



Treball Final de  
Grau en Matemàtiques

---

# Machine learning: mathematical foundations

Alicia Chimeno Sarabia

---

Supervisor  
**Roberto Rubio**

Any  
**2022/23**

Convocatòria  
**Juny**

Draft – v0

# Draft – v0

*To my colleagues,*

...

# Abstract

Exemple d'abstract no definitiu, x ficar algo.

Abstract – In various studies, researchers have described the activation functions that allow multilayer feedforward networks to act as universal approximators. However, we demonstrate that most of the characterizations presented in the literature are specific instances of the following general finding: A typical multilayer feedforward network that employs a locally bounded piecewise continuous activation function can accurately approximate any continuous function to an arbitrary degree if and only if the activation function is non-polynomial.

# Resum

blslalblallb en català

# Preface

Inpirat per ?.  
blablalbla amb glosary [Universitat Autònoma de Barcelona \(UAB\)](#) i ara curt [UAB](#).

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Glossary</b>	<b>iii</b>
<b>Preface</b>	<b>iv</b>
<b>Contents</b>	<b>v</b>
<b>1 Introduction</b>	<b>2</b>
<b>2 Artificial Intelligence</b>	<b>3</b>
2.1 What is Artificial Intelligence . . . . .	3
2.2 What is Machine Learning . . . . .	3
2.3 Types of Learning . . . . .	3
2.3.1 Supervised Learning . . . . .	3
2.3.2 Reinforcement Learning . . . . .	3
2.3.3 Unsupervised Learning . . . . .	3
2.4 Output . . . . .	3
2.4.1 Classification Problem . . . . .	3
2.4.2 Regression Problem . . . . .	3
<b>3 Multilayer Feedforward Networks</b>	<b>4</b>
3.1 Lebesgue measure . . . . .	4
3.2 Neural Networks . . . . .	5
3.3 Theorem . . . . .	6
<b>4 Lemmas and proof</b>	<b>7</b>
4.1 Part 1 . . . . .	7
4.2 Part 2 . . . . .	8
<b>5 Results</b>	<b>12</b>
<b>6 Conclusions</b>	<b>13</b>
6.1 Summary . . . . .	13
6.2 Outlook and Future Work . . . . .	13
<b>7 References</b>	<b>14</b>



# Chapter 1

## Introduction

ficar una quote que quedi bé

— John S. Bell, *Against Measurement*

Early computers were used to perform exact computations with high accuracy and efficiency. Back in 1945, one of the first electronic computer was invented for ballistic calculations during World War II. Computers seemed to be limited to these exact computation tasks. However, over time, researchers started pushing the boundaries of what computers can do, eventually leading to the development of what we call now Artificial Intelligence. AI seeks to make computers do the sorts of things that minds can do.

Some of these (e.g. reasoning) are normally described as “intelligent.” Others (e.g. vision) aren’t. But all involve psychological skills—such as perception, association, prediction, planning, motor control—that enable humans and animals to attain their goals.

3. on les mathematiques prenen lloc ? pq son importants . Can we suggest conjectures, relationships , theorems between fields ??? using ml as a tool to see unexpected relationships.

ML might become a bicycle for the mind !!

MATHS USING MCH LEARNING <-> ML USING MATHS



## Chapter 2

# Artificial Intelligence

### 2.1 What is Artificial Intelligence

### 2.2 What is Machine Learning

Machine Learning is the science of programming computers so they can learn from data.

(with the aim to solve a problem without being explicitly programmed.)

For example,

### 2.3 Types of Learning

We can think about learning as the way we understand it as a human. We can learn . That is how we can classify the Machine Learning problem, based on the degree of feedback.

#### 2.3.1 Supervised Learning

#### 2.3.2 Reinforcement Learning

#### 2.3.3 Unsupervised Learning

### 2.4 Output

#### 2.4.1 Classification Problem

#### 2.4.2 Regression Problem

## Chapter 3

# Multilayer Feedforward Networks

### 3.1 Lebesgue measure

An essential part of measure theory is the calculation of lengths, areas, volumes, etc. We are already familiar with the Riemann integral, and now we aim to introduce a more general integral, the Lebesgue integral, that comes from the Lebesgue measure.

**Definition 1.** A box in  $\mathbb{R}^d$  is a set of the form

$$Q = [a_1, b_1] \times \dots \times [a_d, b_d] = \prod_{i=1}^d [a_i, b_i]$$

The volume of the box is

$$\text{vol}(Q) = (b_1 - a_1) \dots (b_d - a_d) = \prod_{i=1}^d (b_i - a_i)$$

The *exterior measure* (or outer measure) of a set  $E \subseteq \mathbb{R}^d$  is

$$|E|^* = \inf \left\{ \sum_k \text{vol}(Q_k) \right\}$$

where the infimum is taken over all finite or countable collection of boxes  $Q_k$  such that  $E \subseteq \cup_k Q_k$

**Definition 2.** A set  $E \subseteq \mathbb{R}^n$  is *Lebesgue measurable* (or measurable) if  $\forall \epsilon > 0$ , there exist  $U \supseteq E$  open set such that  $|U \setminus E|^* < \epsilon$

**Definition 3.** A function  $u$  defined almost everywhere with respect to Lebesgue measure  $v$  on a measurable set  $\Omega \in \mathbb{R}^n$  is said to be *essentially bounded* on  $\Omega$  if  $|u(x)|$  is bounded almost everywhere on  $\Omega$ . We denote  $u \in L^\infty(\Omega)$  with the norm  $\|u\|_{L^\infty(\Omega)} = \inf(\lambda \mid \{x : |u(x)| \geq \lambda\} = \emptyset) = \text{ess sup}_{x \in \Omega} |u(x)|$

We have that  $L^\infty(\mathbb{R})$  is the space of essentially bounded functions.

Examples and counterexamples of functions essentially bounded.

- $f : \Omega \rightarrow$

**Definition 4.** A function  $u$  defined almost everywhere with respect to Lebesgue measure on a domain  $\Omega$  (a domain is an open set in  $\mathbb{R}^n$ ) is said to be *locally essentially bounded* on  $\Omega$  if for every compact set  $K \subset \Omega$ ,  $u \in L^\infty(K)$ . We denote  $u \in L_{loc}^\infty(K)$

**Definition 5.** We say that a set of functions  $F \subset L_{loc}^\infty(\mathbb{R})$  is dense in  $C(\mathbb{R}^n)$  if for every function  $g \in C(\mathbb{R}^n)$  and for every compact  $K \subset \mathbb{R}^n$ , there exist a sequence of functions  $f_j \in F$  such that  $\lim_{j \rightarrow \infty} \|g - f_j\|_{L^\infty(K)} = 0$

**Definition 6.** Let  $M$  denote the set of functions which are in  $L_{loc}^\infty(\mathbb{R})$  and have the following property. The closure of the set of points of discontinuity of any function in  $M$  is of zero Lebesgue measure.

**Proposition 1.** This implies that for any  $\sigma \in M$ , interval  $[a, b]$ . and  $\delta > 0$ , there exists a finite number of open intervals, the union of which we denote by  $U$ , of measure  $\delta$ , such that  $\sigma$  is uniformly continuous on  $[a, b] \setminus U$ .

**Definition 7.** support

## 3.2 Neural Networks

**Definition 8.** (Multilayer feedforward networks) The general architecture of a multilayer feedforward network, MFN, consist of:

- input layer:  $n$ -input units,  $x$
- one/more hidden layers : intermediate processing units
- output layer:  $m$  output-units  $f(x)$

function that a MFN compute is:

$$f(x) = \sum_{j=1}^k \beta_j \cdot \sigma(w_j \cdot x - \theta_j)$$

- $x = (x_1, \dots, x_n)$  input-vector
- $k$ : # of processing-units in the hidden layer
- $w = (w_1, \dots, w_n)$ : weights vector
- $\sigma : \mathbb{R} \rightarrow \mathbb{R}$  activation function
- $\theta$  treshhold value: ???
- $\beta$

We take  $C(\mathbb{R}^n)$  to be the family of real world functions that one may wish to approximate with feedforward network architectures

**Definition 9.**  $\mathcal{C}_0^\infty$  functions  $\mathcal{C}^\infty$  with compact support.

**Definition 10.**  $\varphi : I \rightarrow \mathbb{R}$  is uniformly continuous on  $I$  if  $\forall \epsilon > 0 \exists \delta > 0$  such that  $|\varphi(x) - \varphi(y)| < \epsilon$  whenever  $|x - y| < \delta$

**Definition 11.** Let  $f, g$  be real-valued functions with compact support. We define the *convolution* of  $f$  with  $g$  as

$$(f * g)(x) = \int f(x - t)g(t) dt$$

### 3.3 Theorem

**Theorem 1.** Let  $\sigma \in M$ . Set

$$\Sigma_n = \text{span}\{\sigma(w \cdot x + \theta) : w \in \mathbb{R}^n, \theta \in \mathbb{R}\}$$

Then  $\Sigma_n$  is dense in  $\mathcal{C}(\mathbb{R}^n)$  if and only if  $\sigma$  is not an algebraic polynomial.

Why this theorem does not contradict the Stone-Weierstrass theroem ? (Enunciat al apendix) Com que  $\sigma$  té grau fix "k", aleshores  $\sigma(wx + \theta)$  grau com a molt k. Per tant el conjunt  $\sigma_n$  és un espai vectorial finit i no pot ser dens.

# Chapter 4

## Lemmas and proof

This chapter/section presents the lemmas that are necessary to prove the main result. In order to prove the theorem, I will first demonstrate the inverse implication. Specifically, I will show that " ".

### 4.1 Part 1

**Lemma 1.** If we have that  $\sigma * \varphi$  is a polynomial for all  $\varphi \in \mathcal{C}_0^\infty$ . Then the degree of the polynomial  $\sigma * \varphi$  is finite, i.e. there exists an  $m \in \mathbb{N}$  such that  $\deg(\sigma * \varphi) \leq m$  for all  $\varphi \in \mathcal{C}_0^\infty$ .

*Proof.* We first prove the claim in the case of  $\varphi \in \mathcal{C}_0^\infty[a, b]$ , where  $\mathcal{C}_0^\infty[a, b]$  is the set of functions  $\mathcal{C}_0^\infty$  with support in  $[a, b]$  for any  $a < b$ .

Let  $\rho$  be a metric on  $\mathcal{C}_0^\infty[a, b]$  defined by

$$\rho(\varphi_1, \varphi_2) = \sum_{n=0}^{\infty} 2^{-n} \frac{\|\varphi_1 - \varphi_2\|_n}{1 + \|\varphi_1 - \varphi_2\|_n}$$

where  $\|\varphi\|_n = \sum_{j=0}^n \sup_{x \in [a, b]} |\varphi^{(j)}(x)|$ ,  $\mathcal{C}_0^\infty[a, b]$  is a complete metric space.

By assumption, we have that  $\sigma * \varphi$  is a polynomial (for any  $\varphi \in \mathcal{C}_0^\infty[a, b]$ ).

Consider the following set (has the property that we want to show).

$V_k = \{\varphi \in \mathcal{C}_0^\infty[a, b] \mid \deg(\sigma * \varphi) \leq k\}$ . This set fulfills the following properties,  $V_k \subset V_{k+1}$ ,  $V_k$  is a closed subspace and  $\cup_{k=0}^{\infty} V_k = \mathcal{C}_0^\infty[a, b]$ . As  $\mathcal{C}_0^\infty[a, b]$  is a complete metric space, then there exists an integer  $m$  such that  $V_m = \mathcal{C}_0^\infty[a, b]$ . Proof ??

For the general case where  $\varphi \in \mathcal{C}_0^\infty$ , we note that the number  $m$  does not depend on the interval  $[a, b]$ . Por tanto hemos visto que el conjunto con la propiedad que queremos es = que el que tenemos. we

□

**Lemma 2.** If  $\sigma * \varphi$  is a polynomial such that  $\deg(\sigma * \varphi) \leq m$  for all  $\varphi \in \mathcal{C}_0^\infty$ , then  $\sigma$  is a polynomial of degree at most  $m$ .

*Proof.* For all  $\varphi \in \mathcal{C}_0^\infty$ , we have that

$$\sigma * \varphi^{(m+1)}(x) = \int \sigma(x-t) \varphi^{(m+1)}(t) dt = 0$$

□

Conclusion: If we have that  $\sigma * \varphi$  is a polynomial then  $\sigma$  is a polynomial. This contradicts the hypothesis. Therefore,  $\sigma * \varphi$  will not be a polynomial.

## 4.2 Part 2

**Lemma 3.** For each  $\varphi \in \mathcal{C}_0^\infty$ ,  $\sigma * \varphi \in \overline{\sum_1}$ .

*Proof.* Consider

$$h_m = \sum_{i=1}^m \varphi(y_i) \Delta y_i \sigma(x - y_i)$$

The sequence  $(h_m)$  satisfies  $h_j \in \sum_1$  for  $j = 1, \dots, m$ . ( $w_i = 1, \theta_i = -y_i, \beta_i = \varphi(y_i) \Delta y_i$ ).

Where  $y_i = -\alpha + \frac{2i\alpha}{m}$ ,  $\Delta y_i = \frac{2\alpha}{m}$  for  $i = 1, \dots, m$ . Partition of the interval  $[-\alpha, \alpha]$

We want to show that  $h_m \rightrightarrows \sigma * \varphi$  in  $[-\alpha, \alpha]$ .

Given  $\epsilon > 0$ , we choose  $\delta > 0$  such that  $10\delta \|\sigma\|_{L^\infty\{-2\alpha, 2\alpha\}} \|\varphi\|_{L^\infty} \leq \frac{\epsilon}{3}$ . Note that ...

We know that  $\sigma \in M$ . Hence, for this given  $\delta > 0$  and  $[-\alpha, \alpha]$  interval, there exists  $r(\delta)$  finite number of intervals the measure of whose union  $\mathcal{U}$  is  $\delta$  such that  $\sigma$  is uniformly continuous on  $[-2\alpha, 2\alpha]$ . We now choose  $m_i$  sufficiently large so that

1.  $m_1 \delta > \alpha r(\delta)$ . We can do this by Archimedes' principle.
2. From the uniform continuity of  $\varphi$
3. From the previous,  $\sigma$  is uniformly continuous on  $[-2\alpha, 2\alpha]$ .

We choose  $m$  such that  $m = \max\{m_1, m_2, m_3\}$ .

Now, fix  $x \in [-\alpha, \alpha]$ . Set  $\Delta_i = [y_{i-1}, y_i]$  where  $y_i$  is defined as above.

First, recall that (for the integral is equal to the sum over intervals of the integrals)

$$\int \sigma(x-y) \varphi(y) dy = \sum_{i=1}^m \int_{\Delta_i} \sigma(x-y) \varphi(y) dy$$

Consider the following difference

$$\begin{aligned}
\left| \int \sigma(x-y)\varphi(y)dy - \sum_{i=1}^m \int_{\Delta_i} \sigma(x-y_i)\varphi(y)dy \right| &= \\
&= \left| \sum_{i=1}^m \int_{\Delta_i} \sigma(x-y)\varphi(y)dy - \sum_{i=1}^m \int_{\Delta_i} \sigma(x-y_i)\varphi(y)dy \right| \\
&= \left| \sum_{i=1}^m \int_{\Delta_i} \varphi(y) \left( \sigma(x-y) - \sigma(x-y_i) \right) dy \right| \\
&\leq \sum_{i=1}^m \int_{\Delta_i} |\varphi(y)| |\sigma(x-y) - \sigma(x-y_i)| dy
\end{aligned}$$

If  $x - \Delta_i \cap U = \emptyset$ . Since  $x - y \notin U$ ,  $x - y_i \notin U$  and  $x - y_i \in [-2\alpha, 2\alpha]$ , bc (2) we have

$$\begin{aligned}
\sum_{i=1}^m \int_{\Delta_i} |\varphi(y)| |\sigma(x-y) - \sigma(x-y_i)| dy &\leq \frac{\epsilon}{\|\varphi\|_{L_1}} \sum_{i=1}^m \int_{\Delta_i} |\varphi(y)| dy \\
&= \frac{\epsilon}{3\|\varphi\|_{L_1}} \int |\varphi(y)| dy \\
&= \frac{\epsilon}{3\|\varphi\|_{L_1}} |\varphi(y)|_{L_1} = \frac{\epsilon}{3}
\end{aligned}$$

If  $x - \Delta_i \cap U \neq \emptyset$

$$\sum_i |\widetilde{\Delta_i}| = \sum_i |(x - \Delta_i \cap U)| \leq |U| + 2|\Delta_i|r(\delta) \leq \delta + 2 \cdot \frac{2\alpha}{m}r(\delta) \leq \delta + 4\delta = 5\delta$$

True by our choice of m, satisfies  $m\delta > \alpha r(\delta) \iff \delta > \frac{\alpha r(\delta)}{m}$

$$\begin{aligned}
\sum_{i=1}^m \int_{\widetilde{\Delta_i}} |\varphi(y)| |\sigma(x-y) - \sigma(x-y_i)| dy &\leq \\
&\leq \sum_{i=1}^m \int_{\widetilde{\Delta_i}} \|\varphi\|_{L^\infty} 2\|\sigma\|_{L^\infty[-2\alpha, 2\alpha]} \\
&= \|\varphi\|_{L^\infty} 2\|\sigma\|_{L^\infty[-2\alpha, 2\alpha]} \sum_i |\widetilde{\Delta_i}| \\
&\leq \|\varphi\|_{L^\infty} 2\|\sigma\|_{L^\infty[-2\alpha, 2\alpha]} 5\delta \leq \epsilon/3
\end{aligned}$$

□

**Lemma 4.** If  $\sigma \in \mathcal{C}^\infty$ , then  $\sum_1$  is dense in  $\mathcal{C}(\mathbb{R})$ .

*Proof.* We recall that set  $\sum_1 = \text{span}\{\sigma(w \cdot x + \theta) : w \in \mathbb{R}, \theta \in \mathbb{R}\}$ . We can write any function  $h \in \sum_1$  as  $h = \sum_i \beta_i \sigma_i(w_i x + \theta_i) = \beta_1 \sigma_1(w_1 x + \theta_1) + \dots$

$\frac{\sigma([w+h]x+\theta)-\sigma(wx+\theta)}{h} \in \sum_1$  because is a linear combination, where  $\beta_1 = \frac{1}{h}, \beta_2 = \frac{-1}{h} \dots$ . By hypothesis, we have  $\sigma \in \mathcal{C}^\infty$ . By definition of derivative we have

$$\frac{d}{dw}\sigma(wx+\theta) = \lim_{h \rightarrow 0} \frac{\sigma([w+h]x+\theta) - \sigma(wx+\theta)}{h} \in \overline{\sum_1}^*$$

Because the limit of a set belongs to the closure of the set.

By the same argument,  $\frac{d^k}{dw^k}\sigma(wx+\theta) \in \overline{\sum_1}$  for all  $k \in \mathbb{N}, w, \theta \in \mathbb{R}$ .

We observe that  $\frac{d}{dw}\sigma(wx+\theta) = \sigma'(wx+\theta) \cdot x$ . If we differentiate this expression  $k$  times, we obtain

$$\frac{d^k}{dw^k}\sigma(wx+\theta) = \sigma^{(k)}(wx+\theta) \cdot x^k$$

Since  $\sigma$  is not a polynomial (theorem hypothesis) then there exists a  $\theta_k \in \mathbb{R}$  such that  $\sigma^{(k)}(\theta_k) \neq 0$

Lets see.\*\*\*\* If  $\sigma$  is not a polynomial and  $\sigma \in \mathcal{C}^\infty$ , lets assume that  $\nexists \theta_k \in \mathbb{R}$  such that  $\sigma^{(k)}(\theta_k) \neq 0$ . This means that the  $k$ -th derivative at every point is 0, i.e,  $\sigma^{(k)}(\theta) = 0 \forall \theta \in \mathbb{R}$ . If we integrate  $k$  times,  $\int \sigma^{(k)} = \int 0 \iff \sigma^{(k-1)} = C$ ,  $\int \sigma^{(k-1)} = \int C \iff \sigma^{(k-2)} = Cw$ , then we end up  $\sigma$  is a polynomial. Contradiction. Therefore, there always exists a point where the derivative does not vanish.

Thus, we evaluate at this point  $\theta_k$  where the derivative does not vanish.

$$\sigma^{(k)}(\theta_k) \cdot x^k = \frac{d^k}{dw^k}\sigma(wx+\theta) \Big|_{w=0, \theta=\theta_k} \in \overline{\sum_1}$$

That implies that  $\overline{\sum_1}$  contains all polynomials, because the expression  $\sigma^{(k)}(\theta_k)x^k$  generates all polynomials. By the Weierstrass theorem, it follows that  $\sum_1$  contains...  
falta mirar.  $\square$

**Lemma 5.** If for some  $\varphi \in \mathcal{C}_0^\infty$  we have that  $\sigma * \varphi$  is not a polynomial, then  $\sum_1$  is dense in  $\mathcal{C}(\mathbb{R})$

*Proof.* From Lemma 3,  $\sigma * \varphi \in$   $\square$

**Lemma 6.** If  $\sum_1$  is dense in  $\mathcal{C}(\mathbb{R})$ , then  $\sum_n$  is dense in  $\mathcal{C}(\mathbb{R}^n)$ .

*Proof.* Let  $V := \text{span}\{f(ax) : a \in \mathbb{R}^n, f \in \mathcal{C}(\mathbb{R})\}$ .  $V$  is dense in  $\mathcal{C}(\mathbb{R}^n)$ .

Let  $g \in \mathcal{C}(\mathbb{R})$ , for any compact subset  $K \subset \mathbb{R}^n$ ,  $V$  dense in  $\mathcal{C}(K)$ . That is, given  $\epsilon > 0$ , there exist  $f_i \in \mathcal{C}(\mathbb{R})$  and  $a_i \in \mathbb{R}^n$   $i = 1, \dots, k$  such that  $\square$

*Proof.*

$\Rightarrow$  To prove the implication, we will use proof by contrapositive. We will see the following. If  $\sigma$  is a polynomial then  $\sum_n$  is not dense in  $\mathcal{C}(\mathbb{R}^n)$ . Let  $\sigma$  be a polynomial of degree  $k$ , then  $\sigma(wx+\theta)$  is a polynomial of degree  $k$  for every  $w, \theta$ . We have  $\sum_n = \text{span}\{\sigma(w \cdot x + \theta) : w \in \mathbb{R}^n, \theta \in \mathbb{R}\}$  that is the set of algebraic polynomials of degree at most  $k$ .

---

\* $\overline{\sum_1}$  denotes the clausure of the set  $\sum_1$



$\sum_n$  is not dens in  $\mathcal{C}(\mathbb{R}^n)$  if for a function  $f(x) \in \mathcal{C}(\mathbb{R}^n)$  we can find  $\epsilon > 0$  and  $K$  such that  $\|p - f\| > \epsilon$  for all  $p$  polynomial of degree  $k$ . For example, let  $f(x) = \cos(x)$ , and  $p(x) = \sigma(wx + \theta)$  that has degree  $k$ . This implies has maximum  $k$  roots. We can find a interval where there is  $k+1$  roots.

$\Leftarrow$

□

Leshno et al. [1993]

## Chapter 5

### Results

$$t = x + y \tag{5.1}$$

# Chapter 6

## Conclusions

It is a mistake to confound strangeness with mystery.

— Sherlock Holmes, *A Study in Scarlet*

### 6.1 Summary

### 6.2 Outlook and Future Work

Hem trobat:

- Aaaaaa
- Bbbbbb

## Chapter 7

### References

M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867, 1993.

# Appendix A

## algo

**Definition 12.** Riemann integral reminder. The Riemann integral is a method for calculating the volume under a curve of a continuous function on a closed, bounded domain in  $\mathbb{R}^n$ . The method involves dividing the domain into smaller subregions and approximating the volume of each subregion with a rectangular solid whose height is the function value at a specific point in the subregion. The Riemann sum is the sum of the volumes of all the rectangular solids, and as the size of the subregions approaches zero, the Riemann sum converges to the Riemann integral.

**Definition 13.** Let  $\Sigma$  be a  $\sigma$ -algebra over a set  $\Omega$ . A *measure* over  $\Omega$  is any function

$$\mu : \Sigma \longrightarrow [0, \infty]$$

satisfying the following properties:

1.  $\mu(\emptyset) = 0$ .
2.  $\sigma$ -*additivity*: If  $(A_n) \in \Sigma$  are pairwise disjoint, then:

$$\mu \left( \bigsqcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \mu(A_n)$$

**Definition 14.** A *metric* (or *distance*) on a set  $X$  is a function  $d : X \times X \rightarrow \mathbb{R}$  such that for all  $s, t \in X$  the following properties are satisfied:

1.  $d(s, t) \geq 0$  and  $d(s, t) = 0$  if and only if  $s = t$ .
2.  $d(s, t) = d(t, s)$ .
3.  $d(s, t) \leq d(s, u) + d(u, t)$  (*triangular inequality*).

**Definition 15.** A *metric space* is a pair  $(X, d)$ , where  $X$  is a set and  $d$  is a distance in  $X$ .

**Definition 16.** (Complete metric space).

---

**Definition 17.** We say that a property holds almost everywhere (a.e.) if the set of points that doesn't hold it is null.

**Theorem 2.** (*Weierstrass approximation theorem*) Let  $f : [a, b] \rightarrow \mathbb{R}$  be a continuous function. Then, there exists polynomials  $p_n \in \mathcal{R}[x]$  such that the sequence  $(p_n)$  converge uniformly to  $f$  on  $[a, b]$ .

**Corollary 1.** The set  $\mathcal{R}^n[x]$  is dense in  $\mathcal{C}(\mathbb{R})^n$