The Stone-Weierstass Theorem and Its Applications

Vinamra Dhoot
vinamradhoot2009@gmail.com

July 11, 2025

Overview

Introduction

Marshall H. Stone

Karl Weierstrass

Weierstrass Approximation Theorem

Theorem 2.1

Let $f \in C([a, b], \mathbb{R})$ be continuous. Then, there is a sequence of polynomials $p_n(x)$ that converge uniformly to f on $[a, b]$. So, for any $\epsilon > 0$, there exists a polynomial P such that

$$|P(x) - f(x)| < \epsilon, \text{ for all } x \in [a, b].$$

What this means is that for any continuous real-valued function f in the closed interval $[a, b]$, the function can be approximated using polynomials p_n arbitrarily well.

The theorem also makes use of the concept called density. Here is how we define it formally.

Definition 2.2

Density

Subset $A \in \mathcal{T}$ is dense in $\mathcal{T}$ if every point in $\mathcal{T}$ belongs to A or is arbitrarily close to A.

Note: $\mathcal{T}$ is a topology and will be defined later.

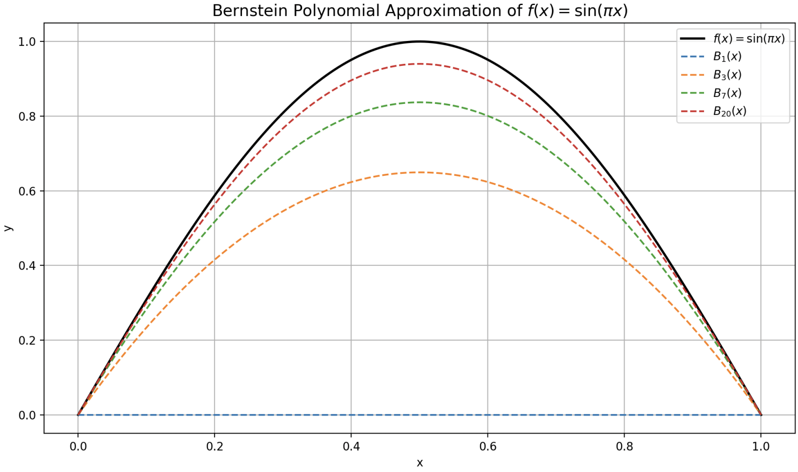
Bernstein polynomials

- Sergei Bernstein provided a constructive and explicit proof of the Theorem.
- He created the Bernstein polynomials which are defined as:

$$B_n(f)(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k}, \text{ for all } x \in [0,1].$$

- We represent $P(x)$ as $B_n(f)(x)$. So, $|B_n(f)(x) - f(x)| < \epsilon$.

Visual Proof of the Weierstrass Approximation Theorem



Definitions

Definition 3.1

Metric Space

A metric space is a set S with a function $d : S \times S \to \mathbb{R}^+$ if any of the following conditions are met:

1. $d(a, b) \geq 0$. $d(a, b) = 0 \Leftrightarrow a = b$.
2. $d(a, b) = d(b, a)$.
3. *The Triangle Inequality:* $d(a, c) \leq d(a, b) + d(b, c)$.

Definition 3.2

A topology, denoted as $\mathcal{T}$, over a set X is a collection of subsets of X such that:

1. $\varnothing$ *and X are in $\mathcal{T}$.*
2. *The intersection of any finite number of elements in $\mathcal{T}$* is in $\mathcal{T}$.
3. *Arbitrary unions of elements are in $\mathcal{T}$.*

Definitions

Definition 3.3

Hausdorff Space

Consider two points $x, y \in X$, where X is a topological space. The two points have a set for example, U of x and V of y, such that they are disjoint ($U \cap V = \varnothing$).

Definition 3.4

Compact space

Compactness generalizes the idea of a closed and bounded subset of a Euclidean space. Similarly, a compact space is one that includes all of the limiting values or points, so it is closed and bounded.

The Stone-Weierstrass Theorem

Stone generalized the Weierstrass Approximation Theorem. It is known as a generalization because both of them are interlinked through the density property.

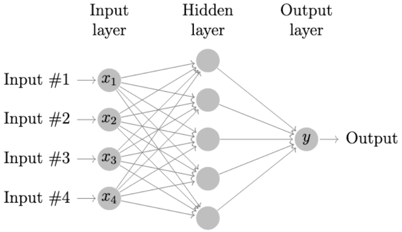
Theorem 3.5

Let X be a compact metric space, and let $\mathcal{A} \subseteq C(X, \mathbb{R})$ be a sub-algebra which separates points of X. Then $\mathcal{A}$ is dense in $C(X, \mathbb{R})$.

Note: All metric spaces here are Hausdorff spaces.

Neural Networks

- A neural network is a model inspired by the structure and function of the human brain.
- We divide a neural network into three parts: an input layer, a hidden layer, and an output layer.



Shallow feedforward network

- A shallow feedforward network (a single hidden layer) can be defined as:

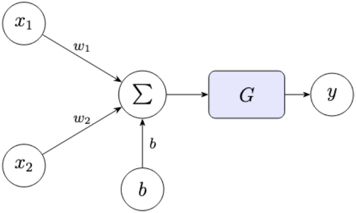
$$\mathcal{F} := \left\{ f : R^n \to R \mid f(x) = \sum_{j=1}^{N} \alpha_j \cdot G(w_j^T x + b_j) \right\}$$

where:

- $x \in R^n$ are the input vectors,
- Each $G(w_j^T x + b_j)$ is a neuron with weights $w_j \in R^n$ and bias $b_j \in R^n$, and activation function G,
- $\alpha \in R^n$ is the output vector.

Neural Network (diagram)

- We can therefore, draw the diagram of a neural network using the following information.



Universal Approximation Theorem

- A really crucial part theorem in Neural Networks is the Unviersal Approximation theorem which makes use of the Stone-Weierstrass theorem.

Theorem 4.1

Let $\sigma : R \to R$ be a nonconstant, bounded, and continuous function. Let $C(K)$ be the space of real-valued continuous functions on a compact subset $K \subset \mathbb{R}^n$. Then for every, function $f \in C(K)$ and $\epsilon > 0$, there exists a neural network function:

$$f_N(x) = \sum_{j=1}^{N} \alpha_j \cdot \sigma(w_j^T x + b_j)$$

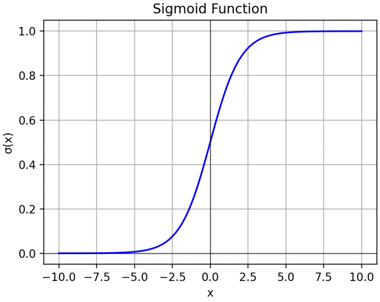
such that,

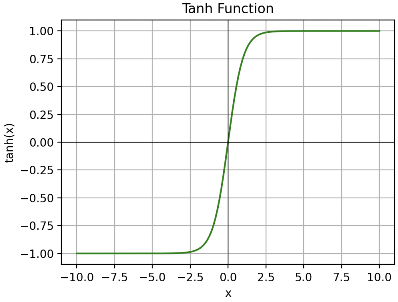
$$\sup_{x \in K} ||f(x) - f_N(x)|| < \epsilon.$$

- The theorem is used in deep learning. It can learn to perform complex tasks like, Natural Language Processing and image recognition.

Activation functions

- In the Stone-Weierstrass interpretation of the Universal Approximation Theorem, we only make use of activation functions like, sigmoid and tanh functions.
- We don't use other types of functions like, ReLU because they are unbounded.





How do we use the Stone-Weierstrass theorem

We want to show that $\mathcal{F}$ meets the criteria of the Stone-Weierstrass Theorem.

- Firstly, we show that it separates points. Firstly, if we have $x \neq y$, we can choose a w and b such that:

  $$w^T x + b \neq w^T y + b.$$

  Since G is injective, we get that

  $$G(w^T x + b) \neq G(w^T y + b).$$

- Secondly, we show that $\mathcal{F}$ contains constants. We show this by taking

  $$w = 0 \in \mathbb{R}^n,$$

  which allows us to show that

  $$f(x) = \alpha \cdot G(b),$$

  which is a constant.

Conclusion

- Weierstrass Approximation Theorem tells us that any continuous function on $[a, b]$ can be uniformly approximated by polynomials.
- The Stone-Weierstrass theorem generalizes this idea to algebras of continuous functions on compact Hausdorff spaces.
- Provides the foundation for the Universal Approximation Theorem in neural networks.
- The Universal Approximation Theorem serves as a crucial theorem to understand neural networks and proves that even a single hidden layer is enough to approximate continuous functions.