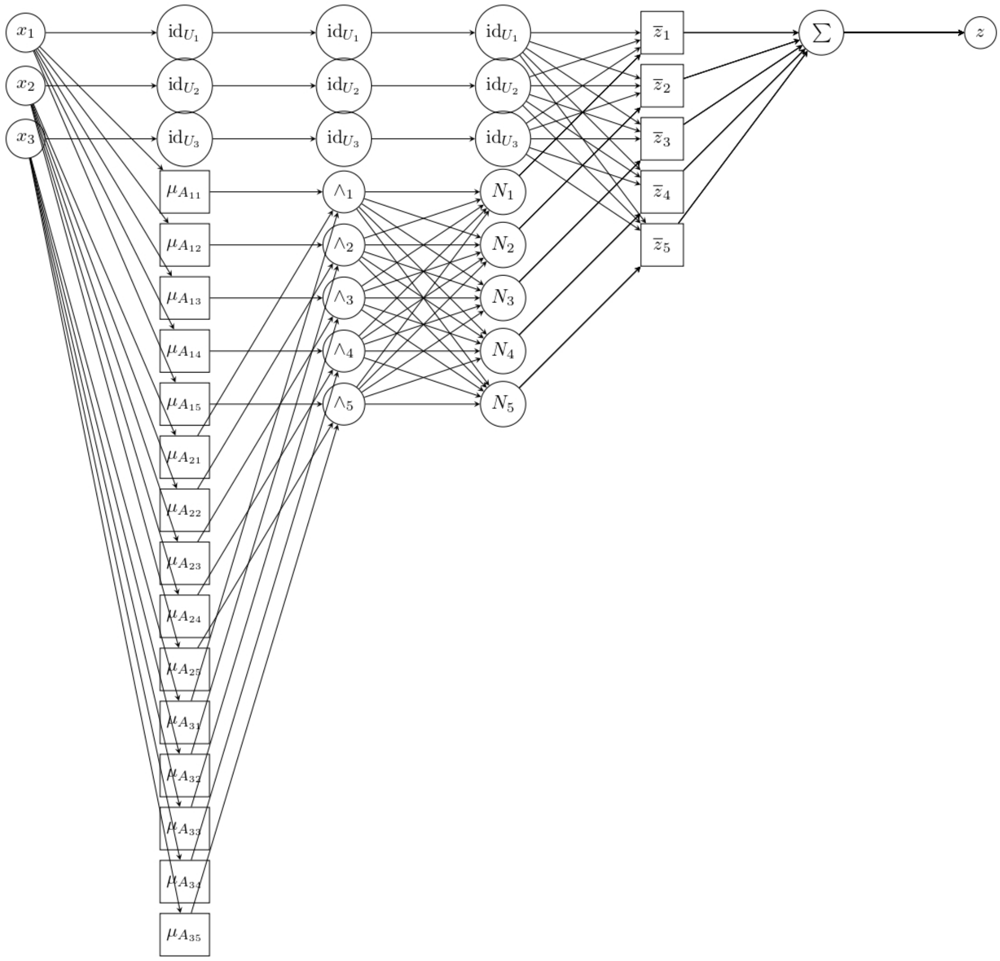
UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II



Introduction to Fuzzy Logic and Neuro Fuzzy systems

Lauriello Lorella & Lucibello Francesco

DEPARTMENT OF MATHEMATICS

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

Naples, 17th of June 2025

Abstract

We give the mathematical foundations of Fuzzy Set Theory and Fuzzy Logic, we present some prominent theoretical achievements of Fuzzy Systems Theory, finally we give a description of ANFIS.

Contents

1 Introduction

Fuzzy Logic is a generalization of classic Logic, embracing the concept of "vagueness" in its theory and being able to deal with statements that are not tractable within classic Logic. Fuzzy Logic can classify statements not only as "true" or "false" as in ordinary Logic, but it's possible to assign a numeric value representing the "degree of truth" of the statement. Moreover Fuzzy Logic is not a branch a Probability Theory and the converse is also true, because Probability Theory deals with "uncertainty". We provide some examples to illustrate the differences.

My dice roll will be 5.

This expression represents an outcome of a dice roll, it is easily evaluated in Probability Theory, since we know that this outcome has a probability of $\frac{1}{6}$, because a dice has 6 faces. It cannot be evaluated by the means of classic Logic, because that expression is not simply true or false; the same holds for the Fuzzy Logic since it's not possible to say if this expression is "partially true". The expression is not vague since we know exactly the possible outcomes $(1, 2, 3, 4, 5, 6)$ but we don't know which is the final outcome until the dice is tossed, so we can't comment on the "truth" of the expression.

The temperature outside is more than $25°C$.

This expression is clearly not vague and there is no uncertainty, meaning that, assuming that you have a thermometer, you can say if this statement is true or false. So this expression can be considered both in classic Logic and in Fuzzy Logic, it doesn't make sense to evaluate it in Probability Theory.[1]

The temperature outside is high.

This expression again is meaningless in the context of Probability, and it is not interpretable in the context of classic Logic (the adjective "high" is vague), while in the context of Fuzzy Logic it's possible to assign a value to this expression representing its level of "truth".

2 Fuzzy set theory

In this section, we will provide an overview of the basic notions of fuzzy set theory, which is a natural extension of classical set theory.

In classical set theory, the fundamental concept is the "set", which is one of the primitive notions, i.e. it doesn't have a definition but is most frequently understood as a collection of objects (elements) having some features distinguishing them from other objects. In the case of classical sets, any element is either a member or not a member of the set, that is, a given object x may belong to a set A (be a member of a set A), or not belong to this set (not be a member of this set). This two possibilities are denoted by $x \in A$ or $x \notin A$.

Human perception, logical thinking and reasoning decisions cannot be modeled using classical set theory, since, in everyday life, people deal with vague concepts that have

[1]It's possible to say that there is a 100% or 0% probability that the temperature outside is more than $25°C$, depending on the case, that makes the matter trivial in the context of Probability Theory.

a rich connotation without an absolute standard of measurement and, as a result, often reflect personal and subjective judgments.
For example, the property of being exactly 35 years old defines a classical set because it divides the universe of individuals into two well-defined and mutually exclusive groups: those who are 35 years old and those who are not.
However, people commonly deal with the concepts of young, old, or middle-aged which are vague and each person does not have the same understanding of these concepts. Moreover, being young (equivalently old) does not define a property that allows a clear and precise distinction among individuals because a person may be considered neither clearly young nor clearly not young. This means that the property of being young does not define a classical set because, in the classical set theory, an element is either entirely in the set or entirely outside of it, there is no ambiguity or partial membership.
On the contrary, fuzzy set theory accepts partial memberships and, therefore, in a sense is a method for rigorously modeling the vagueness and subjectivity in human perception and reasoning, thus breaking away from the deterministic belong-or-don't-belong relationship that characterizes classical set theory. To achieve this, Zadeh proposed the use of membership functions to describe a fuzzy set, that is, an object x may belong to a fuzzy set A with varying membership degrees in the range $[0, 1]$, where 0 and 1 denote, respectively, lack of membership and full membership.

2.1 Fuzzy sets

Definition 2.1 (Fuzzy set)**.** Let $\mu : U \to [0, 1]$ and let

$$A = \{(x, \mu(x)) : x \in U\} .$$

We say that A is a *fuzzy set of* U and μ is the *membership function of* A. We call U the *universe set*, *universe of discourse* or, simply, *universe*.
Remark 2.1. We defined a fuzzy set from a function with domain U and range $[0, 1]$, but it's trivial to get a function of that kind from A, moreover that function is clearly unique, hence we can talk about a fuzzy set and its membership function interchangeably. We usually denote the membership function of a fuzzy set A as μ_A.
We can easily identify a classical set $A \subseteq U$ with the fuzzy set

$$A^* = \{(x, \chi_A(x)) : x \in U\} ,$$

where χ_A is the *characteristic function of* A and it's defined as

$$\chi_A(x) = \begin{cases} 0, & x \notin A \\ 1, & x \in A \end{cases} \quad \forall\, x \in U .$$

Essentially we can identify each subset A of U as a fuzzy set of U with membership function χ_A.
We can imagine the membership function of a fuzzy set as a way to represent "how much an element belongs to a set". For instance we can define the set of "high

temperatures", as the fuzzy set A of $\mathbb{R}$ with membership function

$$\mu_A(x) = \begin{cases} 1, & 50 \leqslant x \\ \frac{x}{50}, & 0 \leqslant x \leqslant 50 \\ 0, & x \leqslant 0 \end{cases} .$$

In this example we are considering temperatures over $50°C$ as being high, the temperatures below 0 as not being high and the temperatures in the middle as being "partially high", for example $25°C$ is 50% high, it belongs to the set of the high temperatures at 50%. Of course the choice of the membership function was arbitrary in this case, but it can be done using expert knowledge or surveys, for example.

Notation 2.2 (Zadeh's representation of a membership function)**.** Let $\mu : U \to [0,1]$, we represent μ using the following notation:

$$\mu = \int_U \frac{\mu(u)}{u} .$$

If $U = \{u_n\}_{n\in\mathbb{N}}$, we can use this notation:

$$\mu = \sum_{i=1}^{\infty} \frac{\mu(u_i)}{u_i} .$$

If $U = \{u_1, \ldots, u_n\}$, we can use this notation:

$$\mu = \sum_{i=1}^{n} \frac{\mu(u_i)}{u_i} = \frac{\mu(u_1)}{u_1} + \cdots + \frac{\mu(u_n)}{u_n} .$$

If $U \subseteq \mathbb{R}$, for each $\mu(u_i) = 0$, you can omit the term $\frac{\mu(u_i)}{u_i}$ in the sum.

These notations are referred to as *Zadeh's representation*, as given in [ZZW23, Def. 3.2].

Notation 2.3 (Vector representation of a fuzzy set)**.** If $U = \{u_1, \ldots, u_n\}$ is a finite set and A is a fuzzy set of U with membership function μ, then we can represent A as follows

$$A = [\mu(u_1) \cdots \mu(u_n)] .$$

2.2 Fuzzy set operations

In this section, we use the definitions and notations as in [Pro+17].

The fuzzy set operations are defined with respect to the sets' membership functions.

Definition 2.4 (Inclusion relation of fuzzy sets)**.** Let A and B be two fuzzy sets of universe U. A is a *subset* of B, denoted as $A \subseteq B$, if

$$\forall x \in U, \mu_A(x) \leqslant \mu_B(x)$$

Definition 2.5 (Equality relation of fuzzy sets)**.** Let A and B be two fuzzy sets of universe U. A and B are *equal* if

$$\forall x \in U, \mu_A(x) = \mu_B(x)$$

Definition 2.6 (Complement of a fuzzy set)**.** Let A be a fuzzy set of universe U. The fuzzy set A' defined as

$$\forall x \in U, \mu_{A'}(x) = 1 - \mu_A(x)$$

is called the *complement* of A.
In order to define union and intersection of two fuzzy sets, we need to define the *t-norm* and *t-conorm*.
Definition 2.7 (Triangular norm)**.** A *Triangular norm* or *t-norm* is a mapping $T : [0,1] \times [0,1] \to [0,1]$ with the following four properties.

- Commutativity: $T(x, y) = T(y, x)$
- Monotonicity: $T(x_1, y_1) \leqslant T(x_2, y_2)$, if $x_1 \leqslant x_2$ and $y_1 \leqslant y_2$
- Associativity: $T(x, T(y, z)) = T(T(x, y), z)$
- Linearity: $T(x, 1) = x$

Notation 2.8. A triangular norm can also be denoted with the symbol $\wedge$, i.e. $T(x, y) = x \wedge y$ are equivalent notations.
Definition 2.9 (Triangular conorm)**.** A *Triangular conorm* or *t-conorm* or *s-norm* is a mapping $C : [0,1] \times [0,1] \to [0,1]$ with the following four properties.

- Commutativity: $C(x, y) = C(y, x)$
- Monotonicity: $C(x_1, y_1) \leqslant C(x_2, y_2)$, if $x_1 \leqslant x_2$ and $y_1 \leqslant y_2$
- Associativity: $C(x, C(y, z)) = C(C(x, y), z)$
- Linearity: $C(x, 0) = x$

Notation 2.10. A triangular conorm can also be denoted with the symbol $\vee$, i.e. $C(x, y) = x \vee y$ are equivalent notations.
There exist various t-norms and t-conorms.
The most common t-norms are:

- *standard intersection* or *minimum t-norm*: $T(x, y) = \min(x, y)$
- *algebraic product*: $T(x, y) = xy$.

Similarly, the most common t-conorms are

- *standard union* or *maximum t-conorm*: $C(x, y) = \max(x, y)$
- *algebraic sum* or *probabilistic sum*: $C(x, y) = x + y - xy$.

In general, the union of two fuzzy sets is described by t-conorms, whereas their intersection is described by t-norms.
Definition 2.11 (Union and intersection of fuzzy sets)**.** Let A and B fuzzy sets of universe U. The *union* of A and B is the fuzzy set $A \cup B$:

$$\forall x \in U, \mu_{A \cup B}(x) = C(\mu_A(x), \mu_B(x)) = \mu_A(x) \vee \mu_B(x)$$

where $C : [0,1] \times [0,1] \to [0,1]$ is a t-conorm.
Using the standard union or the algebraic sum, we have respectively:

$$\forall x \in U, \mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$$

called *maximal operator of fuzzy sets* and

$$\forall x \in U, \mu_{A \cup B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x)\mu_B(x)$$

called *sum operator of fuzzy sets.*
The *intersection* of A and B is the fuzzy set $A \cap B$:

$$\forall x \in U, \mu_{A\cap B}(x) = T(\mu_A(x), \mu_B(x)) = \mu_A(x) \wedge \mu_B(x)$$

where $T : [0,1] \times [0,1] \to [0,1]$ is a t-norm.
Using the standard intersection or algebraic product, we have respectively:

$$\forall x \in U, \mu_{A\cap B}(x) = \min(\mu_A(x), \mu_B(x))$$

called *minimum operator of fuzzy sets* and

$$\forall x \in U, \mu_{A\cap B}(x) = \mu_A(x)\mu_B(x)$$

called *product operator of fuzzy sets.*
Remark 2.2. We can see that the fuzzy set operations defined above are extensions of the classical set operations. Let us consider two crisp sets A and B of a universe U. In the classical set theory, we give the following definitions:

- the *complement* of A is the subset A^c consisting of the elements of U that do not belong to A;
- the *union* of A and B is the subset $A \cup B$ consisting of the elements that belong to at least one of A or B.
- the *intersection* of A and B is the subset $A \cap B$ consisting of the elements that belong to both A and B

Furthermore, the following relations hold between the characteristic functions:

- $\chi_{A^c} = 1 - \chi_A$;
- $\chi_{A\cup B} = \chi_A + \chi_B - \chi_A\chi_B = \max(\chi_A, \chi_B)$.
- $\chi_{A\cap B} = \chi_A\chi_B = \min(\chi_A, \chi_B)$

However, as mentioned earlier, every crisp subset of U can be identified with a fuzzy set of U by defining its membership function as its characteristic function. As a consequence:

- the fuzzy complement of A is $\{(x, \chi_{A^c}(x)) : x \in U\}$
- the fuzzy union of A and B is $\{(x, \chi_{A\cup B}(x)) : x \in U\}$
- the fuzzy intersection of A and B is $\{(x, \chi_{A\cap B}(x)) : x \in U\}$

Definition 2.12 (Cartesian Product of fuzzy sets)**.** Let $A_1, \ldots, A_n$ be fuzzy sets of $U_1, \ldots, U_n$ respectively. A fuzzy set F of $U_1 \times \cdots \times U_n$ of the form

$$\mu_F(x_1, \ldots, x_n) = \mu_{A_1}(x_1) \wedge \cdots \wedge \mu_{A_n}(x_n), \quad \forall\, (x_1, \ldots, x_n) \in U_1 \times \cdots \times U_n$$

where $\wedge$ in a t-norm, is called *cartesian product of* $A_1, \ldots, A_n$ and is denoted by $F = A_1 \times \cdots \times A_n$.
Definition 2.13 (Inner product of fuzzy sets)**.** Let A and B fuzzy sets of universe U. The *inner product* of A and B is

$$A \circ B = \bigvee_{x\in U} (\mu_A(x) \wedge \mu_B(x)).$$

where $\wedge$ and $\vee$ are respectively a t-norm and an s-norm.
Remark 2.3. The inner product of A and B is a scalar; specifically, $A \circ B \in [0,1]$.

2.3 Fuzzy relations

Definition 2.14 (Fuzzy relation)**.** Let U and V be two universe sets, i.e. two non-empty sets. A fuzzy set

$$R = \big\{\big((x, y), \mu_R(x, y)\big) : (x, y) \in U \times V\big\}$$

of the cartesian product $U \times V$ is called *fuzzy relation on* $U \times V$ or *binary fuzzy relation from* U *to* V or *fuzzy relation* for short.
The membership function $\mu_R : U \times V \to [0, 1]$ associates to each pair $(x, y) \in U \times V$ the degree of relationship between x and y.
Similar to the binary fuzzy relation, we can define a *multidimensional fuzzy relation* as

$$R = \big\{\big((x_1, \dots, x_n), \mu_R(x_1, \dots, x_n)\big) : (x_1, \dots, x_n) \in U_1 \times \dots \times U_n\big\}$$

where $\mu_R : U_1 \times \dots \times U_n \to [0, 1]$ is a membership function of an n-dimensional fuzzy set defined in universe $U_1 \times \dots \times U_n$.
Notation 2.15. Suppose $U = \{x_1, \dots, x_n\}$ and $V = \{y_1, \dots, y_m\}$ are finite sets, a fuzzy relation R can be represented by a matrix

$$R = \begin{bmatrix} \mu_R(x_1, y_1) & \mu_R(x_1, y_2) & \cdots & \mu_R(x_1, y_m) \\ \mu_R(x_2, y_1) & \mu_R(x_2, y_2) & \cdots & \mu_R(x_2, y_m) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_R(x_n, y_1) & \mu_R(x_n, y_2) & \cdots & \mu_R(x_n, y_m) \end{bmatrix}$$

Definition 2.16 (Fuzzy relation on fuzzy sets)**.** Let A and B be fuzzy sets of U and V, respectively. A *fuzzy relation on* A *and* B is a fuzzy set

$$R = \big\{\big((x, y), \mu_R(x, y)\big) : (x, y) \in U \times V\big\}$$

such that $\forall\, (x, y) \in U \times V,\ \mu_R(x, y) \leqslant \min(\mu_A(x), \mu_B(y))$.

2.4 Compositions with binary fuzzy relations

Because fuzzy relations are fuzzy sets, they are subject to the same operations as fuzzy sets. Additionally, binary fuzzy relations in different product spaces may be composed. This operation of composition is also called *synthetic operation*. In general, we can define this operation as follows.
Definition 2.17 (Synthetic operation)**.** Let U, V and W be 3 universe sets and let R and S be binary fuzzy relations on $U \times V$ and on $V \times W$, respectively. The *composition*, *synthesis* or *synthetic operation*, of R and S, is the binary relation on $U \times W$:

$$T = R \circ S\,.$$

Meaning

$$\forall\, (x, z) \in U \times W,\ \mu_T(x, z) = \bigvee_{y \in V} \mu_R(x, y) \wedge \mu_S(y, z).$$

Different versions of the composition have been proposed, depending on the choice of the operators $\wedge$ and $\vee$. Frequently used compositions are the so-called *supremum-t-norm composition*, i.e

$$\forall (x,z) \in U \times W, \mu_T(x,z) = \sup_{y \in V} \mu_R(x,y) \wedge \mu_S(y,z)$$

where $\wedge : [0,1] \times [0,1] \to [0,1]$ denotes a t-norm.

- if $\wedge$ is the standard intersection t-norm (i.e. minimum t-norm), the composition is called *max-min composition*
- if $\wedge$ is the product t-norm , the composition is called *max-prod composition.*

Remark 2.4. For relations described by relation matrices, the above compositions can be achieved by multiplication of matrices with multiplication of elements replaced by t-norm and the adding of elements replaced by s-norm.

Definition 2.18 (Zadeh's compositional rule)**.** Let R and A be, respectively, a fuzzy relation on $U \times V$ and a fuzzy set of U. The composition

$$B = A \circ R$$

is called the *conclusion made from the fact A based on the rule R* .

Remark 2.5. $B = A \circ R$ is the synthetic operation and means that

$$\forall y \in V, \mu_B(y) = \bigvee_{x \in U} \mu_A(x) \wedge \mu_R(x,y)\,.$$

3 Fuzzy systems

This section describes the basic concepts of *fuzzy systems.*

In simple terms, a fuzzy system is a computing framework based on the concepts of fuzzy set theory, fuzzy conditional rules and fuzzy reasoning [Alo+21].

In the following, we first formally define a fuzzy system, and then we provide an interpretation of these concepts that will make clear why fuzzy systems can "simulate human thinking procedure".

3.1 Fuzzy conditional rule and Fuzzy inference

Definition 3.1 (Implication Operator)**.** An *implication operator* is a function

$$\phi : [0,1] \times [0,1] \to [0,1]$$

which is a t-norm or has the following properties:

- ϕ is continuous
- $\forall\, a,b,c \in [0,1] : a \leqslant c, \phi(a,b) \geqslant \phi(c,b)$
- $\forall\, a,b,c \in [0,1] : b \leqslant c, \phi(a,b) \leqslant \phi(a,c)$
- $\forall\, b \in [0,1], \phi(0,b) = 1$
- $\forall\, a \in [0,1], \phi(a,1) = 1$
- $\phi(1,0) = 0$

If ϕ is a t-norm we say that ϕ is a *conjunctive implication operator*; otherwise, we speak of *logical implication operator.*

Various implication operators have been defined. For example,

- the standard intersection (minimum) t-norm $\phi(x, y) = \min(x, y)$ is the *minimum (Mamdani) implication operator*
- $\phi(x, y) = \max(1 - x, \min(x, y))$ is the *Early Zadeh implication operator* or *max-min implication operator*
- $\phi(x, y) = \min(1, 1 - x + y)$ is the *Lukasiewicz implication operator*.

Definition 3.2 (Fuzzy conditional rule)**.** Let A and B be fuzzy sets of U and V, respectively. A *fuzzy conditional rule* (*fuzzy implication* or *fuzzy IF-THEN rule*) is a fuzzy relation R on $U \times V$ of the form

$$\mu_R(x, y) = \phi(\mu_A(x), \mu_B(y))$$

where ϕ is an implication operator. In this case, R is denoted by

$$A \implies B$$

and A is called *antecedent* (*premise*), whereas B is called *consequent* (*conclusion*).
A *MISO* (Multiple Inputs Single Output) *fuzzy conditional rule with conjunctive antecedent* (or *canonical fuzzy if-then rule*) is a fuzzy implication of the form

$$A_1 \times A_2 \times \cdots \times A_N \implies B$$

where the antecedent $A \in \mathcal{F}(U_1 \times \cdots \times U_n)$ is a cartesian product of fuzzy sets $A_i \in \mathcal{F}(U_i)$.
Similarly, a *MISO fuzzy conditional rule with disjunctive antecedent* is a fuzzy implication where the antecedent is a fuzzy set A of a universe $U = U_1 \times \cdots \times U_n$ such that

$$\mu_A(x_1, \ldots, x_n) = \mu_{A_1}(x_1) \vee \cdots \vee \mu_{A_n}(x_n)$$

where $\forall i = 1, \ldots, n, A_i$ is a fuzzy set of universe U_i and $\vee$ is a t-conorm.
Remark 3.1. In the context of fuzzy systems, we use only canonical fuzzy if-then rules. For this reason, for brevity, we will refer to this type simply as fuzzy conditional rule/implication/if-then rule, etc.
Remark 3.2. The values $\mu_A(x)$, $\mu_B(y)$ and $\mu_R(x, y)$ can be interpreted respectively as a truth degree of the antecedent in x, a truth degree of the consequent in y and a truth degree of the implication in (x, y). In this sense, a logical implication operator, thanks to the properties $\phi(0, b) = 1$, $\phi(a, 1) = 1$, $\phi(1, 0) = 0$, generalizes the truth table of the implication operator in Boolean logic, that is

μ_A	μ_B	$\mu_{A \implies B}$
0	0	1
0	1	1
1	0	0
1	1	1
x	y	$\phi(x, y)$

However, this doesn't hold in general; for example, it doesn't hold for the minimum implication operator ($\min(0, 0) = 0 \neq 1$). So, in general, a fuzzy implication is not simply a generalization of classical logic implication.

Definition 3.3 (Powerfuzzyset)**.** Let U be the universe of discourse, we denote the set of the fuzzy sets of U as $\mathcal{F}(U)$, formally

$$\mathcal{F}(U) = \left\{\{(x, \mu(x)) \mid x \in U\} \mid \mu : U \to [0, 1]\right\}.$$

In the fuzzy systems literature ([TRK15], [ZZW23], [Pro+17], [Alo+21], [Jan+97]) it's quite common to define fuzzy sets verbally, we try to give a description in mathematical terms.

Definition 3.4 (Fuzzy system)**.** Let $U \subseteq \mathbb{R}^n$ be the *input universe* and $V \subseteq \mathbb{R}^m$ the *output universe*. Let $r_1, r_2, r_3, r_4 \in \mathbb{N}_0$ and let

$$\mathcal{K} = \mathbb{R}^{r_1} \times \mathcal{F}(U)^{r_2} \times \mathcal{F}(V)^{r_3} \times \mathcal{F}(U \times V)^{r_4}.$$

Let $F : U \times \mathcal{K} \to \mathcal{F}(U)^p$ be the *fuzzification algorithm*, $I : \mathcal{F}(U)^p \times \mathcal{K} \to \mathcal{F}(V)^q$ be the *fuzzy inference algorithm* and $D : \mathcal{F}(V)^q \times \mathcal{K} \to V$ be the *defuzzification algorithm*.
Let $K \in \mathcal{K}$ be the *Knowledge Base*, let $F_K = F(\cdot, K)$ be the *fuzzification interface*, or *fuzzification*, let $I_K = I(\cdot, K)$ be the *fuzzy inference machine*, or *fuzzy inference*, and let $D_K = D(\cdot, K)$ be the *defuzzification interface*, or *defuzzification*. Let $f = D_K \circ I_K \circ F_K : U \to V$, then we say that f is a *fuzzy system with Knowledge Base* K or simply a *fuzzy system*.
If $m = 1$ we call f a *MISO (Multiple Inputs Single Output) fuzzy system*.
If $n, m = 1$ we call f a *SISO (Single Input Single Output) fuzzy system*.

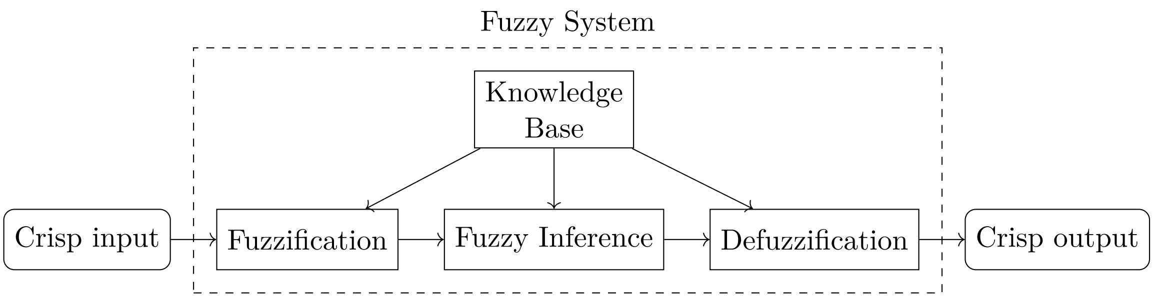


Figure 1: Fuzzy System processing input

In Figure 1 is represented a general fuzzy system.[2]
This definition is problematic because it is possible to prove that any function $f : U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^m$ is a fuzzy system, hence the definition is "too general". In fact, let $F_U : U \to \mathcal{F}(U)$ and $F_V : V \to \mathcal{F}(V)$ be respectively the point fuzzification of U and V, we choose

1. The fuzzification interface $F = F_U$
2. The fuzzy inference machine $I = F_V \circ f \circ F_U^{-1}$
3. As defuzzification interface $D = F_V^{-1}$

notice that none of them depend on any knowledge base. We get that $f = D \circ I \circ F$, so f is a fuzzy system according to this definition.
For this purpose we provide a more specific definition that still includes all the fuzzy systems we consider.

[2]Where "crisp" is used as "non-fuzzy"

Definition 3.5 (Fuzzy system)**.** Let $U \subseteq \mathbb{R}^n$ be the *input universe* and $V \subseteq \mathbb{R}^m$ the *output universe*. Let $r_1, r_2, r_3, r_4 \in \mathbb{N}_0$ and let

$$\mathcal{K} = \mathbb{R}^{r_1} \times \mathcal{F}(U)^{r_2} \times \mathcal{F}(V)^{r_3} \times \mathcal{F}(U \times V)^{r_4}\,.$$

Let $F : U \times \mathcal{K} \to \mathcal{F}(U)$ be the *fuzzification algorithm*, $I : \mathcal{F}(U) \times \mathcal{K} \to \mathcal{F}(V)^q$ be the *fuzzy inference algorithm* and $D : \mathcal{F}(V)^q \times \mathcal{K} \to V$ be the *defuzzification algorithm*. Let $K \in \mathcal{K}$ be the *Knowledge Base*, let $F_K = F(\cdot, K)$ be the *fuzzification interface*, or *fuzzification*, let $I_K = I(\cdot, K)$ be the *fuzzy inference machine*, or *fuzzy inference*, and let $D_K = D(\cdot, K)$ be the *defuzzification interface*, or *defuzzification*. Moreover

$$I(A, K) = (A \circ R_1(A, K), \ldots, A \circ R_q(A, K)) \quad \forall\, A \in \mathcal{F}(U)\ \forall\, K \in \mathcal{K}\,,$$

where $R : \mathcal{F}(U) \times \mathcal{K} \to \mathcal{F}(U \times V)^q$ is the *fuzzy rules generation algorithm*, for each $i \in \{1, \ldots, q\}$ $R_i : \mathcal{F}(U) \times \mathcal{K} \to \mathcal{F}(U \times V)$ is the *i-th fuzzy rule generation algorithm* and $\circ$ is a synthetic operation. Let $R_K = R(\cdot, K)$ be the *fuzzy rules generator* and let $R_{K,i} = R_i(\cdot, K)$ be the *i-th fuzzy rule generator*.
Let $f = D_K \circ I_K \circ F_K : U \to V$, then we say that f is a *fuzzy system with Knowledge Base K* or simply a *fuzzy system*.
If $m = 1$ we call f a *MISO (Multiple Inputs Single Output) fuzzy system*.
If $n, m = 1$ we call f a *SISO (Single Input Single Output) fuzzy system*.
If you want an even more specific definition check Definition A.1.
The are various fuzzification, fuzzy inference and defuzzification algorithms and different choices lead to different types of fuzzy systems.
Definition 3.6 (Point fuzzification)**.** Let U be the universe of discourse, let K be a knowledge base, we define the *point fuzzification algorithm* $F : U \to \mathcal{F}(U)$, so that

$$F(x, K) = F_K(x) = A_x \quad \forall\, K \in \mathcal{K}\,,$$

where A_x is the fuzzy set such that $\mu_{A_x} = \chi_{\{x\}}$.

3.2 Linguistic variables

At this point, we aim to provide an interpretation of the concepts defined so far to make clear why the theories of fuzzy sets, fuzzy logic and fuzzy systems provide a formal mathematical representation of human knowledge, reasoning and decision making about complex problems. In fact, humans are able to control many processes without requiring precise or complete knowledge of the problem or system. Instead, they rely on a form of knowledge often empirical expressed through imprecise natural language terms and conditional rules. A classic example can be found in how a person regulates the temperature of a room. Consider a person entering a room and feeling that it is "a bit cold." Without any precise knowledge about heat exchange or ambient conditions, they might slightly turn up the heater. Later, if the room feels "too warm," they might reduce the heating. These decisions are not based on equations or measurements, but rather on subjective linguistic terms like "cold," "comfortable," or "hot" and still lead to satisfactory temperature regulation.
We can say humans reason in terms of linguistic variables that, informally speaking, are variables whose values are not numbers, but rather words in natural language and can be formally defined in the context of fuzzy set theory as follows:

Definition 3.7 (Linguistic variable)**.** A *linguistic variable* is a quintuple

$$x = (N, U, L, G, M)$$

where:

- N is the name of the variable x,
- U is the universe of discourse, i.e. a crisp or classical set,
- L is the set of linguistic values (terms) of x being a collection of labels for a family of fuzzy sets of U
- G is the set of syntactic rules defined by grammar determing all terms in L,
- M is a semantic rule that defines the meaning of all labels in L, i.e. assigns to each linguistic value in L a fuzzy set of U, i.e. we can see M as a function $M : L \to \mathcal{F}(U)$

Referring back to the initial example, we can consider temperature as name of a linguistic variable, whose values might include *cold*, *warm*, and *hot*, each interpreted as fuzzy sets over the universe of real numbers representing degrees.
Humans use linguistic variable in propositions expressed in natural language, for example "*the temperature is high*". These propositions are represented in fuzzy set theory as linguistic statement.
Definition 3.8 (Elementary linguistic statement)**.** Let $x = (N, U, L, G, M)$ a linguistic variable. An *elementary linguistic statement* or *elementary fuzzy expression* for x is an expression of the form

$$x \text{ is } A$$

where $A = M(l),\ l \in L$ is a fuzzy set of U labeled by l. This elementary statement should be read as: "N *is* l".
For example, let x be a linguistic variable with name N = "temperature" and $A = M(\text{hot})$, then the statement x is A should be read "temperature is hot".
A more complex fuzzy expression can be obtained by combining two or more elementary expressions. It can be presented in the conjunctive form:

$$x_1 \text{ is } A_1 \text{ and } x_2 \text{ is } A_2$$

or disjunctive form:

$$x_1 \text{ is } A_1 \text{ or } x_2 \text{ is } A_2$$

or implication (or IF-THEN) form

$$\text{if } x_1 \text{ is } A_1 \text{ then } x_2 \text{ is } A_2$$

where x_1 and x_2 are linguistic variables, A_1 and A_2 are fuzzy sets in their respective universe. We can generalize from two to an arbitrary number of linguistic variables, combine these forms and obtain, for example:

$$\text{if } x_1 \text{ is } A_1 \text{ and } x_2 \text{ is } A_2 \text{ and } \cdots \text{ and } x_n \text{ is } A_n \text{ then } y \text{ is } B.$$

If the fuzzy sets $A_1, \ldots, A_n, B$ are associated to linguistic values respectively $l_1, \ldots, l_n, l$ the expression can be read as

$$\text{If } N_1 \text{ is } l_1 \text{ and ... and } N_n \text{ is } l_n \text{ then } N \text{ is } l$$

where $N_1, \ldots, N_n, N$ are the name respectively of $x_1, \ldots, x_n, y$.
For example, if N_1 is temperature, N_2 is humidity, N is speed, l_1 is "hot", l_2 is "dry" and l is "fast", we can have read it like this:

If temperature is hot and humidity is dry then speed is fast

In practical applications, fuzzy expressions are always represented as fuzzy sets. For example, conjunctive forms are modeled as the Cartesian product of fuzzy sets, while IF-THEN forms are expressed as fuzzy implications [3].
We will also use the following
Notation 3.9. Let $U_1, \ldots, U_n, V$ be universes of discourse, and let $A_1, \ldots, A_n, B$ be fuzzy sets respectively on $U_1, \ldots, U_n, V$. The notation

$$\text{IF } x_1 \text{ is } A_1 \text{ and } \cdots \text{ and } x_n \text{ is } A_n \text{ THEN } z \text{ is } B\,, \tag{1}$$

stands for a fuzzy implication of the form

$$A_1 \times \cdots \times A_n \implies B\,.$$

This notation allow us to link the intuitive human thinking process, expressed in the form of words, and fuzzy logic. This is what makes fuzzy systems "explainable" compared to other mathematical tools used in the applications.
We will make an analogy with the human brain, to give a better understanding of such design.
We have

1. The crisp input is an element $x_0 \in U$. It represents a measurement of some physical quantity, for example the brightness of a color or the temperature of an object.
2. The Knowledge Base is made up of fuzzy sets (in particular there can be fuzzy relations and fuzzy implications) and real numbers, they are used as "parameters" of the fuzzification, fuzzy inference and defuzzification algorithms. Intuitively, it is a set of information in all the phases of the "thought", for a human it can be thought as its experience and knowledge.
3. The Fuzzification Interface takes as input x_0 and, $F_K(x_0)$ returns a fuzzy set A'. This is similar to a human evaluation of the outer world through its senses, humans get crisp inputs from the surrounding world by our sensory organs and they get "interpreted" by our brain as sensations. Generally speaking something can be interpreted in different ways depending on the knowledge we have about it, for example a baby that never saw fire before may think it is a kind of fun game to play with, but, after the first painful experience, the baby will "interpret" differently fire and will associate a different "meaning" to it. The fuzzy sets here represent how the brain interpreted the physical quantities and they are interpreted in the form of "vague information".
4. The Fuzzy Inference Machine takes as input the fuzzy set A' and $I_K(A')$ returns the fuzzy sets $C_1, \ldots, C_q$. This is similar to a human reasoning process, because our brain "processed" the "vague information" it had and got to a conclusion, that is still in the form of vague information.
5. The Defuzzification Interface D_K takes as input the fuzzy sets $C_1, \ldots, C_q$ and, $D_K(C_1, \ldots, C_q) = y_0$ returns a real number. The human equivalent in this case is performing an action in the real world that best represents the conclusion it got to. The conclusion is "vague", but, in order to perform an action, an actual physical value is needed. For example if a human touches a very hot

[3] See previous section for the definitions of Cartesian product and fuzzy implication

object they will throw it immediately, and the "physical value" that is given as "output" is the electrical impulse given to the arm, that in turns translates to the speed of the arm getting away from the hot object.

4 Fuzzy systems as universal approximators

Notation 4.1 (Universal Approximators)**.** We say that the elements of a class of functions are *universal approximators* if they approximate with an arbitrary degree of accuracy another class of functions, according to some metric.
In this section we expand some known results that prove that fuzzy systems are universal approximators.

4.1 Wang Theorem

The following is a theorem due to [Wan92] and it makes use of the *Stone-Weierstrass Theorem.*
Definition 4.2 (Algebra over a field)**.** Let $\mathbb{K}$ be a field. An *algebra over* $\mathbb{K}$ or a $\mathbb{K}$-*algebra* is a structure $(A, +, \cdot, \star)$ satisfying the following conditions:

1. $(A, +, \cdot)$ is a ring
2. $(A, +, \star)$ is a vector space on $\mathbb{K}$
3. The operations $\cdot : A \times A \to A$ and $\star : \mathbf{K} \times A \to A$ are "compatible", i.e.

$$(a \star x) \cdot (b \star y) = (ab) \star (x \cdot y) \quad \forall\, a, b \in \mathbb{K},\ \forall\, x, y \in A$$

Remark 4.1. There are some generalizations of this definition, for example it's possible to give the definition of an *algebra over a ring.*
Theorem 4.1 (Stone-Weierstrass)**.** *Let $U \subset \mathbb{R}^n$ be a compact set and $Z \subset C(U)$. If*

1. *Z is an algebra over $\mathbb{R}$*
2. *Z separates the points on U i.e.*

$$\forall\, x, y \in U : x \neq y\ \exists\, f \in Z : f(x) \neq f(y)$$

3. *Z vanishes at no point of U, i.e.*

$$\forall\, x \in U \quad \exists f \in Z : f(x) \neq 0$$

then Z is dense in $C(U)$, with respect to the ∞-norm.
We define a set of MISO fuzzy systems $X(U)$ on an input universe $U \subseteq \mathbb{R}^n$ and output universe $\mathbb{R}$, $X(U)$ is parametrized by some "design parameters". For each $i \in \{1, \dots, n\}$,

1. The number $m_i \in \mathbb{N}$
2. The fuzzy sets of the input universe $\{A_i^j \in \mathcal{F}(U_i) : 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant m_i\}$, where $U_i = \{y \in \mathbb{R} : \exists\, x \in U \wedge x_i = y\}$
3. The number of fuzzy output sets $m_0 \in \mathbb{N}$
4. Fuzzy sets of the output universe $\{B^k \in \mathcal{F}(\mathbb{R}) : 1 \leqslant k \leqslant m_0\}$
5. The number of rules $l \in \mathbb{N}$

6. For each $k \in \{1, \ldots, l\}$, a fuzzy rule of the form

$$R_k = \text{IF } x_1 \text{ is } A_1^{j_{1,k}} \text{ and } \cdots \text{ and } x_n \text{ is } A_n^{j_{n,k}} \text{ THEN } z \text{ is } B^{j_{0,k}}\,, \tag{2}$$

where $1 \leqslant j_{i,k} \leqslant m_i$ for every $i \in \{0, \ldots, n\}$.
7. The fuzzy inference algorithm[4] $I : \mathcal{F}(U) \times \mathcal{K} \to \mathcal{F}(\mathbb{R})^{r_3}$
8. The defuzzification algorithm: $D : \mathcal{F}(\mathbb{R})^{r_3} \times \mathcal{K} \to \mathbb{R}$

On the other hand, the fuzzification algorithm F is fixed and it doesn't depend on the knowledge base, so we identify it with the fuzzification interface and it is the point fuzzification.

We construct a subset $Y(U)$ of $X(U)$ for which we will prove the universal approximation property, in particular we will prove that, if U is compact, $Y(U)$ is dense in $C(U)$, the space of continuous real valued functions on a compact $U \subseteq \mathbb{R}^n$ with respect to the ∞-norm.

Let $U \subseteq \mathbb{R}^n$ be a compact, $Y(U) \subset X(U)$ is the function space consisting of all fuzzy systems $f \in X$ such that

- The input fuzzy sets have membership functions that are Gaussian functions of the form
$$\mu_{A_i^j}(x) = a_i^j exp\left(-\frac{1}{2}\left(\frac{x - \overline{x}_i^j}{\sigma_i^j}\right)^2\right) \tag{3}$$
with $\overline{x}_i^j \in \mathbb{R}, 0 < a_i^j \leqslant 1, \sigma_i^j \in (0, \infty)$
- The output fuzzy sets have membership functions that are Gaussian functions of the form
$$\mu_{B^k}(z) = a_0^k exp\left(-\frac{1}{2}\left(\frac{z - \overline{z}^k}{\sigma_0^k}\right)^2\right)$$
with $\overline{z}^k \in \mathbb{R}, 0 < a_0^k \leqslant 1, \sigma_0^k \in (0, \infty)$
- The fuzzy inference algorithm is the product inference
- The defuzzification algorithm is the centroid defuzzification [WM+92]

Lemma 4.2. *Let $U \subset \mathbb{R}^n$ be a compact, then $Y(U)$ is the set consisting of all functions $f : U \to \mathbb{R}$ of the form:*

$$f(x) = \frac{\sum_{k=1}^{l} \overline{z}^{j_{0,k}} \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)}{\sum_{k=1}^{l} \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)} \tag{4}$$

Moreover $Y(U) \subseteq C(U)$.

Proof. From the previous definition of $Y(U)$, an element $f \in Y(U)$ is of the form

$$f = D_K \circ I_K \circ F_K : U \to \mathbb{R}$$

where

[4] At this point we can define our knowledge base

$$K = \{A_i^j : 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant m_i\} \cup \{B^k \in \mathcal{F}(\mathbb{R}) : 1 \leqslant k \leqslant m_0\} \cup \{R_k : 1 \leqslant k \leqslant l\}$$

- $K = \{A_i^j : 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant m_i\} \cup \{B^k \in \mathcal{F}(\mathbb{R}) : 1 \leqslant k \leqslant m_0\} \cup \{R_k = A_1^{j_{1,k}} \times \cdots \times A_n^{j_{n,k}} \implies B^{j_{0,k}} : k = 1, \dots, l\}$
- $F_K : U \to \mathcal{F}(U), I_K : \mathcal{F}(U) \to \mathcal{F}(\mathbb{R})^l, D_K : \mathcal{F}(\mathbb{R})^l \to \mathbb{R}$ are, respectively, the the point fuzzification, the product inference and the centroid defuzzification.

According to product inference algorithm,

$$\forall A \in \mathcal{F}(U), I_K(A) = (A \circ R_1, \dots, A \circ R_l)$$

where $\forall k = 1, \dots, l, \forall z \in \mathbb{R}$

$$\begin{aligned}\mu_{A \circ R_k}(z) &= \sup_{x' \in U}[\mu_A(x')\mu_{R_k}(x', z)] = \\ &= \sup_{x' \in U}[\mu_A(x')\mu_{A_1^{j_{1,k}}}(x'_1) \cdots \mu_{A_n^{j_{n,k}}}(x'_n)\mu_{B^{j_{0,k}}}(z)]\end{aligned}$$

Let $A_x = F_K(x)$ be the fuzzy singleton of x, then $I_K \circ F_K(x) = I_K(A_x) = (A_x \circ R_1, \dots, A_x \circ R_l)$ where $\forall k = 1, \dots, l, \forall z \in \mathbb{R}$

$$\begin{aligned}\mu_{A_x \circ R_k}(z) &= \sup_{x' \in U}[\mu_{A_x}(x')\mu_{A_1^{j_{1,k}}}(x'_1) \cdots \mu_{A_n^{j_{n,k}}}(x'_n)\mu_{B^{j_{0,k}}}(z)] = \\ &= \mu_{A_x}(x)\mu_{A_1^{j_{1,k}}}(x_1) \cdots \mu_{A_n^{j_{n,k}}}(x_n)\mu_{B^{j_{0,k}}}(z) = \\ &= \mu_{A_1^{j_{1,k}}}(x_1) \cdots \mu_{A_n^{j_{n,k}}}(x_n)\mu_{B^{j_{0,k}}}(z) = \\ &= \mu_{B^{j_{0,k}}}(z) \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)\,.\end{aligned}$$

In general, centroid defuzzification $D_K : \mathcal{F}(\mathbb{R})^l \to \mathbb{R}$ takes as input a l-tuple $(C_1, \dots, C_l)$ of fuzzy sets of $\mathbb{R}$ and computes a weighted sum of the centroids [5] of each C_k.

In this case, centroid defuzzification means that the nonfuzzy output of the fuzzy system for input x is a weighted sum of the centroids of $A_x \circ R_1, \dots, A_x \circ R_l$ where the weights are determined by the product inference as $\prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)$.[6]

In conclusion,

$$f(x) = D_K \circ I_K \circ F_K(x) = D_K(A_x \circ R_1, \dots, A_x \circ R_l) = \frac{\sum_{k=1}^{l} c_k \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)}{\sum_{k=1}^{l} \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)}$$

[5]For a continuous membership function $\mu : \mathbb{R} \to [0, 1]$, the *centroid* is

$$\frac{\displaystyle\int_{\mathbb{R}} z\,\mu(z)\,dz}{\displaystyle\int_{\mathbb{R}} \mu(z)\,dz}$$

[6]If we view the fuzzy inference machine and defuzzification interface as an integrated part, then product inference can be explained as that the weight of rule R_k to the contribution of determining the output of fuzzy system for input $x \in U$ equals $\prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)$.

where c_k is the centroid of $A_x \circ R_k$. On the other hand, for each $k = 1, \ldots, l$,

$$\mu_{A_x \circ R_k}(z) = \mu_{B^{j_{0,k}}}(z) \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i) =$$

$$= \left[a_0^{j_{0,k}} \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i) \right] exp \left(-\frac{1}{2} \left(\frac{z - \overline{z}^{j_{0,k}}}{\sigma_0^{j_{0,k}}} \right)^2 \right) > 0$$

since $0 < a_0^{j_{0,k}} \leqslant 1, 0 < \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i) \leqslant 1$.
As a result, $\forall k = 1, \ldots l, c_k = \overline{z}^{j_{0,k}}$ (see Proposition A.1) and

$$f(x) = \frac{\sum_{k=1}^{l} \overline{z}^{j_{0,k}} \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)}{\sum_{k=1}^{l} \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)}.$$

The function f is well-defined and continuous on U because $\forall i = 1, \ldots, n, \forall k = 1, \ldots, l, \mu_{A_i^{j_{i,k}}}(x_i)$ is a Gaussian function so, is non zero and continuous in the i-th component of x. This means $Y(U) \subseteq C(U)$. □

Theorem 4.3 (Wang)**.** *Let $U \subset \mathbb{R}^n$ be a compact universe of discourse, $Y(U)$ is dense in $C(U)$ with respect to ∞-norm, that is*

$$\forall\, g \in C(U)\ \forall \varepsilon > 0\ \exists f \in\ Y(U) : d_\infty(f, g) = \sup_{x \in U} |f(x) - g(x)| < \varepsilon$$

Proof. Let us denote $Y = Y(U)$, we want to prove that Y is not empty. For each $1 \leqslant i \leqslant n$ we can choose $m_i \in \mathbb{N}$ and the parameters $\overline{x}_i^j, \sigma_i^j, a_i^j$ for each $j \in \{1, \ldots, m_i\}$, we define the input fuzzy sets with membership functions $\mu_{A_i^j} : U_i \to \mathbb{R}$ of the form Equation 3; we can choose also $m_0 \in \mathbb{N}$ and $\mu_{B^k} : \mathbb{R} \to \mathbb{R}$ for each $k \in \{1, \ldots, m_0\}$ in the same way. We can choose $l \in \mathbb{N}$ and l rules of the form Equation 2, this way we built the knowledge base

$$K = \{A_i^j : 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant m_i\} \cup \{B^k \in \mathcal{F}(V) : 1 \leqslant k \leqslant m_0\} \cup \{R_k : 1 \leqslant k \leqslant l\}.$$

The fuzzification algorithm F, the fuzzy inference algorithm I, the defuzzification algorithm D are fixed in Y. Since the fuzzification algorithm is independent on the knowledge base we denote the fuzzification interface as F, with K we determine the fuzzy inference machine I_K and the defuzzification function D_K. Finally $f = D_K \circ I_K \circ F_K \in Y$.
For Lemma 4.2, Y is a set of real continuous functions, we want to show that Y satisfies the conditions of the Theorem 4.1.
In order to prove that Y is an algebra on $\mathbb{R}$, we just need to prove that is closed under its operations: sum of real functions, product of real functions and the scalar multiplication of a real function, because the other properties are inherited from the algebra $C(U)$ in which Y is contained. Let $f_1, f_2 \in Y$, then they are of the form Equation 4

$$f_1(x) = \frac{\sum_{k=1}^{l_1} \overline{z}^{j_{0,k}} \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)}{\sum_{k=1}^{l_1} \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)},$$

$$f_2(x) = \frac{\sum_{s=1}^{l_2} \overline{y}^{\,t_{0,s}} \prod_{i=1}^{n} \mu_{C_i^{t_{i,s}}}(x_i)}{\sum_{s=1}^{l_2} \prod_{i=1}^{n} \mu_{C_i^{t_{i,s}}}(x_i)}\,,$$

where A_i^j and C_i^s are the input fuzzy sets, l_1 and l_2 are the number of fuzzy rules, respectively of f_1, f_2. We have For notational simplicity set

$$W_k(x) = \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)\,, \quad V_s(x) = \prod_{i=1}^{n} \mu_{C_i^{t_{i,s}}}(x_i)\,.$$

Then

$$f_1(x) = \frac{\sum_{k=1}^{l_1} \overline{z}^{\,j_{0,k}}\, W_k(x)}{\sum_{k=1}^{l_1} W_k(x)}, \quad f_2(x) = \frac{\sum_{s=1}^{l_2} \overline{y}^{\,t_{0,s}}\, V_s(x)}{\sum_{s=1}^{l_2} V_s(x)},$$

so

$$\begin{aligned}
f_1(x) + f_2(x) &= \frac{\sum_{k=1}^{l_1} \overline{z}^{\,j_{0,k}} W_k(x)}{\sum_{k=1}^{l_1} W_k(x)} + \frac{\sum_{s=1}^{l_2} \overline{y}^{\,t_{0,s}} V_s(x)}{\sum_{s=1}^{l_2} V_s(x)} \\
&= \frac{\left(\sum_{k=1}^{l_1} \overline{z}^{\,j_{0,k}} W_k\right)\left(\sum_{s=1}^{l_2} V_s\right) + \left(\sum_{s=1}^{l_2} \overline{y}^{\,t_{0,s}} V_s\right)\left(\sum_{k=1}^{l_1} W_k\right)}{\sum_{k=1}^{l_1} W_k \;\sum_{s=1}^{l_2} V_s} \\
&= \frac{\displaystyle\sum_{k=1}^{l_1}\sum_{s=1}^{l_2}\left(\overline{z}^{\,j_{0,k}} W_k V_s + \overline{y}^{\,t_{0,s}} W_k V_s\right)}{\displaystyle\sum_{k=1}^{l_1}\sum_{s=1}^{l_2} W_k V_s} = \frac{\displaystyle\sum_{k=1}^{l_1}\sum_{s=1}^{l_2}\left(\overline{z}^{\,j_{0,k}} + \overline{y}^{\,t_{0,s}}\right) W_k V_s}{\displaystyle\sum_{k=1}^{l_1}\sum_{s=1}^{l_2} W_k V_s} \\
&= \frac{\displaystyle\sum_{k=1}^{l_1}\sum_{s=1}^{l_2}\left(\overline{z}^{\,j_{0,k}} + \overline{y}^{\,t_{0,s}}\right)\prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)\mu_{C_i^{t_{i,s}}}(x_i)}{\displaystyle\sum_{k=1}^{l_1}\sum_{s=1}^{l_2}\prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)\mu_{C_i^{t_{i,s}}}(x_i)} \\
&= \frac{\displaystyle\sum_{k=1}^{l_1}\sum_{s=1}^{l_2}\left(\overline{w}^{\,h_{0,k,s}}\right)\prod_{i=1}^{n} \mu_{E_i^{h_{i,k,s}}}(x_i)}{\displaystyle\sum_{k=1}^{l_1}\sum_{s=1}^{l_2}\prod_{i=1}^{n} \mu_{E_i^{h_{i,k,s}}}(x_i)}\,.
\end{aligned}$$

For each $i \in \{1, \ldots, n\}$ and for each pair $(k, s) \in \{1, \ldots, l_1\} \times \{1, \ldots, l_2\}$, we define a new rule whose antecedent $E_i^{h_{i,k,s}}$ has membership function $\mu_{E_i^{h_{i,k,s}}}(x_i) = \mu_{A_i^{j_{i,k}}}(x_i)\,\mu_{C_i^{t_{i,s}}}(x_i) : U_i \to \mathbb{R}$, for Proposition A.2, it is a function of the form Equation 3. For each pair (k, s), we can choose any output fuzzy set $F^{(k,s)}$ such that $\mu_{F^{(k,s)}}$ is a gaussian function of the form Equation 3 and its mean is $\overline{w}^{\,h_{0,k,s}}$, for Proposition A.1, its centroid is $w^{h_{0,k,s}}$, therefore $f_1 + f_2 \in Y$.

Similarly,

$$f_1(x)\,f_2(x) = \frac{\sum_{k=1}^{l_1} \overline{z}^{\,j_{0,k}} W_k}{\sum_{k=1}^{l_1} W_k} \cdot \frac{\sum_{s=1}^{l_2} \overline{y}^{\,t_{0,s}} V_s}{\sum_{s=1}^{l_2} V_s} = \frac{\left(\sum_{k=1}^{l_1} \overline{z}^{\,j_{0,k}} W_k\right)\left(\sum_{s=1}^{l_2} \overline{y}^{\,t_{0,s}} V_s\right)}{\sum_{k=1}^{l_1} W_k \ \sum_{s=1}^{l_2} V_s}$$

$$= \frac{\displaystyle\sum_{k=1}^{l_1}\sum_{s=1}^{l_2} \overline{z}^{\,j_{0,k}} W_k \,\overline{y}^{\,t_{0,s}} V_s}{\displaystyle\sum_{k=1}^{l_1}\sum_{s=1}^{l_2} W_k V_s} = \frac{\displaystyle\sum_{k=1}^{l_1}\sum_{s=1}^{l_2} \left(\overline{z}^{\,j_{0,k}}\ \overline{y}^{\,t_{0,s}}\right) W_k V_s}{\displaystyle\sum_{k=1}^{l_1}\sum_{s=1}^{l_2} W_k V_s}.$$

$$= \frac{\displaystyle\sum_{k=1}^{l_1}\sum_{s=1}^{l_2} \left(\overline{z}^{\,j_{0,k}}\ \overline{y}^{\,t_{0,s}}\right) \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)\mu_{C_i^{t_{i,s}}}(x_i)}{\displaystyle\sum_{k=1}^{l_1}\sum_{s=1}^{l_2}\prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)\mu_{C_i^{t_{i,s}}}(x_i)}$$

$$= \frac{\displaystyle\sum_{k=1}^{l_1}\sum_{s=1}^{l_2} \left(\overline{w}^{\,h_{0,k,s}}\right) \prod_{i=1}^{n} \mu_{E_i^{h_{i,k,s}}}(x_i)}{\displaystyle\sum_{k=1}^{l_1}\sum_{s=1}^{l_2}\prod_{i=1}^{n} \mu_{E_i^{h_{i,k,s}}}(x_i)}$$

We introduce rules with the same antecedents $E_i^{h_{i,k,s}}$ as before and consequents $F^{(k,s)}$ with gaussian membership and such that its centroid is $\overline{w}^{\,h_{0,k,s}} = \overline{z}^{\,j_{0,k}}\ \overline{y}^{\,t_{0,s}}$, we conclude that $f_1 f_2 \in Y$.
Finally, for any $\alpha \in \mathbb{R}$,

$$\alpha\, f_1(x) = \frac{\sum_{k=1}^{l_1} (\alpha\, \overline{z}^{\,j_{0,k}}) \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)}{\sum_{k=1}^{l_1} \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)} = \frac{\sum_{k=1}^{l_1} (\overline{w}^{\,j_{0,k}}) \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)}{\sum_{k=1}^{l_1} \prod_{i=1}^{n} \mu_{A_i^{j_{i,k}}}(x_i)},$$

so we can consider the same fuzzy system, but with consequents of the rules such that their centroids are $\overline{w}^{\,j_{0,k}} = \alpha \overline{z}^{\,j_{0,k}}$, shows $\alpha f_1 \in Y$.
Now we prove that Y separates points on U. Let $x^0, y^0 \in U$ be such that $x^0 \neq y^0$. Then there exist $i \in \{1,\dots,n\}$ such that $x_i^0 \neq y_i^0$. For each $1 \leqslant i \leqslant n$, in the i-th subspace of U, U_i, we define two fuzzy sets with membership functions

$$\mu_{A_i^1}(x_i) = \exp\left[-\frac{(x_i - x_i^0)^2}{2}\right], \quad \mu_{A_i^2}(x_i) = \exp\left[-\frac{(x_i - y_i^0)^2}{2}\right],$$

notice that if $x_i^0 = y_i^0$ then $A_i^1 = A_i^2$. In the output universe $\mathbb{R}$ we define two fuzzy sets with Gaussians memebership functions

$$\mu_{B^j}(z) = \exp\left[-\frac{(z - z^j)^2}{2}\right] \quad \forall\, j \in \{1,2\}$$

where z^1, z^2 will be chosen below. Let the rules consist of

$$R_1:\ A_1^1 \times \cdots \times A_n^1 \implies B^1, \qquad R_2:\ A_1^2 \times \cdots \times A_n^2 \implies B^2,$$

so that $f \in Y$ and has the form

$$f(x) = \frac{z^1 \prod_{k=1}^n \mu_{A_k^1}(x_k) + z^2 \prod_{k=1}^n \mu_{A_k^2}(x_k)}{\prod_{k=1}^n \mu_{A_k^1}(x_k) + \prod_{k=1}^n \mu_{A_k^2}(x_k)}\,.$$

Set

$$\alpha = \frac{1}{1 + \prod_{k=1}^{n} \exp\bigl[-(x_k^0 - y_k^0)^2/2\bigr]},$$

so that at x^0 and y^0 one finds

$$f(x^0) = \frac{z^1 + z^2 \prod_{k=1}^n \mu_{A_k^2}(x_k^0)}{1 + \prod_{k=1}^n \mu_{A_k^2}(x_k^0)} = \alpha\, z^1 + (1-\alpha)\, z^2,$$

$$f(y^0) = \frac{z^1 \prod_{k=1}^n \mu_{A_k^1}(y_k^0) + z^2}{\prod_{k=1}^n \mu_{A_k^1}(y_k^0) + 1} = \alpha\, z^2 + (1-\alpha)\, z^1\,.$$

Notice that

$$x_i^0 = y_i^0 \quad \forall\, i \in \{1, \ldots, n\} \iff \prod_{k=1}^{n} \exp\bigl[-(x_k^0 - y_k^0)^2/2\bigr] = 1 \iff \alpha = \frac{1}{2}\,,$$

so $\alpha \neq 1/2$, that is $\alpha \neq 1 - \alpha$. Choosing $z^1 = 0$, $z^2 = 1$ gives

$$f(x^0) = 1 - \alpha \neq \alpha = f(y^0)\,,$$

so Y separates points on U.
Finally we need to prove that Y vanishes at no point in U. If we choose $\overline{z}^{j_{0,k}} > 0, \forall k = 1, \ldots, l$, then $\forall x \in U$,

$$f(x) = \frac{\sum_{k=1}^l \overline{z}^{j_{0,k}} \prod_{i=1}^n \mu_{A_i^{j_{i,k}}}(x_i)}{\sum_{k=1}^l \prod_{i=1}^n \mu_{A_i^{j_{i,k}}}(x_i)} \neq 0\,.$$

Now we can apply the Theorem 4.1 on (Y, d_∞) and conclude that it is dense in $C(U)$. □

If $U \subset \mathbb{R}^n$ is a compact, then $C(U) \subseteq L_2(U)$ and we can generalize the above theorem to $L_2(U) = \{g : U \to \mathbb{R} : \int_U |g(x)|^2 dx < \infty\}$.
Corollary 4.3.1. *Let $U \subset \mathbb{R}^n$ be a compact, $Y(U)$ is dense in $L_2(U)$ with respect to L_2-norm, i.e. $\forall g \in L_2(U), \forall \varepsilon > 0, \exists f \in Y(U) : (\int_U |f(x) - g(x)|^2 dx)^{\frac{1}{2}} < \varepsilon$*

Proof. Continuous functions on U form a dense subset of $L_2(U)$ with respect to L_2-norm, that is $\forall g \in L_2(U), \forall \varepsilon > 0, \exists\ \overline{g} \in C(U) : (\int_U |\overline{g}(x) - g(x)|^2 dx)^{\frac{1}{2}} < \frac{\varepsilon}{2}$.
On the other hand, if $\overline{g} \in C(U)$, then $\exists f \in Y(U) : \sup_{x \in U} |f(x) - \overline{g}(x)| < \frac{\varepsilon}{2V^{\frac{1}{2}}}$ where $V = \int_U dx < \infty$ because U is compact. Hence, we have

$$\|f - g\|_2 \leqslant \|f - \overline{g}\|_2 + \|\overline{g} - g\|_2 = \left(\int_U |f(x) - \overline{g}(x)|^2 dx\right)^{\frac{1}{2}} + \left(\int_U |\overline{g}(x) - g(x)|^2 dx\right)^{\frac{1}{2}} \leqslant$$

$$\leqslant \left(\int_U (\sup_{x \in U} |f(x) - \overline{g}(x)|)^2 dx\right)^{\frac{1}{2}} + \frac{\varepsilon}{2} \leqslant$$

$$\leqslant \left(\int_U (\frac{\varepsilon}{2V^{\frac{1}{2}}})^2 dx\right)^{\frac{1}{2}} + \frac{\varepsilon}{2} = \left(\frac{\varepsilon^2 V}{2^2 V}\right)^{\frac{1}{2}} + \frac{\varepsilon}{2} = 2\frac{\varepsilon}{2} = \varepsilon$$

□

4.2 Additive Fuzzy System (AFS)

In this section we introduce the class of Additive Fuzzy Systems (AFS) ([TRK15], [WM+92]), prove that they are universal approximators ([Kos94]) and show that Wang's theorem can be seen as a special case.

Definition 4.3 (Additive Centroidal Fuzzy System)**.** An *Additive Centroidal Fuzzy System*, or *Additive fuzzy system* for short, is a rule based fuzzy system $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^p$ that maps inputs to outputs by summing fired then-parts sets and then taking the centroid of the sum.

More precisely, an additive fuzzy system is a fuzzy system

$$f = D_K \circ I_K \circ F_K : U \subseteq \mathbb{R}^n \to \mathbb{R}^p$$

where

- $\mathcal{K} = \mathcal{F}(U \times \mathbb{R}^p)^s$ for some $s \in \mathbb{N}$
- $K \in \mathcal{K}$ is a set of s fuzzy implication relations on $U \times R^p$, i.e.

$$K = \{R_i = A_i \implies B_i : i = 1, \ldots, s\}$$

  where $\forall i = 1, \ldots, s, A_i \in \mathcal{F}(U), B_i \in \mathcal{F}(\mathbb{R}^p)$
- $F_K : U \to \mathcal{F}(U)$ is a certain fuzzification interface
- $I_K : \mathcal{F}(U) \to \mathcal{F}(\mathbb{R}^p)$ is a fuzzy inference that we denote $\forall A \in \mathcal{F}(U)$,

$$I_K(A) = \sum_{i=1}^{s} w_i(A) B_i^A$$

  that is, $\forall y \in \mathbb{R}^p$

$$\mu_{I_K(A)}(y) = \sum_{i=1}^{s} w_i(A) \mu_{B_i^A}(y)$$

  where $\forall i = 1, \ldots, s,\ w_i(A) > 0$ is a scalar depending on A and $B_i^A \in \mathcal{F}(\mathbb{R}^p)$ is a fuzzy set depending on K and A, called *i-th fired then-part set* (or *fired then-part set of i-th rule* R_i or *fired* B_i for short)[7]. We denote $\forall x \in U$

$$\mu_{I_K \circ F_K(x)} = \sum_{i=1}^{s} w_i(x) \mu_{B_i^x}$$

  the output of $I_K \circ F_K$ for input x.
- $D_K : \mathcal{F}(\mathbb{R}^p) \to \mathbb{R}^p$ is the centroid defuzzification, i.e. $B \in \mathcal{F}(\mathbb{R}^p)$

$$D_K(B) = \frac{\displaystyle\int_{\mathbb{R}^p} y \mu_B(y) dy}{\displaystyle\int_{\mathbb{R}^p} \mu_B(y) dy} = \left(\frac{\displaystyle\int_{\mathbb{R}^p} y_1 \mu_B(y) dy}{\displaystyle\int_{\mathbb{R}^p} \mu_B(y) dy}, \ldots, \frac{\displaystyle\int_{\mathbb{R}^p} y_p \mu_B(y) dy}{\displaystyle\int_{\mathbb{R}^p} \mu_B(y) dy} \right).$$

[7]Note that it is not required that the weights $\{w_i(A)\}_i = 1, \ldots, s$ sum to unity. According to the definitions given so far, in general $I_K(A)$ isn't a fuzzy set since $\mu_{I_K(A)}$ is $[0, +\infty)-$valued and not $[0, 1]-$valued. However we can consider a generalized membership function that is a $[0, +\infty)$-valued function.

It is required an additional condition: $\forall x \in U$, $\forall i = 1, \ldots, s$, B_i^x is a fuzzy set of $\mathbb{R}^p$ with integrable membership function and

$$\int_{\mathbb{R}^p} \mu_{B_j^x}(y)dy > 0\,,$$

for some $j = 1, \ldots, s$.

Remark 4.2. Note that the previous condition is required for the subsequent defuzzification interface. In fact, the condition ensures that $\forall x \in U$, $\int_{\mathbb{R}^p} \mu_{I_K \circ F_K(x)}(y)dy > 0$ and it makes sense to consider the centroid of $\mu_{I_K \circ F_K(x)}$. Moreover, this condition can be achieved by requiring a combination of properties on the knowledge base K, the fuzzification interface F_K and the inference I_K.

Putting the pieces together, we have $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^p$ such that $\forall x \in U$

$$f(x) = \frac{\displaystyle\int_{\mathbb{R}^p} \sum_{i=1}^{s} w_i(x) y \mu_{B_i^x}(y)dy}{\displaystyle\int_{\mathbb{R}^p} \sum_{i=1}^{s} w_i(x) \mu_{B_i^x}(y)dy}$$

where $\forall i = 1, \ldots, s,\ w_i(x) > 0$ is a scalar depending on x, $B_i^x \in \mathcal{F}(\mathbb{R}^p)$ is the B_i *fired by* x and $\int_{\mathbb{R}^p} \mu_{B_j^x}(y)dy > 0$ for some $j = 1, \ldots, s$.

This additive structure produces a simple convex-sum structure: outputs are convex combinations of the centroids of the fired then-part sets.

Lemma 4.4. *Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^p$ be an additive fuzzy system.*
If $\forall x \in U, \forall i = 1 \ldots, s$,

$$\int_{\mathbb{R}^p} \mu_{B_i^x}(y)dy > 0$$

[8] *then*
$\forall x \in U$,

$$f(x) = \sum_{i=1}^{s} p_i(x) c_i(x)$$

where $\forall x \in U,\ \forall i = 1, \ldots, s, p_i(x) > 0, \sum_{i=1}^{s} p_i(x) = 1$ [9] and $c_i(x)$ is the centroid of B_i fired by x, i.e.

$$c_i(x) = \frac{\int_{\mathbb{R}^p} y \mu_{B_i^x}(y)dy}{\int_{\mathbb{R}^p} \mu_{B_i^x}(y)dy} \in \mathbb{R}^p$$

Proof. $\forall x \in U$,

$$f(x) = \frac{\displaystyle\int_{\mathbb{R}^p} \sum_{i=1}^{s} w_i(x) y \mu_{B_i^x}(y)dy}{\displaystyle\int_{\mathbb{R}^p} \sum_{i=1}^{s} w_i(x) \mu_{B_i^x}(y)dy} = \frac{\displaystyle\sum_{i=1}^{s} w_i(x) \int_{\mathbb{R}^p} y \mu_{B_i^x}(y)dy}{\displaystyle\sum_{i=1}^{s} w_i(x) \int_{\mathbb{R}^p} \mu_{B_i^x}(y)dy} =$$

$$= \frac{\displaystyle\sum_{i=1}^{s} \left[w_i(x) \int_{\mathbb{R}^p} \mu_{B_i^x}(y)dy \right] c_i(x)}{\displaystyle\sum_{i=1}^{s} w_i(x) \int_{\mathbb{R}^p} \mu_{B_i^x}(y)dy} = \sum_{i=1}^{s} p_i(x) c_i(x)$$

[8]This hypotesis is necessary to ensure that $\forall i = 1, \ldots, s$ is well defined the centroid of $\mu_{B_i^x}$

[9]equivalently, $p_i : U \to [0, 1] : \sum_{i=1}^{s} p_i = 1$

where $\forall i = 1, \ldots, s \ \forall x \in U$

$$p_i(x) = \frac{w_i(x) \int_{\mathbb{R}^p} \mu_{B_i^x}(y)dy}{\sum_{i=1}^{s} w_i(x) \int_{\mathbb{R}^p} \mu_{B_i^x}(y)dy}$$

Obviously, $\forall x \in U, \ \sum_{i=1}^{s} p_i(x) = 1$.
Moreover, $\forall x \in U, p_i(x) > 0$ because $w_i(x), \int_{\mathbb{R}^p} \mu_{B_i^x}(y)dy > 0$ by hypotheses. □

Notation 4.4. The class of all additive fuzzy systems from $U \subseteq \mathbb{R}^n$ to $\mathbb{R}^p$ is denoted by $AFS(U, \mathbb{R}^p)$.
We now consider two particular classes of additive fuzzy systems.
The first one, denoted by $AFS_{prod}(U, \mathbb{R}^p)$, consists of all additive fuzzy systems $f \in AFS(U, \mathbb{R}^p)$ such that:

$$\forall A \in \mathcal{F}(U), \mu_{I_K(A)} = \sum_{i=1}^{s} w_i(A) a_i(A, K) \mu_{B_i}$$

where $\forall i = 1, \ldots, s, \ a_i(A, K) \in [0, 1]$.
The second one, denoted by $AFS_{min}(U, \mathbb{R}^p)$, consists of all additive fuzzy systems $f \in AFS(U, \mathbb{R}^p)$ such that:

$$\forall A \in \mathcal{F}(U), \mu_{I_K(A)} = \sum_{i=1}^{s} w_i(A) \min(a_i(A, K), \mu_{B_i})$$

where $\forall i = 1, \ldots, s, \ a_i(A, K) \in [0, 1]$.
In both cases, the scalars $a_i(A, K) \in [0, 1]$ have the meaning of *activation value of i-th rule R_i in A* or *i-th activation value in A*.
Remark 4.3. Typically, the i-th activation value of A is determined as the inner product between A and the antecedent A_i of i-th rule, that is

$$a_i(A, K) = A \circ A_i = \bigvee_{x \in U} \mu_A(x) \wedge \mu_{A_i}(x)$$

where $\vee$ and $\wedge$ are, respectively, a t-conorm and t-norm.
Theorem 4.5 (Kosko)**.** *Let $U \subset \mathbb{R}^n$ be a compact, then $AFS_{prod}(U, \mathbb{R}^p)$ and $AFS_{min}(U, \mathbb{R}^p)$ are dense in $C(U, \mathbb{R}^p)$ with respect to ∞-norm.*
Remark 4.4. In the proof, we denote by $|\cdot|$ both the 2-norm in $\mathbb{R}^n$ and 2-norm in $\mathbb{R}^p$. It will be clear from the argument if it refers to $\mathbb{R}^n$ or $\mathbb{R}^p$

Proof. Let $f \in C(U, \mathbb{R}^p)$ then f is continuous on U compact, i.e. f is uniformly continuous, that is

$$\forall \varepsilon > 0, \exists \delta > 0 : \forall x_1, x_2 \in U : |x_1 - x_2| < \delta, |f(x_1) - f(x_2)| < \varepsilon$$

Since U is compact, it is possible to cover U with a finite family of open cubes, each having center in U and diameter $< \frac{\delta}{2}$. In fact, $\forall x \in U$, let consider the open cube $M_x = (x_1 - \rho, x_1 + \rho) \times \cdots \times (x_n - \rho, x_n + \rho)$ with $0 < \rho < \frac{\delta}{4\sqrt{n}}$ then we have

- $U \subseteq \bigcup_{x \in U} M_x$

- $\forall x \in U, diam(M_x) = 2\rho\sqrt{n} < 2\sqrt{n}\frac{\delta}{4\sqrt{n}} = \frac{\delta}{2}$

$\{M_x : x \in U\}$ is an open cover of the compact U, then a finite subcover of U exists, i.e. $\exists x_1, \ldots, x_s \in U : M_{x_i} = M_i, U \subseteq \bigcup_{i=1}^{s} M_i$ and $\forall i = 1, \ldots, s,\ M_i$ is an open cube with center $x_i \in U$ and $diam(M_i) < \frac{\delta}{2}$.
As a consequence,

- $\forall i = 1, \ldots, s, \forall u, w \in M_i \cap U, |u - w| \leqslant diam(M_i) < \frac{\delta}{2} < \delta$ and then $|f(u) - f(w)| < \varepsilon$
- $\forall j, k = 1, \ldots, s : M_j \cap M_k \neq \varnothing, \forall u \in M_j \cap U, w \in M_k \cap U, |u - w| < \delta$ and then $|f(u) - f(w)| < \varepsilon$

In particular,

- $\forall x \in U, \exists i = 1, \ldots, s : x \in M_i$ and $\forall i = 1, \ldots, s : x \in M_i, |f(x) - f(x_i)| < \varepsilon$
- $\forall j, k = 1, \ldots, s : M_j \cap M_k \neq \varnothing, |f(x_i) - f(x_j)| < \varepsilon$

Let consider a fuzzy system $F \in AFS_{prod}(U, \mathbb{R}^p)$ with the following properties:

- the knowledge base is $K = \{R_i = A_i \implies B_i : i = 1, \ldots, s\}$ where $\forall i = 1, \ldots, s, A_i \in \mathcal{F}(U) : \mu_{A_i}(x) \neq 0 \iff x \in M_i$ and $B_i \in \mathcal{F}(\mathbb{R}^p) : \mu_{B_i}$ has centroid in $f(x_i)$, i.e.

$$\int_{\mathbb{R}^p} \mu_{B_i}(y)dy > 0$$

  and

$$C(\mu_{B_i}) = \frac{\displaystyle\int_{\mathbb{R}^p} y\mu_{B_i}(y)dy}{\displaystyle\int_{\mathbb{R}^p} \mu_{B_i}(y)dy} = f(x_i)$$

- $F_K : U \to \mathcal{F}(U)$ is the point fuzzification
- I_K is the following inference machine:

$$\forall A \in \mathcal{F}(U), \mu_{I_K(A)} = \sum_{i=1}^{s} \mu_{B_i^A}$$

  where $\forall i = 1, \ldots s, \mu_{B_i^A} = a_i(A, K)\mu_{B_i}$ and $a_i(A, K) = \sup_{x \in U} \mu_A(x)\mu_{A_i}(x)$

It follows that $\forall x' \in U$, let $A_{x'} = F_K(x')$, then $\forall i = 1, \ldots, s$

$$a_i(A_{x'}, K) = \mu_{A_i}(x')$$

and then

$$\mu_{I_K(A_{x'})} = \sum_{i=1}^{s} \mu_{B_i^{x'}} = \sum_{i=1}^{s} \mu_{A_i}(x')\mu_{B_i}$$

Since $\forall x' \in U, \exists i = 1, \ldots, s : x' \in M_i$, it follows that $\exists i = 1, \ldots, s : \mu_{A_i}(x') > 0$ and then

$$\int_{\mathbb{R}^p} \mu_{A_i}(x')\mu_{B_i}(y)dy > 0$$

.

By applying the centroid defuzzification D_K we obtain that

$$\begin{aligned}F(x') &= \frac{\displaystyle\sum_{i=1}^{s}\int_{\mathbb{R}^p} y\mu_{B_i^{x'}}(y)dy}{\displaystyle\sum_{i=1}^{s}\int_{\mathbb{R}^p} \mu_{B_i^{x'}}(y)dy} = \\ &= \frac{\displaystyle\sum_{i=1}^{s}\mu_{A_i}(x')\int_{\mathbb{R}^p} y\mu_{B_i}(y)dy}{\displaystyle\sum_{i=1}^{s}\mu_{A_i}(x')\int_{\mathbb{R}^p} \mu_{B_i}(y)dy} = \\ &= \frac{\displaystyle\sum_{i=1}^{s}\left[\mu_{A_i}(x')\int_{\mathbb{R}^p} \mu_{B_i}(y)dy\right]C(\mu_{B_i})}{\displaystyle\sum_{i=1}^{s}\mu_{A_i}(x')\int_{\mathbb{R}^p} \mu_{B_i}(y)dy} = \\ &= \frac{\displaystyle\sum_{i=1}^{s}\left[\mu_{A_i}(x')\int_{\mathbb{R}^p} \mu_{B_i}(y)dy\right]f(x_i)}{\displaystyle\sum_{i=1}^{s}\mu_{A_i}(x')\int_{\mathbb{R}^p} \mu_{B_i}(y)dy} = \\ &= \sum_{i=1}^{s} c_i(x')f(x_i)\end{aligned}$$

where $\forall i = 1, \ldots, s$

$$0 \leqslant c_i(x') = \frac{\displaystyle\mu_{A_i}(x')\int_{\mathbb{R}^p} \mu_{B_i}(y)dy}{\displaystyle\sum_{i=1}^{s}\mu_{A_i}(x')\int_{\mathbb{R}^p} \mu_{B_i}(y)dy} \leqslant 1$$

and they sum to unity, that is $F(x')$ is a convex combination of the centroids $f(x_i)$. On the other hand, $\mu_{A_i}(x') \neq 0 \iff x' \in M_i$ then $F(x')$ is a convex combination of the centroids $\{C(\mu_{B_i}) = f(x_i) : x' \in M_i\}$ i.e.

$$F(x') = \sum_{\substack{1\leqslant i\leqslant s,\\ x'\in M_i}} c_i(x')f(x_i)$$

with $0 < c_i(x')$ and $\sum_{i=1,x'\in M_i}^{s} c_i(x') = 1$. In conclusion, $\forall x' \in U$

$$\begin{aligned}
|F(x') - f(x')| &= \left| \sum_{\substack{1\leqslant i\leqslant s,\\ x'\in M_i}} c_i(x') f(x_i) - f(x') \right| = \\
&= \left| \sum_{\substack{1\leqslant i\leqslant s,\\ x'\in M_i}} c_i(x') f(x_i) - \sum_{\substack{1\leqslant i\leqslant s,\\ x'\in M_i}} c_i(x') f(x') \right| = \\
&= \left| \sum_{\substack{1\leqslant i\leqslant s,\\ x'\in M_i}} c_i(x')(f(x_i) - f(x')) \right| \leqslant \\
&\leqslant \sum_{\substack{1\leqslant i\leqslant s,\\ x'\in M_i}} c_i(x') \left| f(x_i) - f(x') \right| \leqslant \\
&\leqslant \sum_{\substack{1\leqslant i\leqslant s,\\ x'\in M_i}} c_i(x') \varepsilon \leqslant \\
&\leqslant \varepsilon \sum_{\substack{1\leqslant i\leqslant s,\\ x'\in M_i}} c_i(x') = \\
&= \varepsilon
\end{aligned} \tag{5}$$

This conclude the proof for $AFS_{prod}(U, \mathbb{R}^p)$.
For $AFS_{min}(U, \mathbb{R}^P)$, the proof is the same with the difference that the membership function μ_{B_i} is required to be symmetric and centered on $f(x_i)$ and

$$\forall A \in \mathcal{F}(U), \mu_{I_K(A)} = \sum_{i=1}^{s} \mu_{B_i^A}$$

where $\forall i = 1, \dots s, \mu_{B_i^A} = \min(a_i(A, K), \mu_{B_i})$ and $a_i(A, K) = \sup_{x\in U} \mu_A(x)\mu_{A_i}(x)$
It follows that $\forall x' \in U$, let $A_{x'} = F_K(x')$, then $\forall i = 1, \dots, s$

$$a_i(A_{x'}, K) = \mu_{A_i}(x')$$

and then

$$\mu_{I_K(A_{x'})} = \sum_{i=1}^{s} \mu_{B_i^{x'}} = \sum_{i=1}^{s} \min(\mu_{A_i}(x'), \mu_{B_i})$$

Since $\forall x' \in U, \exists i = 1, \dots, s : x' \in M_i$, it follows that $\exists i = 1, \dots, s : \mu_{A_i}(x') > 0$ and then

$$\int_{\mathbb{R}^p} \min(\mu_{A_i}(x'), \mu_{B_i})(y) dy > 0.$$

From the hypothesis of symmetry of μ_{B_i} it follows that if $\mu_{A_i}(x') > 0$ then $\mu_{B_i^{x'}} =$

$\min(\mu_{A_i}(x'), \mu_{B_i})$ has the same centroid of μ_{B_i} i.e. $C(\mu_{B_i^{x'}}) = f(x_i)$. Then, $\forall x' \in U$

$$F(x') = \frac{\sum_{i=1}^{s} \int_{\mathbb{R}^p} y\mu_{B_i^{x'}}(y)dy}{\sum_{i=1}^{s} \int_{\mathbb{R}^p} \mu_{B_i^{x'}}(y)dy} = $$

$$= \frac{\sum_{i=1}^{s} C(\mu_{B_i^{x'}}) \int_{\mathbb{R}^p} \mu_{B_i^{x'}}(y)dy}{\sum_{i=1}^{s} \int_{\mathbb{R}^p} \mu_{B_i^{x'}}(y)dy} =$$

$$= \frac{\sum_{i=1}^{s} f(x_i) \int_{\mathbb{R}^p} \mu_{B_i^{x'}}(y)dy}{\sum_{i=1}^{s} \int_{\mathbb{R}^p} \mu_{B_i^{x'}}(y)dy} =$$

$$= \sum_{i=1}^{s} c_i(x')f(x_i)$$

where $\forall i = 1, \ldots, s$

$$0 \leqslant c_i(x') = \frac{\int_{\mathbb{R}^p} \mu_{B_i^{x'}}(y)dy}{\sum_{i=1}^{s} \int_{\mathbb{R}^p} \mu_{B_i^{x'}}(y)dy} \leqslant 1$$

and they sum to unity, that is $F(x')$ is a convex combination of the centroids $f(x_i)$. However, $\mu_{A_i}(x') \neq 0 \iff x' \in M_i$ and if $\mu_{A_i}(x') = 0$ then

$$c_i(x') = \int_{\mathbb{R}^p} \mu_{B_i^{x'}}(y)dy = \int_{\mathbb{R}^p} \min(\mu_{A_i}(x'), \mu_{B_i})(y)dy = 0$$

It follows that $F(x')$ is a convex combination of the centroids $\{f(x_i) : x' \in M_i\}$ i.e.

$$F(x') = \sum_{\substack{1 \leqslant i \leqslant s, \\ x' \in M_i}} c_i(x')f(x_i)$$

with $0 \leqslant c_i(x')$ and $\sum_{i=1, x' \in M_i}^{s} c_i(x') = 1$. It is possible to conclude as in Equation 5. $\square$

Remark 4.5. We can observe that in the case of $AFS_{prod}(U, \mathbb{R}^p)$, we can construct a fuzzy system F that approximate the function f also by choosing the membership functions μ_{A_i} such that

$$\mu_{A_i}(x) = \begin{cases} < \dfrac{\varepsilon}{m_i \int_{\mathbb{R}^p} \mu_{B_i}(y)dy} & x \notin M_i \\ \geqslant \gamma_i & \forall x \in U \end{cases}$$

where $m_i = \max_{x \in U} |f(x) - f(x_i)|$,[10] which exists because f is continuous in U compact, and $0 < \gamma_i < 1$. In this case, still holds that $x \in M_i \implies \mu_{A_i}(x) > \gamma_i > 0$ and still holds that $\forall x' \in U$

$$F(x') = \sum_{i=1}^{s} c_i(x') f(x_i)$$

with

$$0 \leqslant c_i(x') = \frac{\mu_{A_i}(x') \displaystyle\int_{\mathbb{R}^p} \mu_{B_i}(y)dy}{\displaystyle\sum_{i=1}^{s} \mu_{A_i}(x') \int_{\mathbb{R}^p} \mu_{B_i}(y)dy} \leqslant 1, \ \sum_{i=1}^{s} c_i(x') = 1$$

but we can't conclude as in Equation 5. However, $\forall x' \in U$ we have that

$$0 \leqslant \sum_{\substack{1 \leqslant i \leqslant s, \\ x' \in M_i}} c_i(x') \leqslant 1$$

and $\forall i = 1, \ldots, s : x' \notin Mi$

$$c_i(x') < \frac{\varepsilon}{m_i \displaystyle\sum_{i=1}^{s} \gamma_i \int_{\mathbb{R}^p} \mu_{B_i}(y)dy}$$

[10]Note that if $m_i = 0$ for some $i = 1, \ldots, s$, f is constant in U and for any choice of μ_{A_i} we have $F = f$

We can conclude, $\forall x' \in U$

$$
\begin{aligned}
|F(x') - f(x')| &= \left| \sum_{i=1}^{s} c_i(x') f(x_i) - f(x') \right| = \\
&= \left| \sum_{i=1}^{s} c_i(x') f(x_i) - \sum_{i=1}^{s} c_i(x') f(x') \right| = \\
&= \left| \sum_{i=1}^{s} c_i(x') \left[f(x_i) - f(x') \right] \right| \leqslant \\
&\leqslant \left| \sum_{\substack{1 \leqslant i \leqslant s, \\ x' \in M_i}} c_i(x') \left[f(x_i) - f(x') \right] \right| + \left| \sum_{\substack{1 \leqslant i \leqslant s, \\ x' \notin M_i}} c_i(x') \left[f(x_i) - f(x') \right] \right| \leqslant \\
&\leqslant \sum_{\substack{1 \leqslant i \leqslant s, \\ x' \in M_i}} c_i(x') \left| f(x_i) - f(x') \right| + \sum_{\substack{1 \leqslant i \leqslant s, \\ x' \notin M_i}} c_i(x') \left| f(x_i) - f(x') \right| \leqslant \\
&\leqslant \sum_{\substack{1 \leqslant i \leqslant s, \\ x' \in M_i}} c_i(x') \varepsilon + \sum_{\substack{1 \leqslant i \leqslant s, \\ x' \notin M_i}} \frac{\varepsilon}{m_i \sum_{i=1}^{s} \gamma_i \int_{\mathbb{R}^p} \mu_{B_i}(y) dy} \left| f(x_i) - f(x') \right| \leqslant \\
&\leqslant \sum_{\substack{1 \leqslant i \leqslant s, \\ x' \in M_i}} c_i(x') \varepsilon + \sum_{\substack{1 \leqslant i \leqslant s, \\ x' \notin M_i}} \frac{\varepsilon}{m_i \sum_{i=1}^{s} \gamma_i \int_{\mathbb{R}^p} \mu_{B_i}(y) dy} m_i \leqslant \\
&\leqslant \varepsilon \sum_{\substack{1 \leqslant i \leqslant s, \\ x' \in M_i}} c_i(x') + \varepsilon \sum_{\substack{1 \leqslant i \leqslant s, \\ x' \notin M_i}} \frac{1}{\sum_{i=1}^{s} \gamma_i \int_{\mathbb{R}^p} \mu_{B_i}(y) dy} \leqslant \\
&\leqslant \varepsilon + \varepsilon \sum_{i=1}^{s} \frac{1}{\sum_{i=1}^{s} \gamma_i \int_{\mathbb{R}^p} \mu_{B_i}(y) dy} = \\
&= \varepsilon \left(1 + \sum_{i=1}^{s} \frac{1}{\sum_{i=1}^{s} \gamma_i \int_{\mathbb{R}^p} \mu_{B_i}(y) dy} \right) = \\
&= \varepsilon \theta
\end{aligned}
$$

where $\theta > 0$ is constant in $x' \in U$.
Moreover, we can use this proof to demonstrate and generalize the previous Wang's theorem without using the Stone-Weierstrass theorem: let $U \subseteq \mathbb{R}^n$ be a compact, Wang's theorem proves the density in $C(U, \mathbb{R})$ for a class of fuzzy systems that form a subclass of $AFS_{prod}(U, \mathbb{R})$. In fact, a fuzzy systems $f \in Y(U)$, in the class considered by Wang's theorem, has as knowledge base a finite set of fuzzy implications of $U \times \mathbb{R}$ such that both membership functions of antecendets and membership functions of consequents are Gaussian functions, i.e

$$K = \{R_i = A_i \implies B_i : i = 1, \ldots, s\}$$

such that $\forall i = 1, \dots, s$

$$\forall x \in U,\ \mu_{A_i}(x) = \alpha_i exp\left(-\frac{1}{2}\left|\frac{x-\overline{x}_i}{\sigma_i}\right|^2\right) =$$

$$= \alpha_i \prod_{k=1}^{n} exp\left(-\frac{1}{2}\left(\frac{x_k-\overline{x}_{i,k}}{\sigma_i}\right)^2\right)$$

$$\forall x \in \mathbb{R},\ \mu_{B_i}(x) = \beta_i exp\left(-\frac{1}{2}\left(\frac{x-\nu_i}{\gamma_i}\right)^2\right)$$

with $0 < \alpha_i, \beta_i \leqslant 1,\ \gamma_i, \sigma_i \in (0, +\infty),\ \nu_i \in \mathbb{R}, \overline{x}_i \in R^n\ \forall i = 1, \dots, s$. Then the output of fuzzification and inference is $\forall x \in U$

$$\mu_{I_K \circ F_K(x)} = \sum_{i=1}^{s} w_i \mu_{A_i}(x) \mu_{B_i}$$

where $\forall i = 1, \dots, s$

$$w_i = \left(\int_{\mathbb{R}} \mu_{B_i}(y)dy\right)^{-1}$$

Finally the output of centroid defuzzification is

$$\frac{\sum_{i=1}^{s} \nu_i \mu_{A_i}(x)}{\sum_{i=1}^{s} \mu_{A_i}(x)}$$

In other words, is the subclass of $AFS_{prod}(U, \mathbb{R})$ where membership functions are Gaussian membership functions, the fuzzification interface is the point fuzzification and in the inference $\forall A \in \mathcal{F}(U),\ w_i(A) = \left(\int_{\mathbb{R}} \mu_{B_i}(y)dy\right)^{-1}$ and $a_i(A, K) = A \circ A_i = \sup_{x \in U} \mu_A(x) \mu_{A_i}(x)$.
We can also consider multivariate Gaussian membership functions for the consequents B_i of the rules and obtain a generalization of the class $Y(U)$, namely $Y(U, \mathbb{R}^p) \subseteq AFS_{prod}(U, \mathbb{R}^p)$ with Gaussian membership functions both for antecendents and consenquets of the rules, point fuzzification, and inference such that $\forall A \in \mathcal{F}(U),\ w_i(A) = \left(\int_{\mathbb{R}^p} \mu_{B_i}(y)dy\right)^{-1}$ and $a_i(A, K) = A \circ A_i = \sup_{x \in U} \mu_A(x) \mu_{A_i}(x)$.
We can prove that if $U \subset \mathbb{R}^n$ is compact, $Y(U, \mathbb{R}^p)$ is dense in $C(U, \mathbb{R}^p)$ with respect to ∞-norm, by adapting the proof of Kosko's theorem as follows:

- Construct the finite cover of open cubes of U as in the proof of Kosko's theorem
- $\forall i = 1, \dots, s$ choose μ_{B_i} as a Gaussian function centered in $f(x_i) \in \mathbb{R}^p$ and μ_{A_i} as Gaussian function such that $\forall x \notin M_i,\ \mu_{A_i}(x) \leqslant \frac{\varepsilon}{m_i}$ where $m_i = \max_{x \in U} |f(x) - f(x_i)|$ [11]
- Observe that since $\forall i = 1, \dots, s$ μ_{A_i} is a Gaussian function, it has non zero minimum $\gamma_i > 0$ in the compact U

[11] It is always possible to consider a Gaussian function that remains below a certain bound outside a bounded set

- Conclude [12] that $\forall x \in U$

$$F(x) = \frac{\sum_{i=1}^{s} C(\mu_{B_i})\mu_{A_i}(x)}{\sum_{i=1}^{s} \mu_{A_i}(x)} = \frac{\sum_{i=1}^{s} f(x_i)\mu_{A_i}(x)}{\sum_{i=1}^{s} \mu_{A_i}(x)}$$

and then

$$|F(x) - f(x)| = \left| \frac{\sum_{i=1}^{s} \mu_{A_i}(x)\,[f(x_i) - f(x)]}{\sum_{i=1}^{s} \mu_{A_i}(x)} \right| \leqslant$$

$$\leqslant \left| \frac{\sum_{\substack{i=1 \\ x \in M_i}}^{s} \mu_{A_i}(x)\,[f(x_i) - f(x)]}{\sum_{i=1}^{s} \mu_{A_i}(x)} \right| + \left| \frac{\sum_{\substack{1 \leqslant i \leqslant s, \\ x \notin M_i}} \mu_{A_i}(x)\,[f(x_i) - f(x)]}{\sum_{i=1}^{s} \mu_{A_i}(x)} \right| \leqslant$$

$$\leqslant \varepsilon + \sum_{\substack{1 \leqslant i \leqslant s, \\ x \notin M_i}} \frac{\varepsilon}{\sum_{i=1}^{s} \gamma_i} \leqslant$$

$$\leqslant \varepsilon + \sum_{i=1}^{s} \frac{\varepsilon}{\sum_{i=1}^{s} \gamma_i} =$$

$$= \varepsilon \left(1 + \sum_{i=1}^{s} \frac{1}{\sum_{i=1}^{s} \gamma_i} \right) \leqslant \varepsilon\theta$$

with $\theta > 0$ is a constant.

4.3 Fuzzy relation based system

Besides those analyzed so far, other types of fuzzy systems have been shown to be universal approximators. In the context of real-valued functions of a real variable, an important result is due to Castro and Delgrado [CD96].

We define two sets of SISO fuzzy systems on an input universe $U \subseteq \mathbb{R}$ and output universe $\mathbb{R}$ that differ in the choice of the fuzzification algorithm.

In particular, given

- a class REL of fuzzy relations on $\mathbb{R} \times \mathbb{R}$ such that for each finite family of squares $\mathcal{I} = \{I_h = (x_h - \delta, x_h + \delta) \times (y_h - \varepsilon, y_h + \varepsilon) : (x_h, y_h) \in \mathbb{R}^2, h = 1, \ldots, n,\ \delta > 0, \varepsilon > 0\}$, for some $n \in \mathbb{N}$, there is a fuzzy relation $R \in REL$ such that $\mu_R(x, y) \neq 0 \iff \exists h \in \{1, \ldots, n\} : (x, y) \in I_h$ or, equivalently, $\{(x, y) \in \mathbb{R}^2 : \mu_R(x, y) \neq 0\} = \bigcup_{h=1}^{n} I_h$

[12] The Proposition A.1 can be generalize to multivariate Gaussian function, the centroid of μ_{B_i} is its center $f(x_i)$

- A t-norm $\wedge : [0,1] \times [0,1] \to [0,1]$
- A t-conorm $\vee : [0,1] \times [0,1] \to [0,1]$ which it makes sense to consider $\bigvee_{x \in X} x$ where $X \subseteq [0,1]$ is infinite[13].
- a defuzzification algorithm $D : \mathcal{F}(\mathbb{R}) \times \mathcal{K} \to \mathbb{R}$ verifying the property of producing a point in the support of the original fuzzy set, that is $\forall B \in \mathcal{F}(\mathbb{R}) : \mu_B \not\equiv 0, \mu_B(D(B,K)) \neq 0$

we denote by F_{point} the set of all fuzzy systems of the form:

$$f = D_K \circ I_K \circ F_K : U \to \mathbb{R}$$

where
- $K = R \in \mathcal{K} = REL$ is a fuzzy relation on $\mathbb{R} \times \mathbb{R}$
- $F_K : U \to \mathcal{F}(U)$ such that $\forall x_0 \in U, \mu_{F_K(x_0)}(x) \neq 0 \iff x = x_0$ [14]
- $I_K : \mathcal{F}(U) \to \mathcal{F}(\mathbb{R})$ is the compositional rule of inference, i.e.

$$\forall A \in \mathcal{F}(U), I_K(A) = A \circ R$$

i.e. $\forall y \in \mathbb{R}$

$$\mu_{A \circ R}(y) = \bigvee_{x \in U} \wedge(\mu_A(x), \mu_R(x,y))$$

The design parameters of a fuzzy system $f \in F_{point}$ are the Knowledge base $K = R \in REL$ and the particular fuzzification interface (i.e. the choice of $\mu_{F_K(x_0)}(x_0)$ $\forall x_0 \in U$).
We denote by F_{approx} the set of fuzzy systems having the same form as the previous ones, with the difference that $F_K : U \to \mathcal{F}(U)$ is the approximate fuzzification, that is, given $\delta > 0, \forall x_0 \in U$

$$\mu_{F_K(x_0)}(x) \neq 0 \iff x \in U : |x - x_0| < \delta$$

or, equivalently

$$\forall x \in U, \mu_{F_K(x_0)}(x) = \mu[x_0, \delta](x)$$

where $\mu[x_0,\delta] : \mathbb{R} \to [0,1]$ is a function such that $\mu[x_0,\delta](x) \neq 0 \iff |x - x_0| < \delta$.
The design parameters of a fuzzy system $f \in F_{approx}$ are the Knowledge base $K = R \in REL$, $\delta > 0$ and $\mu[x,\delta]$ $\forall x \in U$ (i.e. the particular fuzzification interface).

If $U \subset \mathbb{R}$ is a compact, then F_{point} and F_{approx} are dense in $C(U)$ with respect to the ∞-norm.
We first prove the following
Lemma 4.6. *Let $U \subset \mathbb{R}$ be a compact, then*

$$\forall g \in C(U), \forall \varepsilon > 0, \exists f \in F_{approx} : \forall x_0 \in U, \forall y \in \mathbb{R}, \mu_{B_{x_0}}(y)|g(x_0) - y| \leqslant \varepsilon \mu_{B_{x_0}}(y)$$

[13] An example of t-conorm with this property is the standard union t-conorm, i.e. maximum t-conorm: if $X \subseteq [0,1]$, even infinite, $0 \leqslant \bigvee_{x \in X} x = \sup_{x \in X} x \leqslant 1$

[14] Similar to point fuzzification but it is also possibile that $\mu_{F_K(x_0)}(x_0) \neq 1$

where B_{x_0} is the fuzzy output of f (i.e. $I_K \circ F_K(x_0)$) for the input x_0. Moreover, the membership function $\mu_{B_{x_0}}$ is not identically zero.
Similarly,

$$\forall g \in C(U), \forall \varepsilon > 0, \exists f \in F_{point} : \forall x_0 \in U, \forall y \in \mathbb{R}, \mu_{B_{x_0}}(y)|g(x_0) - y| \leqslant \varepsilon \mu_{B_{x_0}}(y)$$

where B_{x_0} is the fuzzy output of f (i.e. $I_K \circ F_K(x_0)$) for the input x_0.Moreover, the membership function $\mu_{B_{x_0}}$ is not identically zero.

Proof. Let $g \in C(U)$ and $\varepsilon > 0$: g is continuous on a compact U, then g is uniformly continuous on U, i.e. $\exists \delta > 0 : \forall x_1, x_2 \in U : |x_1 - x_2| < \delta, |g(x_1) - g(x_2)| < \varepsilon$.
For each $x \in U$ let consider $(x - \delta, x + \delta)$, it is obvious that $U \subseteq \bigcup_{x \in U}(x - \delta, x + \delta)$.
Since U is a compact and $\{(x - \delta, x + \delta)\}_{x \in U}$ is an open cover of U, it exists a finite family of points $x_1, \ldots, x_s \in U$ such that $U \subseteq \bigcup_{l=1}^{s}(x_l - \delta, x_l + \delta)$. From the choice of δ, it follows that $\forall l = 1, \ldots, s, \forall x \in U \cap (x_l - \delta, x_l + \delta), |g(x) - g(x_l)| < \varepsilon$.
Let consider the finite family of squares

$$\mathcal{I} = \{I_l = (x_l - \delta, x_l + \delta) \times (y_l - \varepsilon, y_l + \varepsilon) : l = 1, \ldots, s\}$$

where $\forall l = 1, \ldots s, y_l = g(x_l)$.
From the hypothesis on REL, it follows that

$$\exists R \in REL : \mu_R(x, y) \neq 0 \iff \exists l = 1, \ldots, s : (x, y) \in I_l$$

or, equivalently, $\{(x, y) \in \mathbb{R}^2 : \mu_R(x, y) \neq 0\} = \bigcup_{l=1}^{s} I_l$.
Let consider a $f \in F_{approx}$ corresponding to the choice of $R \in REL$ and such that $\forall x_0 \in U, F_K(x_0) = A_{x_0} : \forall x \in U$

$$\mu_{A_{x_0}}(x) = \begin{cases} 1 & |x - x_0| < \delta \\ 0 & |x - x_0| \geqslant \delta \end{cases}$$

Then $\forall x_0 \in U$ the fuzzy output of f for the input x_0 is $B_{x_0} = A_{x_0} \circ R$, i.e

$$\forall y \in \mathbb{R}, \mu_{B_{x_0}}(y) = \bigvee_{x \in U} \wedge(\mu_{A_{x_0}}(x), \mu_R(x, y))$$

$\mu_{B_{x_0}}$ is not identically zero because $\forall y \in \mathbb{R}$ [15]

$$\begin{aligned} \mu_{B_{x_0}}(y) &\geqslant \wedge(\mu_{A_{x_0}}(x_0), \mu_R(x_0, y)) = \\ &= \wedge(1, \mu_R(x_0, y)) = \\ &= \mu_R(x_0, y) \end{aligned}$$

Let $l = 1, \ldots, s : x_0 \in (x_l - \delta, x_l + \delta)$, then $\forall y \in (y_l - \varepsilon, y_l + \varepsilon), (x_0, y) \in I_l$ and $\mu_{B_{x_0}}(y) \geqslant \mu_R(x_0, y) > 0$
Let $y \in \mathbb{R}$, we distinguish two cases:

- $\mu_{B_{x_0}}(y) > 0$

[15] $\forall x, y \in [0, 1], \vee(x, y) \geqslant x, y$

In this case, $0 < \bigvee_{x\in U} \wedge(\mu_{A_{x_0}}(x), \mu_R(x,y))$ then [16] $\exists x' \in U : \wedge(\mu_{A_{x_0}}(x'), \mu_R(x',y)) > 0$. From the properties of a t-norm [17], it follows that $\mu_{A_{x_0}}(x') > 0$ and $\mu_R(x',y) > 0$. In particular,

- $\mu_{A_{x_0}}(x') > 0 \implies |x' - x_0| < \delta \implies |g(x_0) - g(x')| < \varepsilon$
- $\mu_R(x',y) > 0 \implies \exists l = 1,\dots,s : (x',y) \in I_l$, i.e. $|x' - x_l| < \delta$ and $|y_l - y| < \varepsilon \implies |g(x') - g(x_l)| < \varepsilon$ and $|g(x_l) - y| < \varepsilon$.

Then,

$$|g(x_0) - y| \leqslant |g(x_0) - g(x')| + |g(x') - g(x_l)| + |g(x_l) - y| \leqslant 3\varepsilon$$

Multiplying both sides by $\mu_{B_{x_0}}(y)$,

$$\mu_{B_{x_0}}(y)|g(x_0) - y| \leqslant 3\varepsilon\mu_{B_{x_0}}(y).$$

- $\mu_{B_{x_0}}(y) = 0$

In this case, obviously $\mu_{B_{x_0}}(y)|g(x_0) - y| \leqslant 3\varepsilon\mu_{B_{x_0}}(y)$ since both sides are zero.
Similarly, let consider a $h \in F_{point}$ corresponding to the choice of $R \in REL$ and of point fuzzification as fuzzification interface. Then $\forall x_0 \in U$ the fuzzy output of h for the input x_0 is $B_{x_0} = A_{x_0} \circ R$, i.e.

$$\mu_{B_{x_0}}(y) = \bigvee_{x\in U} \wedge(\mu_{A_{x_0}}(x), \mu_R(x,y))$$

where A_{x_0} is the fuzzy singleton associated to x_0.
$\mu_{B_{x_0}}$ is not identically zero because $\forall y \in \mathbb{R}$

$$\begin{aligned}\mu_{B_{x_0}}(y) &= \bigvee_{x\in U} \wedge(\mu_{A_{x_0}}(x), \mu_R(x,y)) = \\ &= \wedge(\mu_{A_{x_0}}(x_0), \mu_R(x_0,y)) = \\ &= \wedge(1, \mu_R(x_0,y)) = \\ &= \mu_R(x_0,y)\end{aligned}$$

Let $l = 1,\dots,s : x_0 \in (x_l - \delta, x_l + \delta)$, then $\forall y \in (y_l - \varepsilon, y_l + \varepsilon), (x_0,y) \in I_l$ and $\mu_{B_{x_0}}(y) = \mu_R(x_0,y) > 0$
Let $y \in \mathbb{R}$, we distinguish two cases:

- $\mu_{B_{x_0}}(y) > 0$

In this case,

$$\begin{aligned}0 < \bigvee_{x\in U} \wedge(\mu_{A_{x_0}}(x), \mu_R(x,y)) &\implies \exists x' \in U : \wedge(\mu_{A_{x_0}}(x'), \mu_R(x',y)) > 0 \implies \\ &\implies \mu_{A_{x_0}}(x') > 0, \mu_R(x',y) > 0\end{aligned}$$

.

- $\mu_{A_{x_0}}(x') > 0 \implies x' = x_0$
- $\mu_R(x',y)) > 0 \implies \exists l = 1,\dots,s : (x',y) \in I_l$ i.e. $|x' - x_l| < \delta$ and $|y_l - y| < \varepsilon \implies |g(x_0) - g(x_l)| < \varepsilon$ and $|g(x_l) - y| < \varepsilon$.

[16] $\vee$ is a t-conorm: $\forall x \in U, \wedge(\mu_{A_{x_0}}(x'), \mu_R(x',y)) = 0 \implies \bigvee_{x\in U} \wedge(\mu_{A_{x_0}}(x), \mu_R(x,y)) = 0$
[17] $x = 0$ or $y = 0 \implies \wedge(x,y) = 0$

Then,

$$|g(x_0) - y| \leqslant |g(x_0) - g(x_l)| + |g(x_l) - y| \leqslant 2\varepsilon$$

Multiplying both sides by $\mu_{B_{x_0}}(y)$,

$$\mu_{B_{x_0}}(y)|g(x_0) - y| \leqslant 2\varepsilon\mu_{B_{x_0}}(y).$$

- $\mu_{B_{x_0}}(y) = 0$

In this case, obviously $\mu_{B_{x_0}}(y)|g(x_0) - y| \leqslant 2\varepsilon\mu_{B_{x_0}}(y)$ □

Now we can conclude the following

Theorem 4.7 (Castro-Delgado)**.** *Let $U \subset \mathbb{R}$ be a compact, F_{approx} and F_{point} are dense in $C(U)$ with respect to ∞-norm.*

Proof. We want to prove that

$$\forall g \in C(U), \forall\varepsilon > 0, \exists f \in F_{approx}, h \in F_{point} : \sup_U |f - g|, \sup_U |h - g| \leqslant \varepsilon$$

From the previous lemma, we know that

$$\forall g \in C(U), \forall\varepsilon > 0, \exists f \in F_{approx} : \forall x_0 \in U, \forall y \in \mathbb{R}, \mu_{B_{x_0}}(y)|g(x_0) - y| \leqslant \varepsilon\mu_{B_{x_0}}(y)$$

where B_{x_0} is the fuzzy output of f for the input x_0 and $\mu_{B_{x_0}} \not\equiv 0$. Then $\forall x_0 \in U$, let consider $y = f(x_0) = D_K(B_{x_0})$. From the hypothesis on the defuzzification, it follows that $\mu_{B_{x_0}}(f(x_0)) > 0$. From

$$\mu_{B_{x_0}}(f(x_0))|g(x_0) - f(x_0)| \leqslant \varepsilon\mu_{B_{x_0}}(f(x_0))$$

dividing by $\mu_{B_{x_0}}(f(x_0))$, we obtain

$$|g(x_0) - f(x_0)| \leqslant \varepsilon$$

and, then, $\sup_U |f - g| \leqslant \varepsilon$.
Similarly, we know that

$$\forall g \in C(U), \forall\varepsilon > 0, \exists h \in F_{point} : \forall x_0 \in U, \forall y \in \mathbb{R}, \mu_{B_{x_0}}(y)|g(x_0) - y| \leqslant \varepsilon\mu_{B_{x_0}}(y)$$

where B_{x_0} is the fuzzy output of h for the input x_0 and $\mu_{B_{x_0}} \not\equiv 0$. Then $\forall x_0 \in U$, let consider $y = h(x_0) = D_K(B_{x_0})$. Again from the hypotesis on the defuzzification,it follows that $\mu_{B_{x_0}}(h(x_0)) > 0$. From

$$\mu_{B_{x_0}}(h(x_0))|g(x_0) - h(x_0)| \leqslant \varepsilon\mu_{B_{x_0}}(h(x_0))$$

dividing by $\mu_{B_{x_0}}(h(x_0))$, we obtain

$$|g(x_0) - h(x_0)| \leqslant \varepsilon$$

and, then, $\sup_U |h - g| \leqslant \varepsilon$.

□

5 Neuro-Fuzzy systems

5.1 Adaptive networks

An *adaptive network* is a parametrized function $f : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^m$ that can be represented as a directed graph $G = (V, E)$ such that for each node $i \in V$ we have a function associated to that node $f_i : \mathbb{R}^{p_i} \times \mathbb{R}^{r_i} \to \mathbb{R}^{q_i}$, where p_i is the dimension of the input of f_i, q_i is the dimension of the output of f_i and r_i is the number of parameters taken by the function f_i. The graph is connected and there are n nodes that have no incoming edges, those are called *input nodes* and are assigned as an argument x of f along with its parameters P to calculate $f(x, P)$. Similarly there are m nodes that have no outcoming nodes, those are called *output nodes* and contain the result $f(x, P)$. Each directed edge from node i to node j denotes that the output of the node i is passed as argument to the node j. The nodes that don't have parameters are represented with a circle while the ones that have them are said "adaptive nodes" and are represented with squares. An adaptive network can be changed by varying the parameters of the functions of its nodes. An example of adaptive network is given in Figure 2, it's also possible to specify the parameters of an adaptive node as shown in Figure 3, in this case the functions that previously were adaptive are now fixed and the parameters nodes are seen as adaptive nodes that simply return the parameters themselves. So in general a "parameter node" is a node with the identity function $\mathrm{id}_{\mathbb{R}^r}$, where s is the number of parameters.

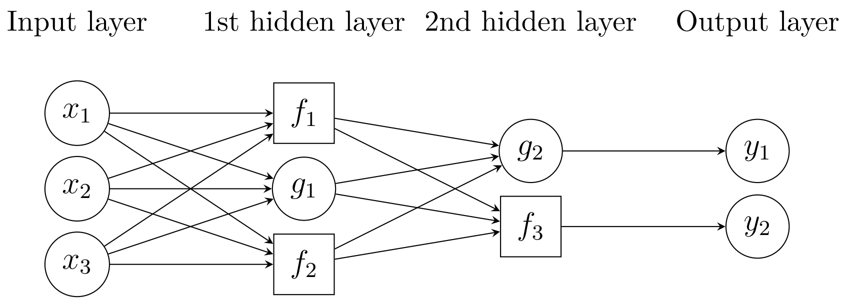


Figure 2: A feedforward adaptive network

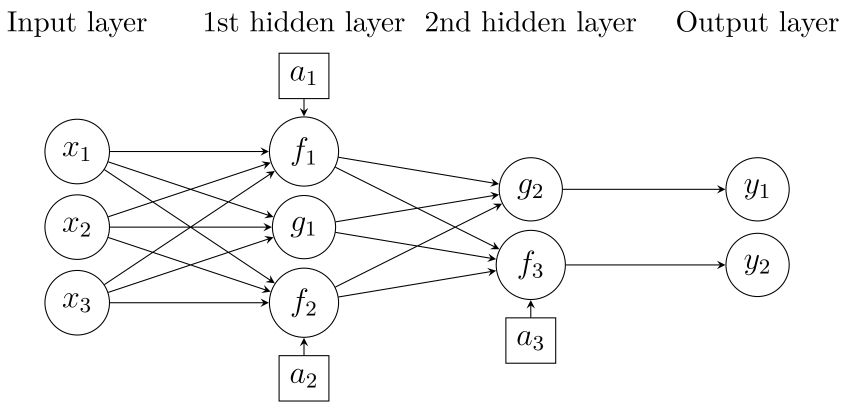


Figure 3: A feedforward adaptive network with explicit parameters

We classify the adaptive networks in *feedforward adaptive networks* if their graph is acyclic and *recurrent adaptive networks* otherwise. Figure 2 is an example of feedforward adaptive network and Figure 4 is an example of recurrent adaptive network.

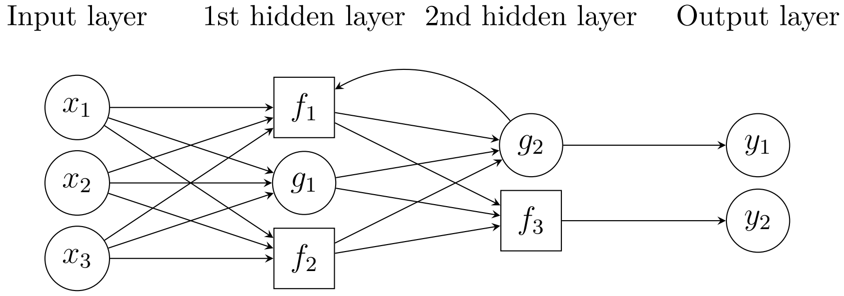


Figure 4: A recurrent adaptive network

In Figure 2, we sorted the adaptive network graph into "layers" such that the outputs of a layer are the inputs of the next layer,[18] this is the most common graphical representation called *layered representation*, but also others exist, like the *topological representation*.
An adaptive network $f : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^m$ can assume the form of different functions $f_P : \mathbb{R}^n \to \mathbb{R}^m$, where $P \in \mathbb{R}^r$, depending on the vector of parameters P. Given a finite set of inputs $X \subset \mathbb{R}^n$ and a set of desired outputs $Y \subset \mathbb{R}^m$ with the same number of elements of X, the process of tuning an adaptive network's parameters to get a set of parameters P^* such that the error measure $e(f_{P*}(X), Y)$ is minimized, or below a certain threshold, is called *learning* and it is carried out with some *learning algorithms* or *learning rules*. We will give a description of a widely known learning algorithm for feedforward adaptive networks: the *steepest descent method*.

5.2 Steepest Descent Method

Let $f : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^m$ be an adaptive network such that its node functions are differentiable and $P = \Big(a_{1,1}, \dots, a_{1,N(1)}, \dots, a_{L,1}, \dots, a_{L,N(L)}\Big) \in \mathbb{R}^r$ is the vector of parameters. We represent f in layered form with $L+2$ layers, where the 0-th layer is the input layer and the $(L+1)$-th layer is the output layer, each layer $l \in \{0, \dots, L+1\}$ has $N(l)$ nodes, we require that $N(L) = N(L+1)$. Let $X_1 = (x_{0,1}, \dots, x_{0,N(0)}) \in \mathbb{R}^{N(0)}$ be the vector of inputs and $Y_1 = (y_1, \dots, y_{N(L)}) \in \mathbb{R}^{N(L)}$ the vector of desired outputs, let $x_{l,i} \in \mathbb{R}$, $f_{l,i}$, $a_{l,i}$ be respectively the output, the function and the vector of parameters of the node $i \in \{1, \dots, N(l)\}$ in the layer $l \in \{0, \dots, L\}$, notice that the output layer $L+1$ doesn't have an output. Then the expression of any output is

$$x_{l+1,i} = f_{l+1,i}(x_l, a_{l+1,i}) \quad \forall\, l \in \{0, \dots, L-1\}, i \in \{1, \dots, N(l)\} \ ,$$

[18]Note that this is not a restriction on the resulting function because you can always add a "chain of identity functions" that pass the output of a non-consecutive layer to desired layer (even "much further" in the layer order).

where $x_l = (x_{l,1}, \dots, x_{l,N(l)})$. We define the error of the function f_P on the input X_1 as

$$E_1 = e(x_L) = \sum_{i=1}^{N(L)} (y_i - x_{L,i})^2$$

this definition is not unique and it might change depending on the context or specific needs. We now define the error signal respect to the node $i \in \{1, \dots, N(l)\}$ in the layer $l \in \{0, \dots, L\}$ as

$$\varepsilon_{l,i} = \frac{\partial e}{\partial x_{l,i}}(x_L)\,,$$

then

$$\varepsilon_{L,i} = \frac{\partial e}{\partial x_{L,i}}(x_L) = 2(x_{L,i} - y_i)\,.$$

Now notice that

$$e(x_L) = \sum_{i=1}^{N(L)} \left(y_i - f_{L,i}(x_{L-1}, a_L)\right)^2\,,$$

so the error depends not only on the last layer but also on the preceding, this reasoning can be extended to all the layers up to the input layer. So

$$\varepsilon_{L-1,i} = \frac{\partial e}{\partial x_{L-1,i}}(x_L) = \nabla e(x_L) \cdot \frac{\partial f_L}{\partial x_{L-1,i}}(x_{L-1}) = \varepsilon_L \cdot \frac{\partial f_L}{\partial x_{L-1,i}}(x_{L-1})\,,$$

in the same way we have $\forall\, l \in \{0, \dots, L-1\}, i \in \{1, \dots, N(l)\}$

$$\varepsilon_{l,i} = \frac{\partial e}{\partial x_{l,i}}(x_L) = \frac{\partial e}{\partial x_{l+1}}(x_L) \cdot \frac{\partial f_{l+1}}{\partial x_{l,i}}(x_l) = \varepsilon_{l+1} \cdot \frac{\partial f_{l+1}}{\partial x_{l,i}}(x_l)\,,$$

Now we can calculate the derivative of the error respect to a parameter $a_{l+1,i}$, of course the only part of the network that depends on that parameter is the node i in the layer $l+1$, so $x_{l+1,i}$ will depend on it as well. We have, $\forall\, l \in \{0, \dots, L-1\}, i \in \{1, \dots, N(l)\}$

$$\frac{\partial e}{\partial a_{l+1,i}}(x_L) = \frac{\partial e}{\partial x_{l+1}}(x_L) \cdot \frac{\partial f_{l+1}}{\partial a_{l+1,i}}(x_l) = \frac{\partial e}{\partial x_{l+1,i}}(x_L)\frac{\partial f_{l+1,i}}{\partial a_{l+1,i}}(x_l) = \varepsilon_{l+1,i}\frac{\partial f_{l+1,i}}{\partial a_{l+1,i}}(x_l)\,,$$

we denote it also as

$$\frac{\partial E_1}{\partial a_{l+1,i}} = \frac{\partial e}{\partial a_{l+1,i}}(x_L)\,.$$

So far we considered just one input vector X_1 and one desired output vector Y_1, but in order to train an adaptive network we need a dataset with multiple input-output couples. So if we have $N \in \mathbb{N}$ input-output couples (X_i, Y_i), we define the total error as

$$E = \sum_{s=1}^{N} E_s\,,$$

where E_s is the error on the s-th input-output couple as defined for E_1. Therefore the derivative of the total error respect to the parameter $a_{l,i}$ is

$$\frac{\partial E}{\partial a_{l,i}} = \sum_{s=1}^{N} \frac{\partial E_s}{\partial a_{l,i}}\,.$$

Finally we define the variation of the parameter $a_{l,i}$ as

$$\Delta a_{l,i} = -\eta \frac{\partial E}{\partial a_{l,i}} ,$$

where $\eta \in \mathbb{R} : \eta > 0$ is chosen, it's called *learning rate* and it's said to be an *hyperparameter of the learning algorithm.*
We just gave a description of a step of the *Steepest Descent Method*, in fact at the step $k \in \mathbb{N}$ we find the parameter variation $\Delta a_{l,i}^{(k)}$ and update

$$a_{l,i}^{(k+1)} = a_{l,i}^{(k)} + \Delta a_{l,i}^{(k)} .$$

So if $P^{(k)} = \left(a_{1,1}^{(k)}, \dots, a_{1,N(1)}^{(k)}, \dots, a_{L,1}^{(k)}, \dots, a_{L,N(L)}^{(k)}\right)$ is the vector of parameters at the k-th iteration, we stop the algorithm when k goes over a max number K of iterations or the total error the k-th iteration $E^{(k)}$ is below a threshold δ. δ and K are hyperparameters of this learning algorithm.
Adaptive networks include some more specific function classes, for example neural networks, we will focus on a specific class of adaptive networks that are connected with fuzzy systems: ANFIS.

5.3 ANFIS

ANFIS, that stands for Adaptive Network-based Fuzzy Inference System or Adaptive Neuro-Fuzzy Inference System, is an adaptive network that is a fuzzy system according to Definition 3.5.
We define a fuzzy system that is usually represented as an ANFIS network, we introduce the necessary notation first.
Notation 5.1. Let $U_1, \dots, U_n, V$ be universes of discourse, and let $A_1, \dots, A_n$ be fuzzy sets respectively on $U_1, \dots, U_n$ and let $f : U_1 \times \cdots \times U_n \to V$. The notation

$$\text{IF } x_1 \text{ is } A_1 \text{ and } \cdots \text{ and } x_n \text{ is } A_n \text{ THEN } z = f(x_1, \dots, x_n) , \tag{6}$$

stands for an implication of the form

$$\text{IF } x_1 \text{ is } A_1 \text{ and } \cdots \text{ and } x_n \text{ is } A_n \text{ THEN } z \text{ is } B ,$$

where B is the fuzzy set given by the point fuzzification of $f(x_1, \dots, x_n)$.
Definition 5.2 (Takagi-Sugeno-Kang fuzzy systems)**.** Let $U_1 \subseteq \mathbb{R}, \dots, U_n \subseteq \mathbb{R}$ universes of discourse, let $U = U_1 \times \cdots \times U_n$, let $A_{11}, ..., A_{1r}, \dots, A_{n1}, ..., A_{nr}$ be fuzzy sets respectively on $U_1, \dots, U_n$, let $a_j \in \mathbb{R}^n$ and $b_j \in \mathbb{R}$ for all $j \in \{1, \dots, r\}$, let $K = \{A_{ij} : 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant r\} \cup \{a_j, b_j : 1 \leqslant j \leqslant r\}$ be the knowledge base. Let the fuzzification interface $F_U : U \to \mathcal{F}(U)$ be the point fuzzification on U, let $F : \mathbb{R} \to \mathcal{F}(\mathbb{R})$ be the point fuzzification in $\mathbb{R}$. Let $R_{K,j} : F_U(U) \to \mathcal{F}(U \times \mathbb{R})$ be the j-th fuzzy rule generator, with $j \in \{1, \dots, r\}$, such that

$$R_{K,j}(A) = A_{1j} \times \cdots \times A_{nj} \implies F(F_U^{-1}(A) \cdot a_j + b_j) ,$$

where the implication is given by the product inference rule. Let $I_K : \mathcal{F}(U) \to \mathcal{F}(\mathbb{R})^r$ the inference machine, such that

$$I_K(A) = \big(A \circ R_{K,1}(A), \dots, A \circ R_{K,r}(A)\big) .$$

Let $D : \mathcal{F}(\mathbb{R}) \to \mathbb{R}$ be the max defuzzification, let $D_K : \mathcal{F}(\mathbb{R})^r \to \mathbb{R}$ be the defuzzification interface such that

$$D_K(C'_1, \ldots, C'_r) = \frac{\sum_{j=1}^r \mu_{C'_j}(D(C'_j))D(C'_j)}{\sum_{j=1}^r \mu_{C'_j}(D(C'_j))} \,.$$

Let $f = D_K \circ I_K \circ F_K$ be the *Takagi-Sugeno-Kang fuzzy system* or *TSK fuzzy system.* The TSK fuzzy system is described as a fuzzy system as per Definition 3.5, but in literature it's written in a simpler form. We prove it in the following.

Proposition 5.1. *Let $f : U \to \mathbb{R}$ be a TSK fuzzy system with the notations in Definition 5.2. Then, for each $x \in U$*

$$f(x) = \frac{\sum_{j=1}^r w_j z_j}{\sum_{j=1}^r w_j} \,,$$

where for each $j \in \{1, \ldots, r\}$

$$w_j = \bigwedge_{1 \leqslant i \leqslant n} \mu_{A_{ij}}(x_i)$$

and $z_j = x \cdot a_j + b_j$.

Remark 5.1. The t-norm $\wedge$, needs to be specified, usually it is the minimum operator.

Proof. Let $x \in U$, let $A = F_U(x)$ be the fuzzy set resulting of the point fuzzification of x. Let $j \in \{1, \ldots, r\}$, let

$$C_j = F(F_U^{-1}(A) \cdot a_j + b_j) = F(x \cdot a_j + b_j) \,,$$

that is the point fuzzification of $x \cdot a_j + b_j$, this also implies that a TSK uses rules of the form Equation 6. Whatever are the chosen norm $\wedge$ and conorm $\vee$, let

$$\mu_{R_j}(u, z) = \mu_{A_{1j} \times \cdots \times A_{nj}}(u)\, \mu_{C_j}(z) = \bigwedge_{1 \leqslant i \leqslant n} \mu_{A_{ij}}(u_i) \cdot \mu_{C_j}(z) \,,$$

in particular,

$$\mu_{R_j}(x, z) = \bigwedge_{1 \leqslant i \leqslant n} \mu_{A_{ij}}(x_i) \cdot \mu_{C_j}(z) = w_j \mu_{C_j}(z) \,,$$

then

$$\mu_{A \circ R_{K,j}(A)}(z) = \bigvee_{u \in U} \wedge(\mu_A(u), \mu_{R_j}(u, z)) = \wedge(\mu_A(x), \mu_{R_j}(x, z)) = \mu_{R_j}(x, z) = w_j\, \mu_{C_j}(z) \,.$$

So

$$\mu_{C'_j}(z) = \mu_{A \circ R_{K,j}(A)}(z) = \begin{cases} w_j & \text{if } z = z_j \\ 0 & \text{if } z \neq z_j \end{cases}$$

and if $w_j \neq 0$ the max defuzzification $D(C'_j) = z_j$ is well defined, otherwise we don't calculate it because in the final sum it doesn't appear. Finally

$$f(x) = \frac{\sum_{j=1}^r \mu_{C'_j}(D(C'_j))D(C'_j)}{\sum_{j=1}^r \mu_{C'_j}(D(C'_j))} = \frac{\sum_{j=1}^r \mu_{C'_j}(z_j) z_j}{\sum_{j=1}^r \mu_{C'_j}(z_j)} = \frac{\sum_{j=1}^r w_j z_j}{\sum_{j=1}^r w_j} \,.$$

The only remaining problem is if all the $w_j = 0$, this is generally not specified in the TSK fuzzy system, but it can be solved assuming, for example, that, for each $1 \leqslant i \leqslant n$, $A_{i1}, \ldots, A_{ir}$ have supports that cover the entire U_i. □

We can represent a TSK fuzzy system as an ANFIS network as in Figure 5. Where

$$N_j(w_1, \dots, w_r) = \frac{w_j}{\sum_{j=1}^{r} w_j} = \overline{w}_j \quad \forall\, j \in \{1, \dots, r\}$$

and

$$\overline{z}_j(x_1, \dots, x_n, \overline{w}_j) = \overline{w}_j(x \cdot a_j + b_j)\,,$$

that has a_j and b_j as parameters.

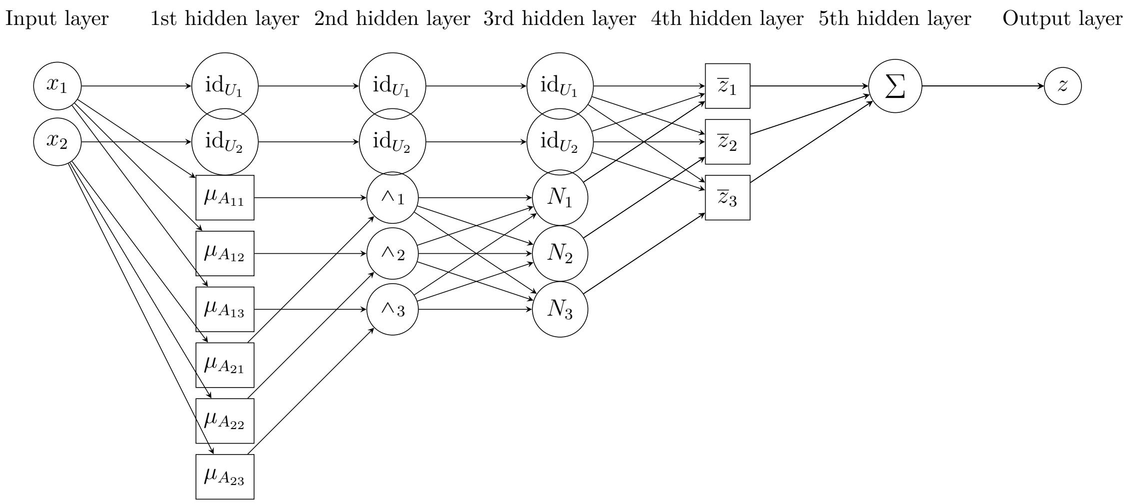


Figure 5: A TSK fuzzy system with 2 inputs and 3 rules represented as an ANFIS network

It's possible to train an ANFIS network as we did in subsection 5.2, but there are other learning algorithms that are more performant.

6 Conclusion

Our work aims to give a rigorous introduction to the fuzzy sets and fuzzy logic theories and to provide an overview of ANFIS networks, we proposed a definition of fuzzy system in Definition 3.5 and expanded known results as in Theorem 4.3, Theorem 4.5 and Theorem 4.7.

A Appendix 1

A.1 Gaussian functions

Proposition A.1. *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a Gaussian function, i.e*

$$\phi(z) = \alpha exp\left(-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right)$$

with $\alpha, \sigma \in \mathbb{R} - \{0\}$ and $\mu \in \mathbb{R}$, then

$$\mu = \frac{\displaystyle\int_{\mathbb{R}} z\,\phi(z)\,dz}{\displaystyle\int_{\mathbb{R}} \phi(z)\,dz}$$

Proof. Recall that $\forall \alpha, \mu \in \mathbb{R}, \sigma \in \mathbb{R} - \{0\}$,

$$\int_{\mathbb{R}} \alpha exp\left(-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right)\,dz = \alpha\sigma\sqrt{2\pi}.$$

It follows that

$$\begin{aligned}
&\int_{\mathbb{R}} z\,\phi(z)\,dz = -\sigma^2 \int_{\mathbb{R}} -\frac{z-\mu+\mu}{\sigma^2}\,\alpha exp\left(-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right)\,dz = \\
&= -\sigma^2 \int_{\mathbb{R}} -\frac{z-\mu}{\sigma^2}\,\alpha exp\left(-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right)\,dz + \sigma^2 \int_{\mathbb{R}} \frac{\mu}{\sigma^2}\,\alpha exp\left(-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right)\,dz = \\
&= -\sigma^2 \int_{\mathbb{R}} \frac{d}{dz}\,\alpha exp\left(-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right)\,dz + \sigma^2\frac{\mu}{\sigma^2} \int_{\mathbb{R}} \alpha exp\left(-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right)\,dz = \\
&= -\sigma^2 \left[\alpha exp\left(-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right)\right]_{-\infty}^{+\infty} + \sigma^2\frac{\mu}{\sigma^2}\alpha\sigma\sqrt{2\pi} = \mu\alpha\sigma\sqrt{2\pi}
\end{aligned}$$

and if $\alpha, \sigma \neq 0$

$$\frac{\displaystyle\int_{\mathbb{R}} z\,\phi(z)\,dz}{\displaystyle\int_{\mathbb{R}} \phi(z)\,dz} = \frac{\mu\alpha\sigma\sqrt{2\pi}}{\alpha\sigma\sqrt{2\pi}} = \mu$$

□

Proposition A.2. *Let ϕ_1, ϕ_2 be two Gaussian functions, then $\phi_1\phi_2$ is also a Gaussian function.*

Proof. Let

$$\phi_1(x) = \alpha_1 exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right),$$

$$\phi_2(x) = \alpha_2 exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right),$$

with $\alpha_1, \alpha_2, \sigma_1, \sigma_2 \in \mathbb{R} - \{0\}$ and $\mu_1, \mu_2 \in \mathbb{R}$, then

$$\begin{aligned}
\phi_1\phi_2(x) &= \alpha_1\alpha_2 exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right) exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right) = \\
&= \alpha_1\alpha_2 exp\left(-\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x-\mu_2}{\sigma_2}\right)^2\right]\right) = \\
&= \alpha_1\alpha_2 exp\left(-\frac{1}{2}\left[\left(\frac{x^2-2x\mu_1+\mu_1^2}{\sigma_1^2}\right) + \left(\frac{x^2-2x\mu_2+\mu_2^2}{\sigma_2^2}\right)\right]\right) = \\
&= \alpha_1\alpha_2 exp\left(-\frac{1}{2}\left[x^2\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right) - 2x\left(\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}\right) + \left(\frac{\mu_1^2}{\sigma_1^2}+\frac{\mu_2^2}{\sigma_2^2}\right)\right]\right),
\end{aligned}$$

by setting $\frac{1}{\sigma^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$

$$\begin{aligned}
\phi_1\phi_2(x) &= \alpha_1\alpha_2 exp\left(-\frac{1}{2}\frac{1}{\sigma^2}\left[x^2 - 2x\sigma^2\left(\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}\right) + \sigma^2\left(\frac{\mu_1^2}{\sigma_1^2}+\frac{\mu_2^2}{\sigma_2^2}\right)\right]\right) = \\
&= \alpha_1\alpha_2 exp\left(-\frac{1}{2}\frac{1}{\sigma^2}\left[x^2 - 2x\sigma^2\left(\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}\right) + \left(\sigma^2\left(\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}\right)\right)^2 - \right.\right. \\
&\left.\left. - \left(\sigma^2\left(\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}\right)\right)^2 + \sigma^2\left(\frac{\mu_1^2}{\sigma_1^2}+\frac{\mu_2^2}{\sigma_2^2}\right)\right]\right) = \\
&= \alpha_1\alpha_2 exp\left(-\frac{1}{2}\frac{1}{\sigma^2}\left[\left(x - \sigma^2\left(\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}\right)\right)^2 - \left(\sigma^2\left(\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}\right)\right)^2 + \sigma^2\left(\frac{\mu_1^2}{\sigma_1^2}+\frac{\mu_2^2}{\sigma_2^2}\right)\right]\right) = \\
&= \alpha_1\alpha_2 exp\left(-\frac{1}{2}\left(\frac{x-\sigma^2\left(\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}\right)}{\sigma}\right)^2\right) \\
&\quad exp\left(-\frac{1}{2}\frac{1}{\sigma^2}\left[-\left(\sigma^2\left(\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}\right)\right)^2 + \sigma^2\left(\frac{\mu_1^2}{\sigma_1^2}+\frac{\mu_2^2}{\sigma_2^2}\right)\right]\right).
\end{aligned}$$

If $\beta = exp\left(-\frac{1}{2}\frac{1}{\sigma^2}\left[-\left(\sigma^2\left(\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}\right)\right)^2 + \sigma^2\left(\frac{\mu_1^2}{\sigma_1^2}+\frac{\mu_2^2}{\sigma_2^2}\right)\right)\right]$ and $\mu = \sigma^2\left(\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}\right)$, we have

$$\phi_1\phi_2(x) = \beta\alpha_1\alpha_2 exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right).$$

□

A.2 Fuzzy Logic

Definition A.1 (Fuzzy system)**.** Let $U \subseteq \mathbb{R}^n$ be the *input universe* and $V \subseteq \mathbb{R}^m$ the *output universe*. Let $r_1, r_2, r_3, r_4 \in \mathbb{N}_0$, let

$$\mathcal{K} = \mathbb{R}^{r_1} \times \mathcal{F}(U)^{r_2} \times \mathcal{F}(V)^{r_3} \times \mathcal{F}(U \times V)^{r_4},$$

and let

$$\mathcal{K}_F = \mathbb{R}^{r_1} \times \mathcal{F}(U)^{r_2}, \quad \mathcal{K}_I = \mathbb{R}^{r_1} \times \mathcal{F}(U \times V)^{r_4}, \quad \mathcal{K}_D = \mathbb{R}^{r_1} \times \mathcal{F}(V)^{r_3}$$

Let $F : U \times \mathcal{K}_F \to \mathcal{F}(U)$ be the *fuzzification algorithm*, $I : \mathcal{F}(U) \times \mathcal{K}_I \to \mathcal{F}(V)^q$ be the *fuzzy inference algorithm* and $D : \mathcal{F}(V)^q \times \mathcal{K}_D \to V$ be the *defuzzification algorithm.* Moreover

$$I(A, K) = (A \circ R_1(A, K), \ldots, A \circ R_q(A, K)) \quad \forall\, A \in \mathcal{F}(U)\ \forall\, K \in \mathcal{K}_I\,,$$

where $R : \mathcal{F}(U) \times \mathcal{K}_I \to \mathcal{F}(U \times V)^q$ is the *fuzzy rules generation algorithm*, for each $i \in \{1, \ldots, q\}$ $R_i : \mathcal{F}(U) \times \mathcal{K}_I \to \mathcal{F}(U \times V)$ is the *i-th fuzzy rule generation algorithm* and $\circ$ is a synthetic operation. Let $R_K = R(\cdot, K)$ be the *fuzzy rules generator* and let $R_{K,i} = R_i(\cdot, K)$ be the *i-th fuzzy rule generator*.
Let $K \in \mathcal{K}$ be the *Knowledge Base*, let K_F, K_I, K_D the components of K such that they belong respectively to $\mathcal{K}_F, \mathcal{K}_I, \mathcal{K}_D$. Let $F_K = F(\cdot, K_F)$ be the *fuzzification interface*, or *fuzzification*, let $I_K = I(\cdot, K_I)$ be the *fuzzy inference machine*, or *fuzzy inference*, and let $D_K = D(\cdot, K_D)$ be the *defuzzification interface*, or *defuzzification.*
Let $f = D_K \circ I_K \circ F_K : U \to V$, then we say that f is a *fuzzy system with Knowledge Base* K or simply a *fuzzy system.*
If $m = 1$ we call f a *MISO (Multiple Inputs Single Output) fuzzy system.*
If $n, m = 1$ we call f a *SISO (Single Input Single Output) fuzzy system.*

References

[Wan92] L-X Wang. “Fuzzy systems are universal approximators”. In: *[1992 proceedings] IEEE international conference on fuzzy systems*. IEEE. 1992, pp. 1163–1170.

[WM+92] Li-Xin Wang, Jerry M Mendel, et al. “Fuzzy basis functions, universal approximation, and orthogonal least-squares learning”. In: *IEEE transactions on Neural Networks* 3.5 (1992), pp. 807–814.

[Kos94] Bart Kosko. “Fuzzy systems as universal approximators”. In: *IEEE transactions on computers* 43.11 (1994), pp. 1329–1333.

[CD96] Juan Luis Castro and Miguel Delgado. “Fuzzy systems with defuzzification are universal approximators”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 26.1 (1996), pp. 149–152.

[Jan+97] Jyh-Shing Roger Jang et al. *Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence*. eng. MATLAB curriculum series. Upper Saddle River, NJ: Prentice Hall, 1997. ISBN: 9780132610667.

[TRK15] Dan Tamir, Naphtali Rishe, and Abraham Kandel. *Fifty Years of Fuzzy Logic and its Applications*. Vol. 326. Jan. 2015. ISBN: 978-3-319-19682-4. DOI: 10.1007/978-3-319-19683-1.

[Pro+17] Piotr Prokopowicz et al. *Theory and Applications of Ordered Fuzzy Numbers A Tribute to Professor Witold Kosiński*. Jan. 2017. ISBN: 9783319596136.

[Alo+21] Jose Alonso et al. *Explainable Fuzzy Systems: Paving the Way from Interpretable Fuzzy Systems to Explainable AI Systems*. Feb. 2021. ISBN: 978-3-030-71097-2. DOI: 10.1007/978-3-030-71098-9.

[ZZW23] Xiu Zhang, Xin Zhang, and Wei Wang. *Intelligent Information Processing with Matlab*. Springer, 2023.

DEPARTMENT OF MATHEMATICS
UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II
Naples, Italy 2025
www.unina.it

UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II