The Stone-Weierstrass Theorem for Neural Networks

Rohan Hitchcock

10 March 2021

These notes are based on the first theorem (unnumbered) in Section II.C.1 of [1] and the following discussion which presents a proof for the theorem that neural networks can uniformly approximate any continuous function on a compact space. This proof uses the versions of the Stone-Weierstrass Theorem and Taylor's Theorem stated below.

Theorem (Taylor's)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a k-differentiable function at $x = 0$. Then there exists $h_k : \mathbb{R} \to \mathbb{R}$ such that*

$$f(x) = \sum_{n=0}^{k} f^{(n)}(0)\frac{x^n}{n!} + h_k(x)x^k$$

and $h_k(x) \to 0$ as $x \to 0$.

Theorem (Stone-Weierstrass)**.** *Let $X \subseteq \mathbb{R}^d$ be compact. Then for any continuous $F : X \to \mathbb{R}$ there exists a polynomial function $P : X \to \mathbb{R}$ and $\epsilon > 0$ such that $|F(x) - P(x)| < \epsilon$ for all $x \in X$.*

We now prove the following theorem.

Theorem. *Let $X \subseteq \mathbb{R}^d$ be compact. Then for any continuous $F : X \to \mathbb{R}$ and $\epsilon > 0$ there exists a neural network function $N : X \to \mathbb{R}$ such that $|F(x) - N(x)| < \epsilon$ for all $x \in X$.*

Proof. It suffices to show that there is a neural network function M which can uniformly approximate multiplication on a compact subset of $\mathbb{R}^2$. This is done in Lemma 1 below. Then, given a polynomial function P approximating F, we can construct N by connecting copies of M so as to approximate P. □

Lemma 1. *Let $X \subseteq \mathbb{R}^2$ be compact. Then given $\epsilon > 0$ there exists a neural network function $M : X \to \mathbb{R}$ such that $|xy - M(x, y)| < \epsilon$ for all $(x, y) \in X$. The activation functions of M may be any function $f : \mathbb{R} \to \mathbb{R}$ which is twice-differentiable at $x = 0$ and $f''(0) > 0$.*

Proof. Let $f : \mathbb{R} \to \mathbb{R}$ be a twice-differentiable function at $x = 0$ such that $f''(0) \neq 0$. Set $f_0 = f(0)$, $f_1 = f'(0)$ and $f_2 = f''(0)$. Then by Taylor's Theorem we have

$$f(x) = f_0 + f_1 x + \frac{1}{2} f_2 x^2 + h(x)x^2$$

for some $h(x) \to 0$ as $x \to 0$. Then define

$$\begin{aligned} m(x, y) &\coloneqq \frac{1}{4f_2}\left(f(x + y) + f(-x - y) - f(x - y) - f(-x + y)\right) \\ &= xy + \frac{1}{4f_2}(x + y)^2\left[h(x + y) + h(-x - y)\right] - \frac{1}{4f_2}(x - y)^2\left[h(x - y) - h(-x + y)\right]. \end{aligned}$$

For any $\epsilon > 0$ we can choose $\delta > 0$ such that if $|x| < \delta$ and $|y| < \delta$ we have $|h(x + y)| < \epsilon \cdot f_2/2$ and $|h(-x - y)| < \epsilon \cdot f_2/2$ and $|h(x - y)| < \epsilon \cdot f_2/2$ and $|h(-x + y)| < \epsilon \cdot f_2/2$. Then $|xy - m(x, y)| < \delta^2\epsilon$.

Now consider $X \subseteq \mathbb{R}^2$ compact. Then there exists some $k > 0$ such that for all $(x, y) \in X$ we have $|x| < k$ and $|y| < k$. Let $\epsilon > 0$ be arbitrary and $\delta > 0$ be such that for $|x| < \delta$ and $|y| < \delta$ we have $|xy - m(x, y)| < \frac{\epsilon}{k^2}$. Set $\lambda = \frac{\delta}{k}$, so for any $(x, y) \in X$ we have $|\lambda x| < \delta$ and $|\lambda y| < \delta$. Then

$$|xy - \lambda^{-2} m(\lambda x, \lambda y)| < \frac{\epsilon\delta^2}{\lambda^2 M^2} = \epsilon.$$

The neural network M in Figure 1 is such that $M(x, y) = \lambda^{-2} m(x, y)$. □

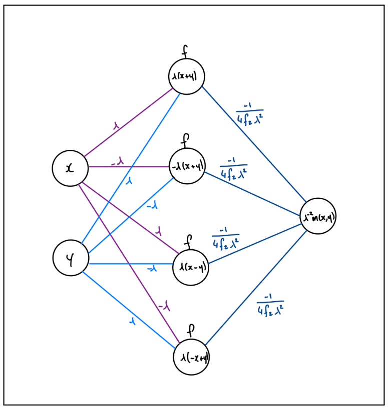

Figure 1: The neural network M.

The size of the approximating neural network Let P be a polynomial approximating $F : \mathbb{R}^d \to \mathbb{R}$, and N a neural network approximating P. We can construct N to have width $d + 1 + 4$. In each layer we have a copy of each input, four neurons for the multiplication currently being performed, plus one neuron to accumulate the value of the polynomial. Once each term has been computed it is added to this neuron.

References

[1] H. W. Lin, M. Tegmark, and D. Rolnick, "Why does deep and cheap learning work so well?" *Journal of Statistical Physics*, vol. 168, no. 6, pp. 1223–1247, Jul. 2017, Publisher: Springer Science and Business Media LLC, ISSN: 1572-9613. DOI: 10.1007/s10955-017-1836-5. [Online]. Available: http://dx.doi.org/10.1007/s10955-017-1836-5.