Lecture 7: Approximation by Neural Networks

Pr. Ismail Merabet

Univ. of K-M-Ouargla

December 8, 2024

Contents

Introduction

The approximation of continuous functions is a cornerstone of mathematical analysis and computational science, with applications spanning from numerical modeling of many real problems.
Neural networks (NNs) have emerged as a powerful tool for approximating continuous functions, thanks to their flexibility, generalization capabilities, and universal approximation properties.
A foundational result is: the Universal Approximation Theorem[1], which states that a feedforward neural network with:

- a single hidden layer,
- equipped with a non-linear activation function [2],

can approximate any continuous function on a compact domain to arbitrary precision, provided the network has a sufficient number of neurons.

[1] Cybenko 89
[2] such as the sigmoid or ReLU

What is an artificial neural networks?

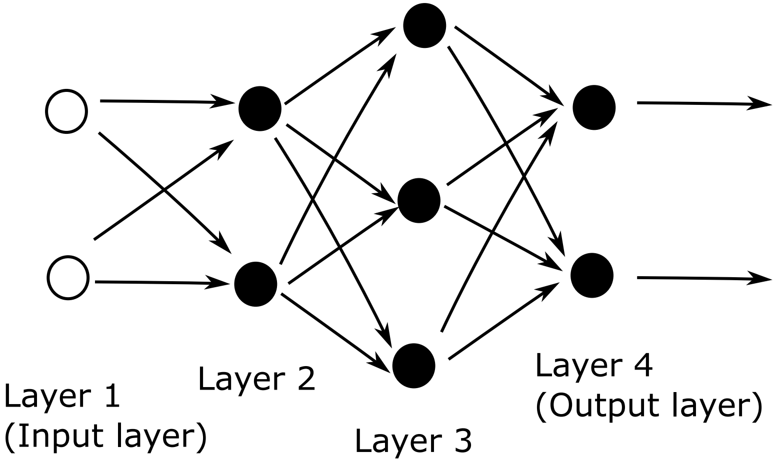

Figure: Artificial Neural Networks

What is an artificial neural networks?

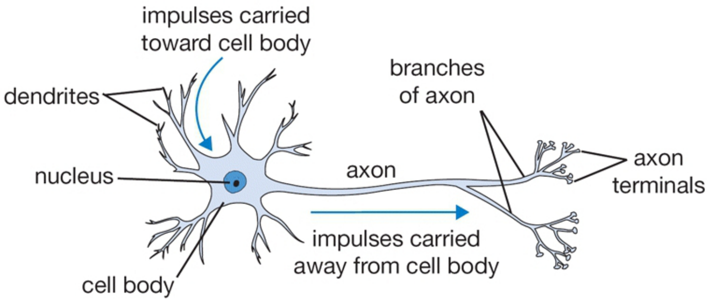

Figure: Natural Neuron

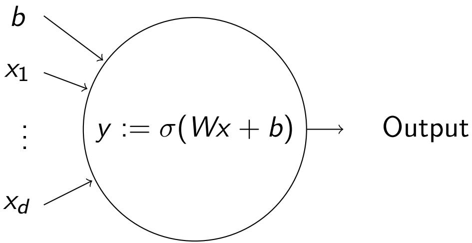

Figure: An artificial neuron

Definition

σ is called the activation function.

Decision Making: Hiring based on experience

σ is the Heaviside function

$$H(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases} \tag{1}$$

Here $x =$ Experience (years): The condition: if the candidate has more than 3 years of experience then say yes.

$$H(x-3) = \begin{cases} 0 & \text{if } x \leq 3 \quad \text{(No)} \\ 1 & \text{if } x > 3 \quad \text{(Yes)} \end{cases} \tag{2}$$

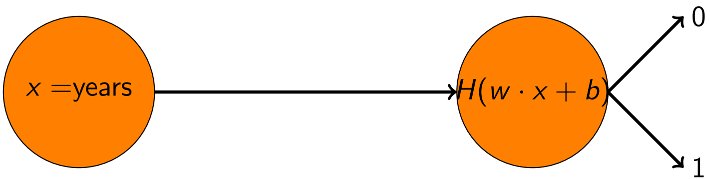

Figure: artificial neural network

Hiring based on experience and education level

Now x_1= experience and x_2= education level.

$$\sigma(w \cdot x + b), \quad x = (x_1, x_2), \quad w = (w_1, w_2)$$

Table:

Candidate	years	degree	score	decision
1	3	Phd (4)	6	Yes
2	6	Master (2)	5	Yes
3	4	Bachelor (1)	1	Yes
4	1	Master (2)	0	No

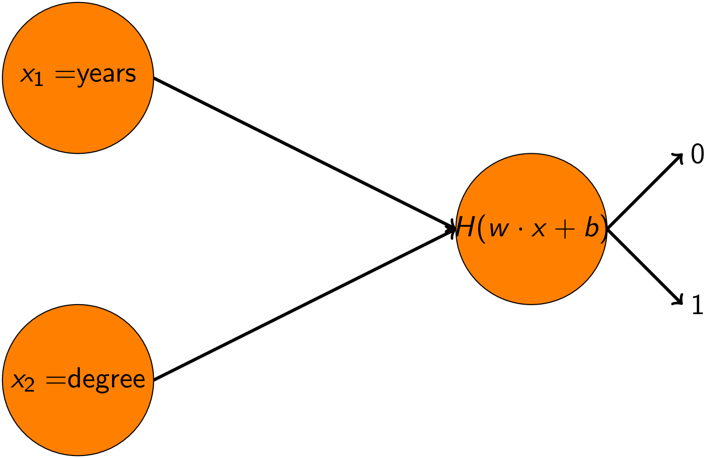

Figure: artificial neural network. $(w_1, w_2, b) = (1, 2, -5)$

Hiring with salary based on experience and education level

$$ReLU(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases} \tag{3}$$

Table:

Candidate	years	degree	salary
1	3	Phd (4)	120000
2	6	Master (2)	10000
3	4	Bachelor (1)	20000
4	1	Master (2)	0

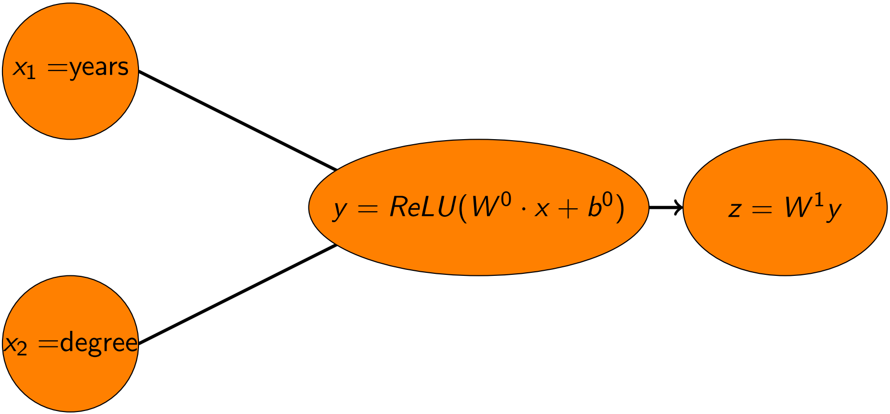

Figure: $\sigma = ReLU, \quad W^0 = (1,2), \quad b^0 = -5, \quad W^1 = 20000$

Some functional analysis theorems

Theorem. Han-Banach : Normed Spaces Version

Let X be a normed vector space, and $Y \subseteq X$ a subspace. If $f : Y \to \mathbb{R}$ is a bounded linear functional, then there exists an extension $F : X \to \mathbb{R}$ such that F is also a bounded linear functional and

$$\|F\| = \|f\|.$$

Theorem. Riesz Representation theorem

Let H be a Hilbert space, and let F be a bounded linear functional on H. Then there exists a unique vector $u \in H$ such that

$$F(v) = (u, v), \quad \forall v \in H,$$

where $(\cdot, \cdot)$ is the inner product on H.

Let I_n denote the n-dimensional unit cube, $[0,1]^n$.

- The space of continuous functions on I_n is denoted by $C(I_n)$ and we use $\|f\|$ to denote the supremum (or uniform) norm of an $f \in C(I_n)$.
- The space of finite, signed regular Borel measures on I_n is denoted by $M(I_n)$.

The main goal of this section is to investigate conditions under which sums of the form

$$G(x) = \sum_{j=1}^{N} \alpha_j \sigma\left(w_j^{\mathrm{T}} x + b_j\right) \tag{4}$$

are dense in $C(I_n)$ with respect to the supremum norm.

Definition

We say that σ is discriminatory if for a measure $\mu \in M(I_n)$

$$\int_{I_n} \sigma\left(w^{\mathrm{T}} x + b\right) d\mu(x) = 0 \tag{5}$$

for all $w \in \mathbb{R}^n$ and $b \in \mathbb{R}$ implies that $\mu = 0$.

Definition

We say that σ is sigmoidal if

$$\sigma(x) \to \begin{cases} 1 & \text{as} \quad x \to +\infty \\ 0 & \text{as} \quad x \to -\infty \end{cases} \tag{6}$$

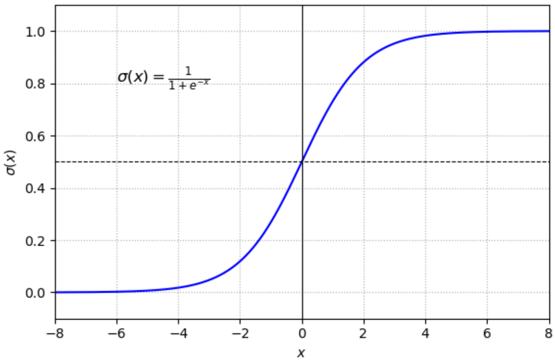

Figure: Example of a sigmoidal function

Theorem. (Cybenko)

Let σ be any continuous discriminatory function. Then finite sums of the form

$$G(x) = \sum_{j=1}^{N} \alpha_j \sigma \left(w_j^{\mathrm{T}} x + b_j \right) \tag{7}$$

are dense in $C(I_n)$.

In other words, given any $f \in C(I_n)$ and $\varepsilon > 0$, there is a sum, $G(x)$, of the above form, for which

$$|G(x) - f(x)| < \varepsilon \quad \text{for all} \quad x \in I_n. \tag{8}$$

Remark

Note that this theorem is the analogue of the Weirstrass theorem seen in Lecture 1 of this cours.

Proof.

Let $S \subset C(I_n)$ be the set of functions of the form $G(x)$ as in (7).
Clearly S is a linear subspace of $C(I_n)$.

> We claim that the closure of S is all of $C(I_n)$.

Assume that $\bar{S}$ the closure of S is not all of $C(I_n)$.
Then $\bar{S}$, is a closed proper subspace of $C(I_n)$.
By the Hahn-Banach theorem, there is a bounded linear functional on $C(I_n)$, call it L, with the property that :

$$L \neq 0 \text{ but } L(\bar{S}) = L(S) = 0.$$

By the Riesz Representation Theorem, this bounded linear functional, L, is of the form

$$L(h) = \int_{I_n} h(x) d\mu(x) \tag{9}$$

for some $\mu \in M(I_n)$, for all $h \in C(I_n)$, where $M(I_n)$ is the space of finite, signed regular Borel measures on I_n.

In particular, since $\sigma\left(w^{\mathrm{T}} x + b\right)$ is in $\bar{S}$ for all w and b, we must have that

$$\int_{I_n} \sigma\left(w^{\mathrm{T}} x + b\right) d\mu(x) = 0 \tag{10}$$

for all w and b.

However, we assumed that σ was discriminatory so that this condition implies that $\mu = 0$ contradicting our assumption. Hence, the subspace S must be dense in $C\left(I_n\right)$.
This demonstrates that sums of the form

$$G(x) = \sum_{j=1}^{N} \alpha_j \sigma\left(w_j^\top x + b_j\right) \tag{11}$$

are dense in $C\left(I_n\right)$ providing that σ is continuous and discriminatory.

However, we assumed that σ was discriminatory so that this condition implies that $\mu = 0$ contradicting our assumption. Hence, the subspace S must be dense in $C(I_n)$.

This demonstrates that sums of the form

$$G(x) = \sum_{j=1}^{N} \alpha_j \sigma\left(w_j^\top x + b_j\right) \tag{11}$$

are dense in $C(I_n)$ providing that σ is continuous and discriminatory.

Lemma

Any bounded, measurable sigmoidal function, σ, is discriminatory. In particular, any continuous sigmoidal function is discriminatory.

Error estimate

In this section we consider the degree of approximation of continuous functions by superpositions of a sigmoidal function. We compare these results with the classical Jackson and Bernstein theorems seen in Lecture 2 of this cours.

Let $I = [0, 1]$, σ be a sigmoidal function and $f \in C(I)$. We introduce the set of functions of the form

$$V_\sigma^n := \left\{ v(x) = \alpha_0 + \sum_{i=1}^{n} \alpha_i \sigma(w_i x + b_i) : \quad w_i, \quad b_i, \quad \alpha_i \in \mathbb{R} \right\} \tag{12}$$

We define the error of approximation of f from V_σ^n by

$$E_{n,\sigma}(f) := \operatorname{dist}(f, V_\sigma^n) = \inf_{g \in V_\sigma^n} \|f - g\|_\infty \tag{13}$$

Theorem

Let σ be a bounded sigmoidal function.

1. If f is a continuous function on $[0, 1]$, then

$$E_{n,\sigma}(f) \leq \|\sigma\| \omega\left(f, \frac{1}{n+1}\right) \tag{14}$$

   where :

   - $\|\sigma\| = \sup\limits_{x \in \mathbb{R}} |\sigma(x)|$
   - $\omega(f, \delta) = \sup\left\{|f(x) - f(y)| : |x - y| \leq \delta, \quad x, y \in [0, 1]\right\}.$

2. If f is continuously differentiable on $[0, 1]$, then

$$\omega\left(f, \frac{1}{n+1}\right) \leq \frac{1}{n+1} \left\|f'\right\|.$$