Abstract Algebra

Reading Project: Summer of Science
Indian Institute of Technology Bombay

Swayam Chube
Mentor: Eeshan Jain

Last Compiled: May 5, 2022

Contents

Chapter 1

Universal Approximation and Barron's Theorem

1.1 Universal Approximation

A *deep network* may be represented by the following function:

$$f(x; w) = \sigma_L(w_L \sigma_{L-1}(\ldots \sigma_1(w_1 x + b_1)\ldots))$$

where each σ_i is a vectorized *non-linear activation function.* Some examples are the Rectified Linear Unit (ReLU) activation function given by $\max\{0, x\}$ and the Sigmoid function given by $\frac{1}{1+e^{-x}}$.

We define the *depth* as L and the *width* as $\max_i d_i$.

We shall show in this section that *all continuous functions* can be expressed with deep networks to sufficient accuracy.

Definition 1.1 (Universal Approximator). A class of functions $\mathcal{H}$ is a *universal approximator* if for any continuous function g and compact domain D and any $\varepsilon > 0$, there is $f \in \mathcal{H}$ such that

$$|f(x) - g(x)| \leq \varepsilon \qquad \forall x \in D$$

The following theorem helps characterize universal approximators:

Theorem 1.2. Suppose $\mathcal{H}$ satisfies the following propositions:

1. each $f \in H$ is continuous over D
2. for all $x \in D$, there is $f \in \mathcal{H}$ with $f(x) \neq 0$
3. for all $x \neq x' \in D$, there is $f \in \mathcal{H}$ with $f(x) \neq f(x')$
4. H is closed under $\times$ and vector space operations

then $\mathcal{H}$ is a universal approximator.

A major theorem related to this is due to Hornik et al.

Theorem 1.3 (Hornik et al. 1989). For any continuous activation function $\sigma\colon \mathbb{R} \to \mathbb{R}$ with

$$\lim_{z\to-\infty} \sigma(z) = 0 \qquad \lim_{z\to\infty} \sigma(z) = 1$$

the depth two feedforward netwroks of unbounded width are universal approximators.

Theorem 1.4 (Park et al.). For any square integrable function $f : \mathbb{R}^d \to \mathbb{R}$ and any $\varepsilon > 0$, there is a width $d + 1$ network with ReLU activations such that

$$\int |f(x) - \hat{f}(x)|^2 \, dx \leq \varepsilon$$

1.2 Barron's Theorem