\},$$

where $t^{(l)} \in (\mathbb{R} \cup \{-\infty\})^{n_l}$ is called a threshold vector.

Henceforth all neural networks in our subsequent discussions will be assumed to satisfy (a)–(c).

4 Tropical algebra of neural networks

Comment 4.1. Consider the output from the first layer in a neural network

$$\nu(x) = \max\{Ax + b, t\};$$

where $A \in \mathbb{Z}^{p\times d}, b \in \mathbb{R}^p$, and $t \in (\mathbb{R} \cup \{Ax + b, t\})$. We will decompose A as a difference of two nonnegative integervalued matrices, $A = A_+ - A_-$ with $A_+, A_- \in \mathbb{N}^{p\times d}$; e.g., in the standard way with entries

$$a_{ij}^+ := \max\{a_{ij}, 0\},\ a_{ij}^- := \max\{-a_i j, 0\}$$

respectively. Since

$$\max\{Ax + b, t\} = \max\{A_+x + b, A_-x + t\} - A_{-x},$$

we see that every coordinate of one-layer neural network is a difference of two tropical polynomials.

For networks with more layers, we apply this decomposition recursively to obtain the following result

Proposition 4.2. Let $A \in \mathbb{Z}^{m\times n}, b \in \mathbb{R}^m$ be the parameters of the $(l + 1)$th layer, and let $t \in (\mathbb{R} \cup -\infty)^m$ be the threshold vector in the $(l + 1)$th layer. If the nodes of the lth layer are given by tropical rational functions,

$$\nu^{(l)}(x) = F^{(l)}(x) \oslash G^{(l)}(x) = F^{(l)}(x) - G^{(l)}(x),$$

i.e., each coordinate of $f^{(l)}$ and $G^{(l)}$ is a tropical polynomial in x, then the outputs of the preactivation and of the $(l + 1)$th layer are given by tropical rational functions

$$\rho^{(l+1)} \circ \nu^{(l)}(x) = H^{(l+1)}(x) - G^{(l+1)}(x),$$
$$\nu^{(l+1)}(x) = \sigma \circ \rho^{(l+1)} \circ \nu^{(l)}(x) = F^{(l+1)}(x) - G^{(l+1)}(x),$$

respectively, where

$$F^{(l+1)}(x) = \max\{H^{(l+1)}(x), G(l+^{)}(x) + t\},$$
$$G^{(l+1)}(x) = A_x G^{(l)}(x) + A_- F^{(l)}(x),$$
$$H^{(l+1)}(x) = A_+ F^{(l)}(x) + A_- G(l)(x) + b.$$

Induction yields the following.

Theorem 4.3. (Tropical characterization of neural networks).
A feedforward neural network under assumptions (a)–(c) is a function $\nu : \mathbb{R}^d \to \mathbb{R}^p$ whose coordinates are tropical rational functions of the input, i.e.,

$$\nu(x) = F(x) \oslash G(x) = F(x) - G(x)$$

where F and G are tropical polynomial maps. Thus ν is a tropical rational map.

Comment 4.4. Note that the tropical rational functions above have real coefficients, not integer coefficients. The integer weights $A^(l) \in \mathbb{Z}^{n_l \times n_{l-1}}$ have gone into the powers of tropical monomials in f and g, which is why we require our weights to be integer-valued, although as we have explained, this requirement imposes little loss of generality

By setting $t^{(1)} = \cdots = t^{(L-1)} = 0$ and $t^{(l)} = -\infty$, we obtain the following corollary.

Corollary 4.5. Let $\nu : \mathbb{R}^d \to \mathbb{R}$ be an ReLU activated feedforward neural network with integer weights and linear output. Then ν is a tropical rational function.

Theorem 4.6. (Equivalence of neural networks and tropical rational functions).

(i) Let $\nu : \mathbb{R}^d \to \mathbb{R}$. Then ν is a tropical rational function if and only if ν is a feedforward neural network satisfying assumptions (a)–(c).

(ii) A tropical rational function $f \oslash g$ can be represented as an L-layer neural network, with

$$L \leq \max\{\lceil \log_2 r_f \rceil, \lceil \log_2 r_g \rceil\} + 2,$$

where r_f and r_g are the number of monomials in the tropical polynomials f and g respectively.

Proposition 4.7. Let $\nu : \mathbb{R}^d \to \mathbb{R}$. Then ν is a continuous piecewise linear function with integer coefficients if and only if ν is a tropical rational function.

Comment 4.8. Corollary 5.3, Theorem 5.4, and Proposition 5.5 collectively imply the equivalence of

(i) tropical rational functions,

(ii) continuous piecewise linear functions with integer coefficients,

(iii) neural networks satisfying assumptions (a)-(c).

Proposition 4.9. Every feedforward neural network with ReLU activation is a tropical rational signomial map.

References

Boyd, S., Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Gritzmann, P., & Sturmfels, B. (1993). Minkowski addition of polytopes: computational complexity and applications to gröbner bases. *SIAM Journal on Discrete Mathematics*, *6*(2), 246–269.

Hartman, P., et al. (1959). On functions representable as a difference of convex functions. *Pacific Journal of Mathematics*, *9*(3), 707–713.

Maclagan, D., & Sturmfels, B. (2015). *Introduction to tropical geometry* (Vol. 161). American Mathematical Soc.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*(4), 115–133.

Palay, S. L. (1956). Synapses in the central nervous system. *The Journal of biophysical and biochemical cytology*, *2*(4), 193.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, *323*(6088), 533–536.

Widrow, B., & Hoff, M. E. (1960). *Adaptive switching circuits* (Tech. Rep.). Stanford Univ Ca Stanford Electronics Labs.

Zhang, L., Naitzat, G., & Lim, L.-H. (2018). Tropical geometry of deep neural networks. *arXiv preprint arXiv:1805.07091*.