Multiplication and Universal Approximation

On the Philosophy of Lines

Oscar I. Hernandez

Department of Mathematics & Statistics
University of Alaska Fairbanks

February 1, 2022

Goal

Let $k, m, n \in \mathbb{N}$. Let K be a compact subspace of $\mathbb{R}^m$, e.g. $[0, 1]^m$.

Idea

Let $\sigma : \mathbb{R} \to \mathbb{R}$ and $f : K \to \mathbb{R}^n$ be continuous. Suppose σ is not a polynomial, i.e. $\sigma(x) \neq a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$. Then, we can use it in a small (shallow feed-forward) neural network to approximate f as well as we want.

Theorem ([1])

Let $\sigma : \mathbb{R} \to \mathbb{R}$ and $f : K \to \mathbb{R}^n$ be continuous. The function σ is not a polynomial iff for any $\varepsilon > 0$ there are a pair of affine transformations $A_1 : \mathbb{R}^m \to \mathbb{R}^k$, $A_2 : \mathbb{R}^k \to \mathbb{R}^n$ such that $N = A_2 \circ \vec{\sigma} \circ A_1$ satisfies:

$$||N - f||_\infty < \varepsilon. \tag{1}$$

Outline

Linearity: Matrix Multiplication, Affine Transformation

Let $W : \mathbb{R}^2 \to \mathbb{R}^4$ be defined below.

$$W : \begin{pmatrix} u \\ v \end{pmatrix} \mapsto \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} w_{11}u + w_{12}v \\ w_{21}u + w_{22}v \\ w_{31}u + w_{32}v \\ w_{41}u + w_{42}v \end{pmatrix} \tag{2}$$

For any $b \in \mathbb{R}^4$, we obtain an *affine transformation* $A : \mathbb{R}^2 \to \mathbb{R}^4$.

$$A : \begin{pmatrix} u \\ v \end{pmatrix} \mapsto W \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} = \begin{pmatrix} w_{11}u + w_{12}v + b_1 \\ w_{21}u + w_{22}v + b_2 \\ w_{31}u + w_{32}v + b_3 \\ w_{41}u + w_{42}v + b_4 \end{pmatrix} \tag{3}$$

Non-Linearity: Activation

Let $\sigma : \mathbb{R} \to \mathbb{R}$ be a continuous function with two derivatives, e.g. $\sigma(x) = \frac{1}{1+e^{-(x)}}$. Also need σ to not be a polynomial.

Let $\vec{\sigma} : \mathbb{R}^k \to \mathbb{R}^k$ apply to each component, as shown below.

$$\vec{\sigma} : \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_k \end{pmatrix} \mapsto \begin{pmatrix} \sigma(x_1) \\ \sigma(x_2) \\ \dots \\ \sigma(x_k) \end{pmatrix} \tag{4}$$

Taylor Polynomial and Multiplication Claim

By Taylor's theorem, we obtain:

$$\sigma(x) = \sigma(0) + x\sigma'(0) + \frac{x^2}{2}\sigma''(0) + o(x^3). \tag{5}$$

Test: https://www.desmos.com/calculator/5mo5qmpyui

Lemma ([2])

$$\frac{\sigma(u+v) + \sigma(-u-v) - \sigma(u-v) - \sigma(-u+v)}{4\sigma''(0)} \to uv \text{ as } u, v \to 0 \tag{6}$$

Multiplication Approximation

Proof.

$$\frac{8uv + o((u+v)^3) + o((-u-v)^3) + o((u-v)^3) + o((-u+v)^3)}{8} \tag{7}$$

$$= uv\left[1 + o(u^2+v^2)\right] \to uv \text{ as } u, v \to 0. \tag{8}$$

□

$$\frac{1}{4\sigma''(0)}\left(\begin{pmatrix}1 & 1 & 1 & 1\end{pmatrix}\vec{\sigma}\left(\begin{pmatrix}+1 & +1\\ -1 & -1\\ +1 & -1\\ -1 & +1\end{pmatrix}\begin{pmatrix}u\\ v\end{pmatrix} + \begin{pmatrix}0\\0\\0\\0\end{pmatrix}\right) + (0)\right) \tag{9}$$

Test of Convergence

See computational notebook.

Continuous k-ary Multiplication

"Why does deep and cheap learning work so well?"

Outlook

We can approximate the following.

1. additions of multiples, i.e. linear combinations, i.e. polynomials $p : \mathbb{R}^m \to \mathbb{R}$
2. vector-valued polynomials $\vec{p} : \mathbb{R}^m \to \mathbb{R}^n$
3. any set in which vector-valued polynomials are dense, i.e. continuous functions $f : \mathbb{R}^m \to \mathbb{R}^n$

Theorem ([1])

Let $\sigma : \mathbb{R} \to \mathbb{R}$ and $f : \mathbb{R}^m \to \mathbb{R}^n$ be continuous. The function σ is non-linear iff for any $\varepsilon > 0$ there are a pair of affine transformations $A_1 : \mathbb{R}^m \to \mathbb{R}^k$, $A_2 : \mathbb{R}^k \to \mathbb{R}^n$ such that $N = A_2 \circ \vec{\sigma} \circ A_1$ satisfies:

$$||N - f||_\infty < \varepsilon. \tag{10}$$

Generalization

- $\sigma''(0) \neq 0$
- $0 \leq |u|, |v| \ll 1$

References

George Cybenko, *Approximation by superpositions of a sigmoidal function*, Mathematics of control, signals and systems **2** (1989), no. 4, 303–314.

Henry W Lin, Max Tegmark, and David Rolnick, *Why does deep and cheap learning work so well?*, Journal of Statistical Physics **168** (2017), no. 6, 1223–1247.