

# Introduction to Machine Learning

## Universal Approximation Theorem of the Multilayer Perceptron

Andres Mendez-Vazquez

June 30, 2018

# Outline

## 1 Introduction

- The Representation of Functions

## 2 Basic Definitions

- Topology
- Compactness
- Continuous Functions
- Bounding Continuous Functions
- About Density in a Topology
- Density Concept
- Having a Nice Space
  - Hausdorff Space
- Measures
- The Borel Measure
- Discriminatory Functions
- The Important Theorem
- Universal Representation Theorem



# Outline

## 1 Introduction

- The Representation of Functions

## 2 Basic Definitions

- Topology
- Compactness
- Continuous Functions
- Bounding Continuous Functions
- About Density in a Topology
- Density Concept
- Having a Nice Space
  - Hausdorff Space
- Measures
- The Borel Measure
- Discriminatory Functions
- The Important Theorem
- Universal Representation Theorem



# Introduction

## Representation of functions

The main result in multi-layer perceptron is its power of representation.

### Furthermore

After all, it is quite striking if we can represent continuous functions of the form  $f : \mathbb{R}^n \mapsto \mathbb{R}$  as a finite sum of simple functions.



# Introduction

## Representation of functions

The main result in multi-layer perceptron is its power of representation.

## Furthermore

After all, it is quite striking if we can represent continuous functions of the form  $f : \mathbb{R}^n \mapsto \mathbb{R}$  as a finite sum of simple functions.



# Therefore

## Our main goal

We want to know under which conditions the sum of the form:

$$G(\mathbf{x}) = \sum_{j=1}^N \alpha_j f(\mathbf{w}^T \mathbf{x} + \theta_j) \quad (1)$$

can represent continuous functions in a specific domain.



# Setup of the problem

## Definition of $I_n$

It is an  $n$ -dimensional unit cube  $[0, 1]^n$

In addition, we have the following set of functions

$$\mathcal{C}(I_n) = \{f : I_n \rightarrow \mathbb{R} \mid f \text{ is a continuous function}\} \quad (2)$$



# Setup of the problem

## Definition of $I_n$

It is an  $n$ -dimensional unit cube  $[0, 1]^n$

In addition, we have the following set of functions

$$C(I_n) = \{f : I_n \rightarrow \mathbb{R} \mid f \text{ is a continuous function}\} \quad (2)$$





# Outline

## 1 Introduction

- The Representation of Functions

## 2 Basic Definitions

- **Topology**
  - Compactness
  - Continuous Functions
  - Bounding Continuous Functions
  - About Density in a Topology
  - Density Concept
  - Having a Nice Space
    - Hausdorff Space
  - Measures
  - The Borel Measure
  - Discriminatory Functions
  - The Important Theorem
  - Universal Representation Theorem



## Definition (Topological Space)

A topological space is then a set  $X$  together with a collection of subsets of  $X$ , called **open sets** and satisfying the following axioms:

- 1. The empty set and  $X$  itself are open.
- 2. Any union of open sets is open.
- 3. The intersection of any finite number of open sets is open.



## Definition (Topological Space)

A topological space is then a set  $X$  together with a collection of subsets of  $X$ , called **open sets** and satisfying the following axioms:

- 1 The empty set and  $X$  itself are open.
- 2 Any union of open sets is open.
- 3 The intersection of any finite number of open sets is open.

## Remarks

- This is quite axiomatic... because any set in the collection of  $X$  is open...



## Definition (Topological Space)

A topological space is then a set  $X$  together with a collection of subsets of  $X$ , called **open sets** and satisfying the following axioms:

- 1 The empty set and  $X$  itself are open.
- 2 Any union of open sets is open.
- 3 The intersection of any finite number of open sets is open.

- This is quite axiomatic... because any set in the collection of  $X$  is open...



## Definition (Topological Space)

A topological space is then a set  $X$  together with a collection of subsets of  $X$ , called **open sets** and satisfying the following axioms:

- 1 The empty set and  $X$  itself are open.
- 2 Any union of open sets is open.
- 3 The intersection of any finite number of open sets is open.

- This is quite axiomatic... because any set in the collection of  $X$  is open...



## Definition (Topological Space)

A topological space is then a set  $X$  together with a collection of subsets of  $X$ , called **open sets** and satisfying the following axioms:

- 1 The empty set and  $X$  itself are open.
- 2 Any union of open sets is open.
- 3 The intersection of any finite number of open sets is open.

## Remark

- This is quite axiomatic... because any set in the collection of  $X$  is open...



## Example

### We have

- Given any set  $X$ , one can define a topology on  $X$  where every subset of  $X$  is an open set.

### Also

- Let  $(X, d)$  be a metric space. The sets (called open Balls) are a Topology

$$S(x_0, r) = \{x \in X \mid d(x_0, x) < r\} \text{ where } r > 0 \text{ and } x_0 \in X$$



## Example

### We have

- Given any set  $X$ , one can define a topology on  $X$  where every subset of  $X$  is an open set.

### Also

- Let  $(X, d)$  be a metric space. The sets (called open Balls) are a Topology

$$S(x_0, r) = \{x \in X \mid d(x_0, x) < r\} \text{ where } r > 0 \text{ and } x_0 \in X$$





# Outline

## 1 Introduction

- The Representation of Functions

## 2 Basic Definitions

- Topology
- **Compactness**
- Continuous Functions
- Bounding Continuous Functions
- About Density in a Topology
- Density Concept
- Having a Nice Space
  - Hausdorff Space
- Measures
- The Borel Measure
- Discriminatory Functions
- The Important Theorem
- Universal Representation Theorem



# Definition

## Open Cover Definition

- A topological space  $X$  is called compact if each of its open covers has a finite subcover.

In Our Case:

- $I_n$  is compact



# Definition

## Open Cover Definition

- A topological space  $X$  is called compact if each of its open covers has a finite subcover.

## In Our Case

- $I_n$  is compact



# Compactness

## Theorem

A compact set is closed and bounded.

Proof

$I_n$  is a compact set in  $\mathbb{R}^n$ .



# Compactness

## Theorem

A compact set is closed and bounded.

## Thus

$I_n$  is a compact set in  $\mathbb{R}^n$ .



# Why Compactness?

## Basically

- Given that in Topology we care in how something behaves in open sets!!!

## Compactness

- Establish some sort of "fitness" in a Topological sense

## Therefore

- There are only finitely many possible behaviors.



# Why Compactness?

## Basically

- Given that in Topology we care in how something behaves in open sets!!!

## Compactness

- Establish some sort of “fitness” in a Topological sense

## Therefore

- There are only finitely many possible behaviors.



# Why Compactness?

## Basically

- Given that in Topology we care in how something behaves in open sets!!!

## Compactness

- Establish some sort of “fitness” in a Topological sense

## Therefore

- There are only finitely many possible behaviors.





# Outline

## 1 Introduction

- The Representation of Functions

## 2 Basic Definitions

- Topology
- Compactness
- **Continuous Functions**
- Bounding Continuous Functions
- About Density in a Topology
- Density Concept
- Having a Nice Space
  - Hausdorff Space
- Measures
- The Borel Measure
- Discriminatory Functions
- The Important Theorem
- Universal Representation Theorem



# Continuous Functions

## Theorem

- A function  $f$  from a topological space  $X$  into another topological space  $Y$  is continuous if and only if every open set  $V$  in  $Y$ ,

$$f^{-1}(V) = \{x \mid f(x) \in V\}$$

Example: Is  $f: [0,1] \rightarrow \mathbb{R}, f(x) = x^2$  a continuous function in  $\mathbb{R}$ ?



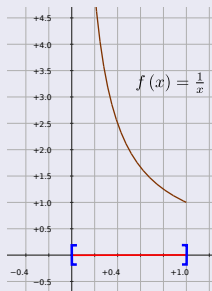
# Continuous Functions

## Theorem

- A function  $f$  from a topological space  $X$  into another topological space  $Y$  is continuous if and only if every open set  $V$  in  $Y$ ,

$$f^{-1}(V) = \{x \mid f(x) \in V\}$$

Example, Is  $f : [0, 1] \rightarrow \mathbb{R}$ ,  $f(x) = 1/x$  a continuous function in  $[0, 1]$ ?



It is not!!

Define with  $g$  a continuous function

$$B_{\mathbb{R}}(\epsilon, g(x_0)) = \{y \in \mathbb{R} \mid \|y - g(x_0)\| < \epsilon\}$$

Therefore, its pre-image is open

$$f^{-1}(B_{\mathbb{R}}(\epsilon, g(x_0)))$$

Therefore exist a ball around  $B_{[0,1]}(\delta, x_0)$

$$B_{[0,1]}(\delta, x_0) \subseteq f^{-1}(B_{\mathbb{R}}(\epsilon, g(x_0)))$$



It is not!!

Define with  $g$  a continuous function

$$B_{\mathbb{R}}(\epsilon, g(x_0)) = \{y \in \mathbb{R} \mid \|y - f(x_0)\| < \epsilon\}$$

Therefore, its pre-image is open

$$f^{-1}(B_{\mathbb{R}}(\epsilon, g(x_0)))$$

Therefore exist a ball around  $(\delta, x_0)$

$$B_{[0,1]}(\delta, x_0) \subseteq f^{-1}(B_{\mathbb{R}}(\epsilon, g(x_0)))$$



It is not!!

Define with  $g$  a continuous function

$$B_{\mathbb{R}}(\epsilon, g(x_0)) = \{y \in \mathbb{R} \mid \|y - f(x_0)\| < \epsilon\}$$

Therefore, its pre-image is open

$$f^{-1}(B_{\mathbb{R}}(\epsilon, g(x_0)))$$

Therefore exist a ball around  $B_{[0,1]}(\delta, x_0)$

$$B_{[0,1]}(\delta, x_0) \subseteq f^{-1}(B_{\mathbb{R}}(\epsilon, g(x_0)))$$



## The well know $\epsilon - \delta$ definition

### What about the $+\infty$ and our original $f$

- You need to use a sequence  $\{x_n\}$  such that  $x_n \rightarrow +\infty$  when  $n \rightarrow \infty$

Therefore, we have for  $\delta > 0$

- We have that  $\lim_{n \rightarrow \infty} B(\epsilon, f(x_n)) = \lim_{n \rightarrow \infty} x_n$

Therefore

$$\lim_{n \rightarrow \infty} f^{-1}(B_{\mathbb{R}}(\epsilon, g(x_n))) = \{0\}$$



## The well know $\epsilon - \delta$ definition

What about the  $+\infty$  and our original  $f$

- You need to use a sequence  $\{x_n\}$  such that  $x_n \rightarrow +\infty$  when  $n \rightarrow \infty$

Therefore, we have for some  $\epsilon >$

- We have that  $\lim_{n \rightarrow \infty} B(\epsilon, f(x_n)) = \lim_{n \rightarrow \infty} x_n$

Therefore

$$\lim_{n \rightarrow \infty} f^{-1}(B_{\mathbb{R}}(\epsilon, g(x_n))) = \{0\}$$





## The well know $\epsilon - \delta$ definition

What about the  $+\infty$  and our original  $f$

- You need to use a sequence  $\{x_n\}$  such that  $x_n \rightarrow +\infty$  when  $n \rightarrow \infty$

Therefore, we have for some  $\epsilon >$

- We have that  $\lim_{n \rightarrow \infty} B(\epsilon, f(x_n)) = \lim_{n \rightarrow \infty} x_n$

Therefore

$$\lim_{n \rightarrow \infty} f^{-1}(B_{\mathbb{R}}(\epsilon, g(x_n))) = \{0\}$$



Therefore

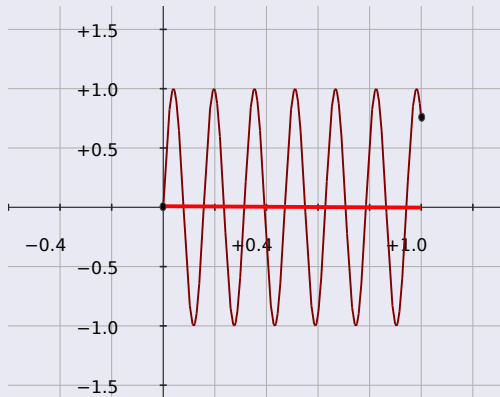
The pre-image is closed

- The function  $f$  is not continuous!!!



# All the continuous functions are bounded

## For Example



# Outline

## 1 Introduction

- The Representation of Functions

## 2 Basic Definitions

- Topology
- Compactness
- Continuous Functions
- **Bounding Continuous Functions**
- About Density in a Topology
- Density Concept
- Having a Nice Space
  - Hausdorff Space
- Measures
- The Borel Measure
- Discriminatory Functions
- The Important Theorem
- Universal Representation Theorem



# Thus

## Theorem

- Let  $K$  be a nonempty subset of  $\mathbb{R}^n$ , where  $n > 1$ . If  $K$  is compact, then every continuous real-valued function defined on  $K$  is **bounded**.

## Definition (Supremum Norm)

- Let  $X$  be a topological space and let  $F$  be the space of all bounded complex-valued continuous functions defined on  $X$ .
  - The supremum norm is the norm defined on  $F$  by

$$\|f\| = \sup_{x \in X} |f(x)| \quad (3)$$



# Thus

## Theorem

- Let  $K$  be a nonempty subset of  $\mathbb{R}^n$ , where  $n > 1$ . If  $K$  is compact, then every continuous real-valued function defined on  $K$  is **bounded**.

## Definition (Supremum Norm)

- Let  $X$  be a topological space and let  $F$  be the space of all bounded complex-valued continuous functions defined on  $X$ .
  - The supremum norm is the norm defined on  $F$  by

$$\|f\| = \sup_{x \in X} |f(x)| \quad (3)$$



# Outline

## 1 Introduction

- The Representation of Functions

## 2 Basic Definitions

- Topology
- Compactness
- Continuous Functions
- Bounding Continuous Functions
- **About Density in a Topology**
- Density Concept
- Having a Nice Space
  - Hausdorff Space
- Measures
- The Borel Measure
- Discriminatory Functions
- The Important Theorem
- Universal Representation Theorem



# I give you an idea

We would like, given  $C(I_n)$

- To prove that there is a function

$$\sum_{j=1}^N \alpha_j f(\mathbf{w}^T \mathbf{x} + \theta_j)$$

Nearby any  $f \in C(I_n)$

- Basically, we want a set

$$R = \left\{ G(\mathbf{x}) \mid G(\mathbf{x}) = \sum_{j=1}^N \alpha_j f(\mathbf{w}^T \mathbf{x} + \theta_j) \right\}$$

such that  $R \subseteq C(I_n)$  and given  $G \in R$ , for all  $\epsilon > 0$ ,  
 $\sup_{\mathbf{x} \in I_n} |G(\mathbf{x}) - f(\mathbf{x})| < \epsilon$ .



# I give you an idea

We would like, given  $C(I_n)$

- To prove that there is a function

$$\sum_{j=1}^N \alpha_j f(\mathbf{w}^T \mathbf{x} + \theta_j)$$

Nearby any  $f(\mathbf{x}) \in C[I_n]$

- Basically, we want a set

$$R = \left\{ G(\mathbf{x}) \mid G(\mathbf{x}) = \sum_{j=1}^N \alpha_j f(\mathbf{w}^T \mathbf{x} + \theta_j) \right\}$$

such that  $R \subseteq C[I_n]$  and given  $G \in R$ , for all  $\epsilon > 0$ ,  
 $\sup_{\mathbf{x} \in I_n} |G(\mathbf{x}) - f(\mathbf{x})| < \epsilon$ .

# Limit Points

## Definition

If  $X$  is a topological space and  $p$  is a point in  $X$ , a neighborhood of  $p$  is a subset  $V$  of  $X$  that includes an open set  $U$  containing  $p$ ,  $p \in U \subseteq V$ .

- This is also equivalent to  $p \in X$  being in the interior of  $V$ .

## Example in a metric space

In a metric space  $(X, d)$ , a set  $V$  is a neighborhood of a point  $p$  if there exists an open ball with center at  $p$  and radius  $r > 0$ , such that

$$B_r(p) = B(p; r) = \{x \in X \mid d(x, p) < r\} \quad (4)$$

is contained in  $V$ .



# Limit Points

## Definition

If  $X$  is a topological space and  $p$  is a point in  $X$ , a neighborhood of  $p$  is a subset  $V$  of  $X$  that includes an open set  $U$  containing  $p$ ,  $p \in U \subseteq V$ .

- This is also equivalent to  $p \in X$  being in the interior of  $V$ .

## Example in a metric space

In a metric space  $(X, d)$ , a set  $V$  is a neighborhood of a point  $p$  if there exists an open ball with center at  $p$  and radius  $r > 0$ , such that

$$B_r(p) = B(p; r) = \{x \in X \mid d(x, p) < r\} \quad (4)$$

is contained in  $V$ .



# Limit Points

## Definition of a Limit Point

Let  $S$  be a subset of a topological space  $X$ . A point  $x \in X$  is a limit point of  $S$  if every neighborhood of  $x$  contains at least one point of  $S$  different from  $x$  itself.

## Example in $\mathbb{R}$

Which are the limit points of the set  $\left\{\frac{1}{n}\right\}_{n=1}^{\infty}$ ?



# Limit Points

## Definition of a Limit Point

Let  $S$  be a subset of a topological space  $X$ . A point  $x \in X$  is a limit point of  $S$  if every neighborhood of  $x$  contains at least one point of  $S$  different from  $x$  itself.

## Example in $\mathbb{R}$

Which are the limit points of the set  $\left\{ \frac{1}{n} \right\}_{n=1}^{\infty}$  ?



# Outline

## 1 Introduction

- The Representation of Functions

## 2 Basic Definitions

- Topology
- Compactness
- Continuous Functions
- Bounding Continuous Functions
- About Density in a Topology
- **Density Concept**
- Having a Nice Space
  - Hausdorff Space
- Measures
- The Borel Measure
- Discriminatory Functions
- The Important Theorem
- Universal Representation Theorem



# This allows to define the idea of density

## Something Notable

A subset  $A$  of a topological space  $X$  is dense in  $X$ , if for any point  $x \in X$ , any neighborhood of  $x$  contains at least one point from  $A$ .

## Classic Example

The real numbers with the usual topology have the rational numbers as a countable dense subset.

- Why do you believe the floating-point numbers are rational?

## In addition

Also the irrational numbers.



# This allows to define the idea of density

## Something Notable

A subset  $A$  of a topological space  $X$  is dense in  $X$ , if for any point  $x \in X$ , any neighborhood of  $x$  contains at least one point from  $A$ .

## Classic Example

The real numbers with the usual topology have the rational numbers as a countable dense subset.

- Why do you believe the floating-point numbers are rational?

Also the irrational numbers.





# This allows to define the idea of density

## Something Notable

A subset  $A$  of a topological space  $X$  is dense in  $X$ , if for any point  $x \in X$ , any neighborhood of  $x$  contains at least one point from  $A$ .

## Classic Example

The real numbers with the usual topology have the rational numbers as a countable dense subset.

- Why do you believe the floating-point numbers are rational?

## In addition

Also the irrational numbers.



# From this, you have the idea of closure

## Definition

The closure of a set  $S$  is the set of all points of closure of  $S$ , that is, the set  $S$  together with all of its limit points.

## Example

The closure of the following set  $(0, 1) \cup \{2\}$

## Meaning

Not all points in the closure are limit points.



# From this, you have the idea of closure

## Definition

The closure of a set  $S$  is the set of all points of closure of  $S$ , that is, the set  $S$  together with all of its limit points.

## Example

The closure of the following set  $(0, 1) \cup \{2\}$

## Warning

Not all points in the closure are limit points.



# From this, you have the idea of closure

## Definition

The closure of a set  $S$  is the set of all points of closure of  $S$ , that is, the set  $S$  together with all of its limit points.

## Example

The closure of the following set  $(0, 1) \cup \{2\}$

## Meaning

Not all points in the closure are limit points.



# Outline

## 1 Introduction

- The Representation of Functions

## 2 Basic Definitions

- Topology
- Compactness
- Continuous Functions
- Bounding Continuous Functions
- About Density in a Topology
- Density Concept
- **Having a Nice Space**
  - Hausdorff Space
- Measures
- The Borel Measure
- Discriminatory Functions
- The Important Theorem
- Universal Representation Theorem



# What Characteristics we would like to have

## First

- We would love to be able to say that separation exist!!!
- Given two functions, we can say they are different if their mappings are different!!!



# What Characteristics we would like to have

## First

- We would love to be able to say that separation exist!!!
- Given two functions, we can say they are different if their mappings are different!!!

## Second

- We want to define a way to measure the open spaces and their pre-images under continuous functions:
  - So we can integrate them!!!



# What Characteristics we would like to have

## First

- We would love to be able to say that separation exist!!!
- Given two functions, we can say they are different if their mappings are different!!!

## Second

- We want to define a way to measure the open spaces and their pre-images under continuous functions:

» So we can integrate them!!!





# What Characteristics we would like to have

## First

- We would love to be able to say that separation exist!!!
- Given two functions, we can say they are different if their mappings are different!!!

## Second

- We want to define a way to measure the open spaces and their pre-images under continuous functions:
  - ▶ So we can integrate them!!!



# Hausdorff Space

## Definition of Separation

Points  $x$  and  $y$  in a topological space  $X$  can be separated by neighborhoods if there exists a neighborhood  $U$  of  $x$  and a neighborhood  $V$  of  $y$  such that  $U$  and  $V$  are disjoint.

### Definition

$X$  is a Hausdorff space if any two distinct points of  $X$  can be separated by neighborhoods.

### What solve the first issue?

- We can identify different functions... by open sets



# Hausdorff Space

## Definition of Separation

Points  $x$  and  $y$  in a topological space  $X$  can be separated by neighborhoods if there exists a neighborhood  $U$  of  $x$  and a neighborhood  $V$  of  $y$  such that  $U$  and  $V$  are disjoint.

## Definition

$X$  is a Hausdorff space if any two distinct points of  $X$  can be separated by neighborhoods.

This solves the first issue!

- We can identify different functions... by open sets



# Hausdorff Space

## Definition of Separation

Points  $x$  and  $y$  in a topological space  $X$  can be separated by neighborhoods if there exists a neighborhood  $U$  of  $x$  and a neighborhood  $V$  of  $y$  such that  $U$  and  $V$  are disjoint.

## Definition

$X$  is a Hausdorff space if any two distinct points of  $X$  can be separated by neighborhoods.

## This solve the first issue!!!

- We can identify different functions... by open sets



# Outline

## 1 Introduction

- The Representation of Functions

## 2 Basic Definitions

- Topology
- Compactness
- Continuous Functions
- Bounding Continuous Functions
- About Density in a Topology
- Density Concept
- Having a Nice Space
  - Hausdorff Space
- **Measures**
- The Borel Measure
- Discriminatory Functions
- The Important Theorem
- Universal Representation Theorem



We have then

Look at what we have

①  $C(I_n)$  is compact

② The continuous functions there are bounded!!!



We have then

Look at what we have

- 1  $C(I_n)$  is compact
- 2 The continuous functions there are bounded!!!



We have then

Look at what we have

- 1  $C(I_n)$  is compact
- 2 The continuous functions there are bounded!!!





Therefore

We can integrate  $f \in C(I_n)$

- However, we need to have a measure  $\mu$  to integrate such functions

Why?

- We want to construct an existence theorem by contradiction and integration is necessary



Cinvestav

# Therefore

We can integrate  $f \in C(I_n)$

- However, we need to have a measure  $\mu$  to integrate such functions

## Why?

- We want to construct an existence theorem by contradiction and integration is necessary



## Now, the Measure Concept

### Definition of $\sigma$ -algebra

Let  $\mathcal{A} \subset \mathcal{P}(X)$ , we say that  $\mathcal{A}$  to be an algebra if

- $\emptyset, X \in \mathcal{A}$ .
- $A, B \in \mathcal{A}$  then  $A \cup B \in \mathcal{A}$ .
- $A \in \mathcal{A}$  then  $A^c \in \mathcal{A}$ .

# Now, the Measure Concept

## Definition of $\sigma$ -algebra

Let  $\mathcal{A} \subset \mathcal{P}(X)$ , we say that  $\mathcal{A}$  to be an algebra if

- 1  $\emptyset, X \in \mathcal{A}$ .
- 2  $A, B \in \mathcal{A}$  then  $A \cup B \in \mathcal{A}$ .
- 3  $A \in \mathcal{A}$  then  $A^c \in \mathcal{A}$ .

## Definition

An algebra  $\mathcal{A}$  in  $\mathcal{P}(X)$  is said to be a  $\sigma$ -algebra, if for any sequence  $\{A_n\}$  of elements in  $\mathcal{A}$ , we have  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ .

# Now, the Measure Concept

## Definition of $\sigma$ -algebra

Let  $\mathcal{A} \subset \mathcal{P}(X)$ , we say that  $\mathcal{A}$  to be an algebra if

- 1  $\emptyset, X \in \mathcal{A}$ .
- 2  $A, B \in \mathcal{A}$  then  $A \cup B \in \mathcal{A}$ .

3  $A \in \mathcal{A}$  then  $A^c \in \mathcal{A}$ .

## Definition

An algebra  $\mathcal{A}$  in  $\mathcal{P}(X)$  is said to be a  $\sigma$ -algebra, if for any sequence  $\{A_n\}$  of elements in  $\mathcal{A}$ , we have  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ .

## Example

In  $X = [0, 1)$ , the class  $\mathcal{A}_0$  consisting of  $\emptyset$ , and all finite unions  $A = \bigcup_{i=1}^n [a_i, b_i)$  with  $0 \leq a_i < b_i \leq a_{i+1} \leq 1$  is an algebra.

# Now, the Measure Concept

## Definition of $\sigma$ -algebra

Let  $\mathcal{A} \subset \mathcal{P}(X)$ , we say that  $\mathcal{A}$  to be an algebra if

- 1  $\emptyset, X \in \mathcal{A}$ .
- 2  $A, B \in \mathcal{A}$  then  $A \cup B \in \mathcal{A}$ .
- 3  $A \in \mathcal{A}$  then  $A^c \in \mathcal{A}$ .

## Definition

An algebra  $\mathcal{A}$  in  $\mathcal{P}(X)$  is said to be a  $\sigma$ -algebra, if for any sequence  $\{A_n\}$  of elements in  $\mathcal{A}$ , we have  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ .

## Example

In  $X = [0, 1)$ , the class  $\mathcal{A}_0$  consisting of  $\emptyset$ , and all finite unions  $A = \bigcup_{i=1}^p [a_i, b_i)$  with  $0 \leq a_i < b_i \leq a_{i+1} \leq 1$  is an algebra.

## Now, the Measure Concept

### Definition of $\sigma$ -algebra

Let  $\mathcal{A} \subset \mathcal{P}(X)$ , we say that  $\mathcal{A}$  to be an algebra if

- 1  $\emptyset, X \in \mathcal{A}$ .
- 2  $A, B \in \mathcal{A}$  then  $A \cup B \in \mathcal{A}$ .
- 3  $A \in \mathcal{A}$  then  $A^c \in \mathcal{A}$ .

### Definition

An algebra  $\mathcal{A}$  in  $\mathcal{P}(X)$  is said to be a  $\sigma$ -algebra, if for any sequence  $\{A_n\}$  of elements in  $\mathcal{A}$ , we have  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$

EXAMPLE  
In  $X = [0, 1)$ , the class  $\mathcal{A}_0$  consisting of  $\emptyset$ , and all finite unions  $A = \bigcup_{i=1}^n [a_i, b_i)$  with  $0 \leq a_i < b_i \leq a_{i+1} \leq 1$  is an algebra.

## Now, the Measure Concept

### Definition of $\sigma$ -algebra

Let  $\mathcal{A} \subset \mathcal{P}(X)$ , we say that  $\mathcal{A}$  to be an algebra if

- 1  $\emptyset, X \in \mathcal{A}$ .
- 2  $A, B \in \mathcal{A}$  then  $A \cup B \in \mathcal{A}$ .
- 3  $A \in \mathcal{A}$  then  $A^c \in \mathcal{A}$ .

### Definition

An algebra  $\mathcal{A}$  in  $\mathcal{P}(X)$  is said to be a  $\sigma$ -algebra, if for any sequence  $\{A_n\}$  of elements in  $\mathcal{A}$ , we have  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$

### Example

In  $X = [0, 1)$ , the class  $\mathcal{A}_0$  consisting of  $\emptyset$ , and all finite unions  $A = \bigcup_{i=1}^n [a_i, b_i)$  with  $0 \leq a_i < b_i \leq a_{i+1} \leq 1$  is an algebra.



## Now, the Measure Concept

### Definition of additivity

Let  $\mu : \mathcal{A} \rightarrow [0, +\infty]$  be such that  $\mu(\emptyset) = 0$ , we say that  $\mu$  is  $\sigma$ -additive if for any  $\{A_i\}_{i \in I} \subset \mathcal{A}$  (Where  $I$  can be finite or infinite countable) of mutually disjoint sets such that  $\cup_{i \in I} A_i \in \mathcal{A}$ , we have that

$$\mu\left(\cup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i) \quad (5)$$

### Definition of Measurability

Let  $\mathcal{A}$  be a  $\sigma$ -algebra of subsets of  $X$ , we say that the pair  $(X, \mathcal{A})$  is a measurable space where a  $\sigma$ -additive function  $\mu : \mathcal{A} \rightarrow [0, +\infty]$  is called a measure on  $(X, \mathcal{A})$ .



## Now, the Measure Concept

### Definition of additivity

Let  $\mu : \mathcal{A} \rightarrow [0, +\infty]$  be such that  $\mu(\emptyset) = 0$ , we say that  $\mu$  is  $\sigma$ -additive if for any  $\{A_i\}_{i \in I} \subset \mathcal{A}$  (Where  $I$  can be finite or infinite countable) of mutually disjoint sets such that  $\cup_{i \in I} A_i \in \mathcal{A}$ , we have that

$$\mu\left(\cup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i) \quad (5)$$

### Definition of Measurability

Let  $\mathcal{A}$  be a  $\sigma$ -algebra of subsets of  $X$ , we say that the pair  $(X, \mathcal{A})$  is a measurable space where a  $\sigma$ -additive function  $\mu : \mathcal{A} \rightarrow [0, +\infty]$  is called a measure on  $(X, \mathcal{A})$ .



# Outline

## 1 Introduction

- The Representation of Functions

## 2 Basic Definitions

- Topology
- Compactness
- Continuous Functions
- Bounding Continuous Functions
- About Density in a Topology
- Density Concept
- Having a Nice Space
  - Hausdorff Space
- Measures
- **The Borel Measure**
- Discriminatory Functions
- The Important Theorem
- Universal Representation Theorem



# A Borel Measure

## Definition

The Borel  $\sigma$ -algebra is defined to be the  $\sigma$ -algebra generated by the open sets (or equivalently, by the closed sets).

# A Borel Measure

## Definition

The Borel  $\sigma$ -algebra is defined to be the  $\sigma$ -algebra generated by the open sets (or equivalently, by the closed sets).

## Definition of a Borel Measure

If  $\mathcal{F}$  is the Borel  $\sigma$ -algebra on some topological space, then a measure  $\mu : \mathcal{F} \rightarrow \mathbb{R}$  is said to be a Borel measure (or Borel probability measure). For a Borel measure, all continuous functions are measurable.

# A Borel Measure

## Definition

The Borel  $\sigma$ -algebra is defined to be the  $\sigma$ -algebra generated by the open sets (or equivalently, by the closed sets).

## Definition of a Borel Measure

If  $\mathcal{F}$  is the Borel  $\sigma$ -algebra on some topological space, then a measure  $\mu : \mathcal{F} \rightarrow \mathbb{R}$  is said to be a Borel measure (or Borel probability measure). For a Borel measure, all continuous functions are measurable.

## Definition of a signed Borel Measure

A signed Borel measure  $\mu : \mathcal{B}(X) \rightarrow \mathbb{R}$  is a measure such that

- $\mu(\emptyset) = 0$ .
- $\mu$  is  $\sigma$ -additive.
- $\sup_{A \in \mathcal{B}(X)} |\mu(A)| < \infty$

# A Borel Measure

## Definition

The Borel  $\sigma$ -algebra is defined to be the  $\sigma$ -algebra generated by the open sets (or equivalently, by the closed sets).

## Definition of a Borel Measure

If  $\mathcal{F}$  is the Borel  $\sigma$ -algebra on some topological space, then a measure  $\mu : \mathcal{F} \rightarrow \mathbb{R}$  is said to be a Borel measure (or Borel probability measure). For a Borel measure, all continuous functions are measurable.

## Definition of a signed Borel Measure

A signed Borel measure  $\mu : \mathcal{B}(X) \rightarrow \mathbb{R}$  is a measure such that

①  $\mu(\emptyset) = 0$ .

②  $\mu$  is  $\sigma$ -additive.

③  $\sup_{A \in \mathcal{B}(X)} |\mu(A)| < \infty$

# A Borel Measure

## Definition

The Borel  $\sigma$ -algebra is defined to be the  $\sigma$ -algebra generated by the open sets (or equivalently, by the closed sets).

## Definition of a Borel Measure

If  $\mathcal{F}$  is the Borel  $\sigma$ -algebra on some topological space, then a measure  $\mu : \mathcal{F} \rightarrow \mathbb{R}$  is said to be a Borel measure (or Borel probability measure). For a Borel measure, all continuous functions are measurable.

## Definition of a signed Borel Measure

A signed Borel measure  $\mu : \mathcal{B}(X) \rightarrow \mathbb{R}$  is a measure such that

- 1  $\mu(\emptyset) = 0$ .
- 2  $\mu$  is  $\sigma$ -additive.

$$\sup_{A \in \mathcal{B}(X)} |\mu(A)| < \infty$$



# A Borel Measure

## Definition

The Borel  $\sigma$ -algebra is defined to be the  $\sigma$ -algebra generated by the open sets (or equivalently, by the closed sets).

## Definition of a Borel Measure

If  $\mathcal{F}$  is the Borel  $\sigma$ -algebra on some topological space, then a measure  $\mu : \mathcal{F} \rightarrow \mathbb{R}$  is said to be a Borel measure (or Borel probability measure). For a Borel measure, all continuous functions are measurable.

## Definition of a signed Borel Measure

A signed Borel measure  $\mu : \mathcal{B}(X) \rightarrow \mathbb{R}$  is a measure such that

- 1  $\mu(\emptyset) = 0$ .
- 2  $\mu$  is  $\sigma$ -additive.
- 3  $\sup_{A \in \mathcal{B}(X)} |\mu(A)| < \infty$

# A Borel Measure

## Regularity

A measure  $\mu$  is Borel regular measure:

- For every Borel set  $B \subseteq \mathbb{R}^n$  and  $A \subseteq \mathbb{R}^n$ ,  
$$\mu(A) = \mu(A \cap B) + \mu(A - B).$$
- For every  $A \subseteq \mathbb{R}^n$ , there exists a Borel set  $B \subseteq \mathbb{R}^n$  such that  $A \subseteq B$  and  $\mu(A) = \mu(B)$ .



# A Borel Measure

## Regularity

A measure  $\mu$  is Borel regular measure:

- 1 For every Borel set  $B \subseteq \mathbb{R}^n$  and  $A \subseteq \mathbb{R}^n$ ,  
$$\mu(A) = \mu(A \cap B) + \mu(A - B).$$

- 2 For every  $A \subseteq \mathbb{R}^n$ , there exists a Borel set  $B \subseteq \mathbb{R}^n$  such that  $A \subseteq B$  and  $\mu(A) = \mu(B)$ .



# A Borel Measure

## Regularity

A measure  $\mu$  is Borel regular measure:

- 1 For every Borel set  $B \subseteq \mathbb{R}^n$  and  $A \subseteq \mathbb{R}^n$ ,  
$$\mu(A) = \mu(A \cap B) + \mu(A - B).$$
- 2 For every  $A \subseteq \mathbb{R}^n$ , there exists a Borel set  $B \subseteq \mathbb{R}^n$  such that  $A \subseteq B$  and  $\mu(A) = \mu(B)$ .



# Outline

## 1 Introduction

- The Representation of Functions

## 2 Basic Definitions

- Topology
- Compactness
- Continuous Functions
- Bounding Continuous Functions
- About Density in a Topology
- Density Concept
- Having a Nice Space
  - Hausdorff Space
- Measures
- The Borel Measure
- **Discriminatory Functions**
- The Important Theorem
- Universal Representation Theorem



# Discriminatory Functions

## Definition

Given the set  $M(I_n)$  of signed regular Borel measures, a function  $f$  is discriminatory if for a measure  $\mu \in M(I_n)$

$$\int_{I_n} f(\mathbf{w}^T \mathbf{x} + \theta) d\mu = 0 \quad (6)$$

for all  $\mathbf{w} \in \mathbb{R}^n$  and  $\theta \in \mathbb{R}$  **implies that**  $\mu = 0$

## Definition

We say that  $f$  is sigmoidal if

$$f(t) \rightarrow \begin{cases} 1 & \text{as } t \rightarrow +\infty \\ 0 & \text{as } t \rightarrow -\infty \end{cases}$$

# Discriminatory Functions

## Definition

Given the set  $M(I_n)$  of signed regular Borel measures, a function  $f$  is discriminatory if for a measure  $\mu \in M(I_n)$

$$\int_{I_n} f(\mathbf{w}^T \mathbf{x} + \theta) d\mu = 0 \quad (6)$$

for all  $\mathbf{w} \in \mathbb{R}^n$  and  $\theta \in \mathbb{R}$  **implies that**  $\mu = 0$

## Definition

We say that  $f$  is sigmoidal if

$$f(t) \rightarrow \begin{cases} 1 & \text{as } t \rightarrow +\infty \\ 0 & \text{as } t \rightarrow -\infty \end{cases}$$

# Outline

## 1 Introduction

- The Representation of Functions

## 2 Basic Definitions

- Topology
- Compactness
- Continuous Functions
- Bounding Continuous Functions
- About Density in a Topology
- Density Concept
- Having a Nice Space
  - Hausdorff Space
- Measures
- The Borel Measure
- Discriminatory Functions
- **The Important Theorem**
- Universal Representation Theorem





# The Important Theorem

## Theorem 1

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

$$G(\mathbf{x}) = \sum_{j=1}^N \alpha_j f(\mathbf{w}_j^T \mathbf{x} + \theta_j), \quad (7)$$

where  $\mathbf{w}_j \in \mathbb{R}^n$  and  $\alpha_j, \theta_j \in \mathbb{R}$  are fixed, are dense in  $C(I_n)$



# Meaning

## In other words

Given any  $g \in C(I_n)$  and  $\epsilon > 0$ , there is a sum,  $G(\mathbf{x})$ , of the above form, for which

$$|G(\mathbf{x}) - g(\mathbf{x})| < \epsilon \quad \forall \mathbf{x} \in I_n \quad (8)$$



# Proof

Let  $S \subset C(I_n)$  be the set of functions of the form  $G(\mathbf{x})$

First,  $S$  is a linear subspace of  $C(I_n)$

## Definition

A subset  $V$  of  $\mathbb{R}^n$  is called a linear subspace of  $\mathbb{R}^n$  if  $V$  contains the zero vector, and is closed under vector addition and scaling. That is, for  $X, Y \in V$  and  $c \in \mathbb{R}$ , we have  $X + Y \in V$  and  $cX \in V$ .

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

Assume that the closure of  $S$  is not all of  $C(I_n)$



# Proof

Let  $S \subset C(I_n)$  be the set of functions of the form  $G(\mathbf{x})$

First,  $S$  is a linear subspace of  $C(I_n)$

## Definition

A subset  $V$  of  $\mathbb{R}^n$  is called a linear subspace of  $\mathbb{R}^n$  if  $V$  contains the zero vector, and is closed under vector addition and scaling. That is, for  $X, Y \in V$  and  $c \in \mathbb{R}$ , we have  $X + Y \in V$  and  $cX \in V$ .

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

Assume that the closure of  $S$  is not all of  $C(I_n)$



# Proof

Let  $S \subset C(I_n)$  be the set of functions of the form  $G(\mathbf{x})$

First,  $S$  is a linear subspace of  $C(I_n)$

## Definition

A subset  $V$  of  $\mathbb{R}^n$  is called a linear subspace of  $\mathbb{R}^n$  if  $V$  contains the zero vector, and is closed under vector addition and scaling. That is, for  $X, Y \in V$  and  $c \in \mathbb{R}$ , we have  $X + Y \in V$  and  $cX \in V$ .

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

Assume that the closure of  $S$  is not all of  $C(I_n)$



# Proof

Then

The closure of  $S$ , say  $R$ , is a **closed proper subspace** of  $C(I_n)$

# Proof

Then

The closure of  $S$ , say  $R$ , is a **closed proper subspace** of  $C(I_n)$

We use the Hahn-Banach Theorem

If  $p : V \rightarrow \mathbb{R}$  is a sub-linear function (i.e. you have  $p(x + y) \leq p(x) + p(y)$  and the product against scalar is the same), and  $\varphi : U \rightarrow \mathbb{R}$  is a linear functional on a linear subspace  $U \subseteq V$  which is dominated by  $p$  on  $U$ , i.e.  $\varphi(x) \leq p(x) \forall x \in U$ .

# Proof

Then

The closure of  $S$ , say  $R$ , is a **closed proper subspace** of  $C(I_n)$

We use the Hahn-Banach Theorem

If  $p : V \rightarrow \mathbb{R}$  is a sub-linear function (i.e. you have  $p(x + y) \leq p(x) + p(y)$  and the product against scalar is the same), and  $\varphi : U \rightarrow \mathbb{R}$  is a linear functional on a linear subspace  $U \subseteq V$  which is dominated by  $p$  on  $U$ , i.e.  $\varphi(x) \leq p(x) \forall x \in U$ .

Then

There exists a linear extension  $\psi : V \rightarrow \mathbb{R}$  of  $\varphi$  to the whole space  $V$ , i.e., there exists a linear functional  $\psi$  such that

- $\psi(x) = \varphi(x) \forall x \in U.$

- $\psi(x) \leq p(x) \forall x \in V.$



# Proof

Then

The closure of  $S$ , say  $R$ , is a **closed proper subspace** of  $C(I_n)$

We use the Hahn-Banach Theorem

If  $p : V \rightarrow \mathbb{R}$  is a sub-linear function (i.e. you have  $p(x + y) \leq p(x) + p(y)$  and the product against scalar is the same), and  $\varphi : U \rightarrow \mathbb{R}$  is a linear functional on a linear subspace  $U \subseteq V$  which is dominated by  $p$  on  $U$ , i.e.  $\varphi(x) \leq p(x) \forall x \in U$ .

Then

There exists a linear extension  $\psi : V \rightarrow \mathbb{R}$  of  $\varphi$  to the whole space  $V$ , i.e., there exists a linear functional  $\psi$  such that

①  $\psi(x) = \varphi(x) \forall x \in U.$

②  $\psi(x) \leq p(x) \forall x \in V.$

# Proof

Then

The closure of  $S$ , say  $R$ , is a **closed proper subspace** of  $C(I_n)$

We use the Hahn-Banach Theorem

If  $p : V \rightarrow \mathbb{R}$  is a sub-linear function (i.e. you have  $p(x + y) \leq p(x) + p(y)$  and the product against scalar is the same), and  $\varphi : U \rightarrow \mathbb{R}$  is a linear functional on a linear subspace  $U \subseteq V$  which is dominated by  $p$  on  $U$ , i.e.  $\varphi(x) \leq p(x) \forall x \in U$ .

Then

There exists a linear extension  $\psi : V \rightarrow \mathbb{R}$  of  $\varphi$  to the whole space  $V$ , i.e., there exists a linear functional  $\psi$  such that

- 1  $\psi(x) = \varphi(x) \forall x \in U.$
- 2  $\psi(x) \leq p(x) \forall x \in V.$

# Proof

It is possible to construct sub-linear function defined as follow

We define the following linear functional

$$T(f) = \begin{cases} f & \text{if } f \in C(I_n) - R \\ 0 & \text{if } f \in R \end{cases} \quad (9)$$

Then

- Using  $T$  as  $p$  and  $\varphi$
- $V = C(I_n)$
- $U = R$



# Proof

It is possible to construct sub-linear function defined as follow

We define the following linear functional

$$T(f) = \begin{cases} f & \text{if } f \in C(I_n) - R \\ 0 & \text{if } f \in R \end{cases} \quad (9)$$

Then

- Using  $T$  as  $p$  and  $\varphi$
- $V = C(I_n)$
- $U = R$



# Therefore

We have

There is a bounded linear functional called  $L \neq 0$

- The  $\psi$  in the Hahn-Banach Theorem
- With  $L(R) = L(S) = 0$ , but  $L(C(I_n) - R) \neq 0$



## Proof

Now, we use the Riesz Representation Theorem

Let  $X$  be a locally compact Hausdorff space. For any positive linear functional  $\psi$  on  $C(X)$ , there is a unique regular Borel measure  $\mu$  on  $X$  such that

$$\psi = \int_X f(x) d\mu(x) \quad (10)$$

for all  $f$  in  $C(X)$

We can then do the following

$$L(h) = \int_{I_n} h(x) d\mu(x) \quad (11)$$

where

For some  $\mu \in M(I_n)$ , for all  $h \in C(I_n)$

## Proof

Now, we use the Riesz Representation Theorem

Let  $X$  be a locally compact Hausdorff space. For any positive linear functional  $\psi$  on  $C(X)$ , there is a unique regular Borel measure  $\mu$  on  $X$  such that

$$\psi = \int_X f(x) d\mu(x) \quad (10)$$

for all  $f$  in  $C(X)$

We can then do the following

$$L(h) = \int_{I_n} h(\mathbf{x}) d\mu(\mathbf{x}) \quad (11)$$

where

For some  $\mu \in M(I_n)$ , for all  $h \in C(I_n)$

## Proof

Now, we use the Riesz Representation Theorem

Let  $X$  be a locally compact Hausdorff space. For any positive linear functional  $\psi$  on  $C(X)$ , there is a unique regular Borel measure  $\mu$  on  $X$  such that

$$\psi = \int_X f(x) d\mu(x) \quad (10)$$

for all  $f$  in  $C(X)$

We can then do the following

$$L(h) = \int_{I_n} h(\mathbf{x}) d\mu(\mathbf{x}) \quad (11)$$

Where?

For some  $\mu \in M(I_n)$ , for all  $h \in C(I_n)$



# Proof

## In particular

Given that  $f(w^T x + \theta)$  is in  $R$  for all  $w$  and  $\theta$

We must have that

$$\int_{I_n} f(w^T x + \theta) d\mu(x) = 0 \quad (12)$$

for all  $w$  and  $\theta$

But we assumed that  $L$  is a discriminator!!!

- Then...  $\mu = 0$  contradicting the fact that  $L \neq 0$ !!! In  $f \in C(I_n) - R$
- We have a contradiction!!!



# Proof

In particular

Given that  $f(\mathbf{w}^T \mathbf{x} + \theta)$  is in  $R$  for all  $\mathbf{w}$  and  $\theta$

We must have that

$$\int_{I_n} f(\mathbf{w}^T \mathbf{x} + \theta) d\mu(\mathbf{x}) = 0 \quad (12)$$

for all  $\mathbf{w}$  and  $\theta$

But we assumed that  $L$  is discriminative!!!

- Then...  $\mu = 0$  contradicting the fact that  $L \neq 0$ !!! In  $f \in C(I_n) - R$
- We have a contradiction!!!



# Proof

## In particular

Given that  $f(\mathbf{w}^T \mathbf{x} + \theta)$  is in  $R$  for all  $\mathbf{w}$  and  $\theta$

## We must have that

$$\int_{I_n} f(\mathbf{w}^T \mathbf{x} + \theta) d\mu(\mathbf{x}) = 0 \quad (12)$$

for all  $\mathbf{w}$  and  $\theta$

## But we assumed that $f$ is discriminatory!!!

- Then...  $\mu = 0$  contradicting the fact that  $L \neq 0$ !!! In  $f \in C(I_n) - R$
- We have a contradiction!!!



# Proof

Finally

The subspace  $S$  of sums of the form  $G$  is dense!!!



# Now, we deal with the sigmoidal function

## Lemma 1

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

## Proof

I will leave this to you... it is possible I will get a question from this proof for the first midterm.



# Now, we deal with the sigmoidal function

## Lemma 1

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

## Proof

I will leave this to you... it is possible I will get a question from this proof for the first midterm.



# Outline

## 1 Introduction

- The Representation of Functions

## 2 Basic Definitions

- Topology
- Compactness
- Continuous Functions
- Bounding Continuous Functions
- About Density in a Topology
- Density Concept
- Having a Nice Space
  - Hausdorff Space
- Measures
- The Borel Measure
- Discriminatory Functions
- The Important Theorem
- **Universal Representation Theorem**



We have the theorem finally!!!

## Universal Representation Theorem for the multi-layer perceptron

Let  $f$  be any continuous sigmoidal function. Then finite sums of the form

$$G(\mathbf{x}) = \sum_{j=1}^N \alpha_j f(\mathbf{w}^T \mathbf{x} + \theta_j) \quad (13)$$

are dense in  $C(I_n)$ .

In other words

Given any  $g \in C(I_n)$  and  $\epsilon > 0$ , there is a sum  $G(\mathbf{x})$  of the above form, for which

$$|G(\mathbf{x}) - g(\mathbf{x})| < \epsilon \quad \forall \mathbf{x} \in I_n \quad (14)$$





# We have the theorem finally!!!

## Universal Representation Theorem for the multi-layer perceptron

Let  $f$  be any continuous sigmoidal function. Then finite sums of the form

$$G(\mathbf{x}) = \sum_{j=1}^N \alpha_j f(\mathbf{w}^T \mathbf{x} + \theta_j) \quad (13)$$

are dense in  $C(I_n)$ .

## In other words

Given any  $g \in C(I_n)$  and  $\epsilon > 0$ , there is a sum  $G(\mathbf{x})$  of the above form, for which

$$|G(\mathbf{x}) - g(\mathbf{x})| < \epsilon \quad \forall \mathbf{x} \in I_n \quad (14)$$



# Proof

## Simple

Combine the theorem and lemma 1... and because the continuous sigmoidals satisfy the conditions of the lemma

- We have our representation!!!

