

# Term Paper Report

on

## A Mathematical Study of Neural Networks for Function Approximation and Data Fitting

Submitted to



Amity University Uttar Pradesh  
in partial fulfillment of the requirements for the award of the degree of  
B.Sc. (Mathematics) (Honours/Honours with Research)  
By

**Yugam Sehgal**  
**A044154423003**

*Under the Supervision of Internal Guide:*

**Dr. Jayanti Tripathi Pandey**

Assistant Professor III

Department of Mathematics

Amity Institute of Applied Sciences

Amity University Noida, Uttar Pradesh

**Batch: 2024-2028**

**Date submitted: June 2, 2026**



AMITY UNIVERSITY  
— UTTAR PRADESH —

## Amity Institute of Applied Sciences

### Declaration

I, **Yugam Sehgal**, student of **B.Sc. (Mathematics) (Honours/Honours with Research)** hereby declare that the Term Paper titled “**A Mathematical Study of Neural Networks for Function Approximation and Data Fitting**” which is submitted by me to *Department of Mathematics, Amity Institute of Applied Sciences*, Science and Technology Domain, Amity University Uttar Pradesh, in partial fulfillment of requirement for the award of the degree of **B.Sc. (Mathematics) (Honours/Honours with Research)**, has not been previously formed the basis for the award of any degree, diploma or other similar title or recognition.

Noida,  
**Date:** June 2, 2026

---

**Signature of Student**



# AMITY UNIVERSITY

— UTTAR PRADESH —

## Amity Institute of Applied Sciences

### Certificate

On the basis of declaration submitted by **Yugam Sehgal**, student of **B.Sc. (Mathematics) (Honours/Honours with Research)**, I hereby certify that the Term Paper Report titled “**A Mathematical Study of Neural Networks for Function Approximation and Data Fitting**” which is submitted to *Department of Mathematics, Amity Institute of Applied Sciences*, Science and Technology Domain, Amity University Uttar Pradesh, in partial fulfillment of the requirement for the degree of **B.Sc. (Mathematics) (Honours/Honours with Research)**, is an original contribution with existing knowledge and faithful record of work carried out by him under my guidance and supervision.

To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Noida,  
**Date:** June 2, 2026

---

**Dr. Jayanti Tripathi Pandey**  
Assistant Professor III  
Department of Mathematics  
Amity Institute of Applied Sciences  
Amity University Noida, Uttar Pradesh



**AMITY UNIVERSITY**  
— UTTAR PRADESH —

## **Amity Institute of Applied Sciences**

### **Plagiarism Certificate**

This is to certify that the Term Paper report entitled “**A Mathematical Study of Neural Networks for Function Approximation and Data Fitting**” submitted by **Yugam Sehgal** for the partial fulfilment of the degree of **B.Sc. (Mathematics) (Honours/Honours with Research)** has been checked by Turnitin software for plagiarism. The Term Paper report has 5 percent plagiarism.

---

**Signature of Faculty Guide**



**AMITY UNIVERSITY**  
— UTTAR PRADESH —

## **Amity Institute of Applied Sciences**

### **Acknowledgment**

It is high privilege for me to express my deep sense of gratitude to those entire faculty members who helped me in the completion of the “**A Mathematical Study of Neural Networks for Function Approximation and Data Fitting**” under supervision of my internal guide **Dr. Jayanti Tripathi Pandey**.

My special thanks to all other faculty members, batchmates and seniors of **Amity Institute of Applied Sciences**, Amity University Uttar Pradesh for helping me in the completion of project work and its report submission.

**Student Name:** Yugam Sehgal

**Enrollment Number:** A044154423003

## **Abstract**

This paper bridges classical approximation theory with modern deep learning by providing a rigorous mathematical study of neural networks through the dual lenses of function approximation and data fitting.

# Table of contents

<b>Declaration</b>	<b>ii</b>
<b>Certificate</b>	<b>iii</b>
<b>Plagiagrism Certificate</b>	<b>iv</b>
<b>Acknowledgment</b>	<b>v</b>
<b>Abstract</b>	<b>vi</b>
<b>Table of contents</b>	<b>vii</b>
<b>1 Introduction: The Function Approximation Problem</b>	<b>1</b>
<b>2 Expressivity and the Universal Approximation Theorem</b>	<b>3</b>
2.1 Arbitrary Width and the Hahn-Banach Application . . . . .	3
2.2 Depth and Bounded Width . . . . .	5

# Chapter 1

## Introduction: The Function Approximation Problem

At its core, supervised machine learning is a rigorous exercise in functional analysis and statistical approximation. We define the input space as a compact subset  $X \subset \mathbb{R}^d$  and the output space as  $Y \subset \mathbb{R}$ . We assume that the observational data  $(x, y)$  are drawn i.i.d. from an unknown, underlying joint probability distribution  $\mathbb{P}$  supported on  $X \times Y$ .

The fundamental objective is to identify a measurable mapping  $f : X \rightarrow Y$  that minimizes the expected risk, defined as:

$$\mathcal{R}(f) = \mathbb{E}_{(x,y) \sim \mathbb{P}}[\ell(f(x), y)]$$

where  $\ell : Y \times Y \rightarrow \mathbb{R}^+$  is a strictly convex loss function, such as the squared error  $\ell(y', y) = (y' - y)^2$ . Because the true probability measure  $\mathbb{P}$  is fundamentally inaccessible, we are restricted to a finite dataset  $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$ . Consequently, the theoretical problem shifts to Empirical Risk Minimization (ERM), where we minimize the empirical risk:

$$\hat{\mathcal{R}}_n(f_\theta) = \frac{1}{n} \sum_{i=1}^n \ell(f_\theta(x_i), y_i)$$

Unlike classical non-parametric methods (such as smoothing splines or orthogonal Fourier bases) which project the target function onto a fixed, predetermined basis, deep learning models learn the basis dynamically. We define  $f_\theta$  as belonging to a

parameterized hypothesis space  $\mathcal{H}$ , where  $\theta \in \mathbb{R}^W$  represents the weights and biases of a neural network architecture. The central mathematical inquiries therefore become two-fold: first, understanding the topological richness of  $\mathcal{H}$  (expressivity), and second, analyzing how gradient-based methods navigate the non-convex landscape of  $\hat{\mathcal{R}}_n$  to find optimal parameters (data fitting).

## Chapter 2

# Expressivity and the Universal Approximation Theorem

Before interrogating how a neural network learns, one must establish what it is mathematically capable of representing. This is the domain of approximation theory. The primary objective is to determine if the hypothesis class  $\mathcal{H}$  is dense in the target function space.

### 2.1 Arbitrary Width and the Hahn-Banach Application

The historical foundation of neural network expressivity is the Universal Approximation Theorem (UAT). Let  $C(X)$  denote the Banach space of continuous functions on a compact domain  $X \subset \mathbb{R}^d$ , equipped with the uniform (supremum) norm:

$$\|f\|_\infty = \sup_{x \in X} |f(x)|$$

A standard two-layer (shallow) feedforward neural network with a single scalar output evaluates to:

$$f_N(x) = \sum_{i=1}^N v_i \sigma(w_i^T x + b_i)$$

where  $w_i \in \mathbb{R}^d$ ,  $b_i, v_i \in \mathbb{R}$ , and  $\sigma : \mathbb{R} \rightarrow \mathbb{R}$  is a non-linear activation function. Let  $\mathcal{M}(\sigma)$  be the linear span of all such possible parameterized functions:

$$\mathcal{M}(\sigma) = \text{span}\{x \mapsto \sigma(w^T x + b) : w \in \mathbb{R}^d, b \in \mathbb{R}\}$$

The foundational theorems by Cybenko (1989) and Hornik (1991) state that if  $\sigma$  is a continuous sigmoidal function, then  $\mathcal{M}(\sigma)$  is dense in  $C(X)$ . This guarantees that for any target function  $g \in C(X)$  and any  $\epsilon > 0$ , there exists an  $f \in \mathcal{M}(\sigma)$  such that  $\|f - g\|_\infty < \epsilon$ .

**Derivation Sketch (via Functional Analysis):**

The mathematical rigor of this proof relies heavily on duality. The proof proceeds by contradiction. Suppose the closure  $\overline{\mathcal{M}(\sigma)}$  is a proper subspace of  $C(X)$ . By the Hahn-Banach Theorem, there must exist a bounded, continuous linear functional  $L$  on  $C(X)$ , which is strictly non-zero ( $L \neq 0$ ), but annihilates the subspace, meaning  $L(f) = 0$  for all  $f \in \mathcal{M}(\sigma)$ .

By the Riesz Representation Theorem, the continuous dual space of  $C(X)$  is isometrically isomorphic to the space of finite, signed, regular Borel measures on  $X$ , denoted  $M(X)$ . Therefore, the functional  $L$  can be represented by integration against a specific non-zero measure  $\mu \in M(X)$ :

$$L(f) = \int_X f(x) d\mu(x) = 0 \quad \forall f \in \mathcal{M}(\sigma)$$

Substituting the fundamental basis functions of our neural network, we obtain:

$$\int_X \sigma(w^T x + b) d\mu(x) = 0 \quad \text{for all } w \in \mathbb{R}^d \text{ and } b \in \mathbb{R}$$

Through the application of the Fourier transform and analyzing the characteristic function of the measure, it can be mathematically demonstrated that for a discriminatory function  $\sigma$ , this integral condition implies that  $\mu$  must be the zero measure ( $\mu = 0$ ). This directly contradicts our Hahn-Banach deduction that  $L \neq 0$ . Therefore, the initial assumption must be false; the closure  $\overline{\mathcal{M}(\sigma)}$  must be the entirety of  $C(X)$ .

It is worth noting the subsequent generalization by Leshno et al. (1993), which proved that the necessary and sufficient condition for  $\sigma$  to be a universal approximator is simply that it is *not a polynomial*. If  $\sigma$  were a polynomial of degree  $k$ ,  $\mathcal{M}(\sigma)$  would be restricted to the finite-dimensional space of polynomials of maximum degree  $k$ , which cannot be

dense in  $C(X)$ .

## 2.2 Depth and Bounded Width

While the classical UAT establishes the expressivity of arbitrarily wide networks, it requires the number of neurons  $N \rightarrow \infty$ . Modern deep learning architectures, however, favor depth over width.

Consider a deep feedforward network where the layer widths are bounded by  $d_{max}$ . The network evaluates as a composition of affine transformations and point-wise nonlinearities:

$$f_{\theta}(x) = W_L \sigma(W_{L-1} \sigma(\dots \sigma(W_1 x + b_1) \dots) + b_{L-1}) + b_L$$

Recent mathematical extensions of the UAT analyze this deep regime. It has been proven that bounded-width networks achieve universality provided their depth  $L$  is allowed to grow arbitrarily large. For instance, a network utilizing the ReLU activation function,  $\sigma(x) = \max(0, x)$ , requires a minimum width of precisely  $d + 4$  (where  $d$  is the dimension of the input space  $X$ ) to uniformly approximate any Lebesgue integrable function in  $L^1(\mathbb{R}^d)$ .

Topologically, a deep ReLU network operates by executing a sequence of piecewise-linear homeomorphisms, recursively folding and partitioning the input space. This shift in the mathematical paradigm—from infinite-dimensional linear spans to deep functional composition—is critical. It sets the theoretical stage for understanding how networks overcome the “curse of dimensionality.”