Data Processing and Deep Learning: a Computational Intelligence based Perspective

Fuzzy Systems

Angelo Ciaramella

Soft Computing

- Motivation
  - All complex human actions are decisions based on such concepts:
    - driving and parking a car
    - financial/business decisions
    - law and justice
    - giving a lecture
    - listening to the professor/tutor
  - Computers need a mathematical model to express and process such complex semantics
  - Concepts in classical mathematics are inadequate for such models

Fuzziness vs Uncertainty

- Any notation is said to be fuzzy when its meaning is not fixed by sharp boundaries
  - the statement can be applied fully, to a certain degree, or not at all
  - the gradual degrees of this membership is also called fuzziness
    - very, rather, or almost

In brief …

- Boolean logic
  - Boole (1854)
- Classical set theory (1900)
  - traditional sets (boolean belonging) and set operations
- Multivariate logic
  - Russell (1920)
  - Lukasiewicz (1930)
- Fuzzy Logic theory
  - Lotfi Asker Zadeh Zadeh (1965)
    - extension of traditional sets (non boolean belonging) and operations on the elements

Zadeh in 2004 (born 1921)

In brief …

- Moreover
  - Type-2 fuzzy sets – Zadeh (1975)
  - Intuitionistics fuzzy sets – Atanassov (1986)
  - Rough sets – Pawlack (1991)
  - Neutrosophic logic – Sarandache (1998)
  - Granular computing – Pedrycz et al. (2001)
  - Explainable Fuzzy Systems – …

Fuzzy sets

- Any notation is said to be **fuzzy** when its meaning is not fixed by sharp boundaries
  - the statement can be applied fully, to a **certain degree**, or not at all
  - the **gradual degrees** of this membership is also called **fuzziness**
    - very, rather, or almost

- **Sorites Paradox**

  If a sand dune is small, adding one grain of sand to it leaves it small.
  A sand dune with a single grain is small.

  Hence all sand dunes are small.

- **Fuzzy set theory does not assume any threshold!**

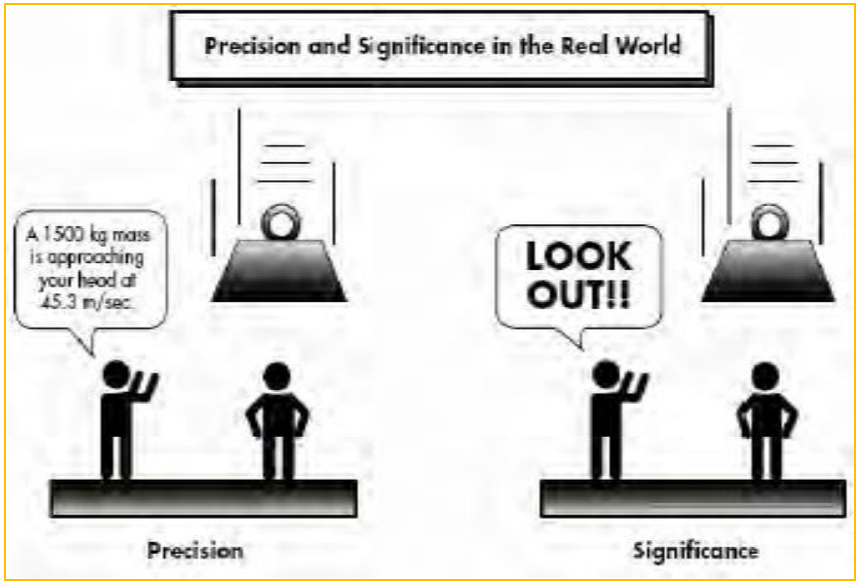
Uncertainty

- Uncertainty describes the probability of a well-defined proposition
  - Rolling a die will either lead to exactly 6 or not, but not something around 6
- Uncertainty is different from Imprecision
  - Uncertainty comes e.g. from randomness or subjective belief
- There are lots of non-standard calculi for handling uncertainty e.g.:
  - belief functions
  - possibility theory

Principle of Incompatibility

- Lotfi A. Zadeh's Principle of Incompatibility

  "Stated informally, the essence of this principle is that as the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost mutually exclusive characteristics."



Crisp Sets

- Crisp set
  - Collection of distinct objects
  - Let X be a universe of discourse and A crisp set

- *Characteristic function*

$$\mu_A(x) = \begin{cases} 1 & if\ and\ only\ if\ x \in A \\ 0 & if\ and\ only\ if\ x \notin A \end{cases}$$

$$\mu_A : X \to \{0,1\}$$

Fuzzy Sets

- Fuzzy set

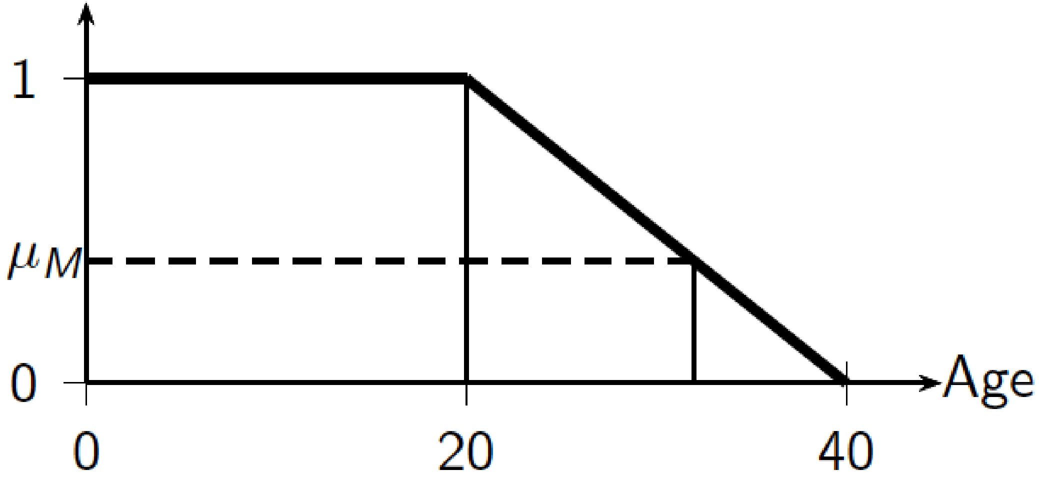
$$A = \{(x, \mu_A(x)) | x \in X\}$$

- *Membership function*
  - *generalization of the characteristic function*
  - *grade* (*or* *degree*) *of membership of* *x* *in* *A*
  - *the degree that x belongs to A*

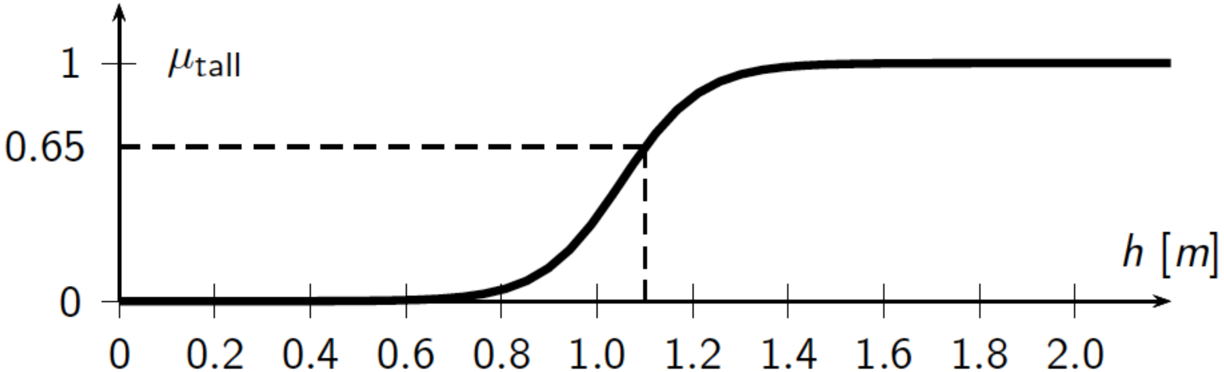
$$\mu_A: X \rightarrow [0,1]$$

Membership function

- Membership degrees



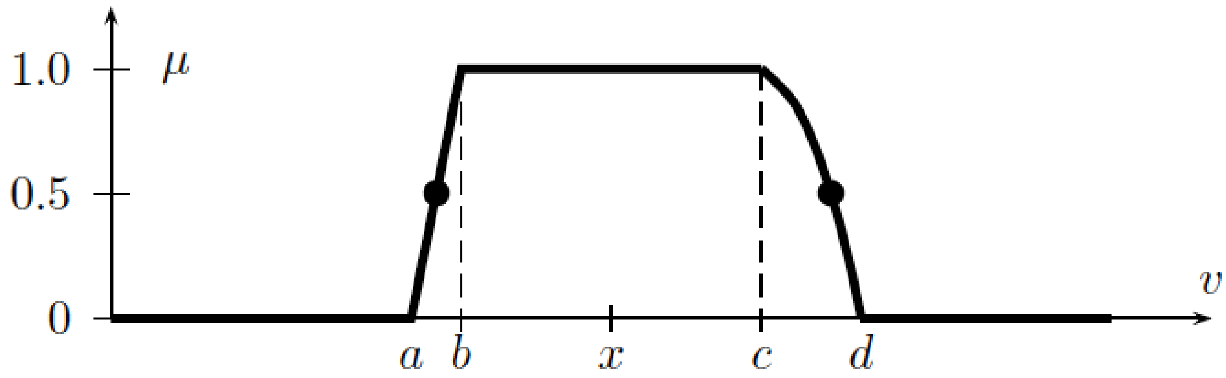
Representing young in “a young person”



Body Height of 4 Year Old Boys

Membership function

- Membership degrees



Fuzzy set *μ* characterizing velocity of rotating hard disk

$$Supp(A) = \{x \in X \mid \mu_A(x) > 0\}$$ [a,d]

$$Core(A) = \{x \in X \mid \mu_A(x) = 1\}$$ [b,c]

$$Height(A) = \sup_{x \in X} \{\mu_A(x)\}$$

Membership function

- Singleton
  - A fuzzy set A whose support is a single point in X
- Crossover
  - The element *x* in which the membershio is 0.5
- Representation of a fuzzy set

$$A = \mu_1/x_1 + \mu_2/x_2 + \cdots + \mu_i/x_i + \cdots + \mu_n/x_n = \sum_{i=1}^{n} \mu_i/x_i$$

$$A = \int_X \mu_A(x)/x$$

Semantic of Fuzzy Sets

- Fuzzy sets are relevant in three types of information-driven tasks
  - classification and data analysis
  - decision-making problems
  - approximate reasoning

- These three tasks exploit three semantics of membership grades
  - similarity
  - preference
  - possibility

Traditional logic

- There are traditional, linguistic, psychological, epistemological and mathematical schools
  - Traditional logic has been founded by Aristotle (384-322 B.C.)
  - Aristotlelian logic can be seen as formal approach to human reasoning
  - It's still used today in Artificial Intelligence for knowledge representation and reasoning about knowledge

Detail of "The School of Athens" by R. Sanzio (1509) showing Plato (left) and his student Aristotle (right).

Classical Logic

- Logic
  - studies methods/principles of reasoning
  - classical logic deals with propositions (either true or false)
  - The propositional logic handles combination of logical variables
    - Key idea - how to express *n*-ary logic functions with logic primitives
      - e.g. $\neg$, $\wedge$, $\vee$, $\rightarrow$
  - A set of logic primitives is complete if any logic function can be composed by a finite number of these primitives
    - e.g. $\{\neg, \wedge, \vee\}$, $\{\neg, \wedge\}$, $\{\neg, \rightarrow\}$, $\{\downarrow\}$ (NOR), $\{|\}$ (NAND)

Inference rules

- When a variable represented by logical formula is (*thruth values*)
  - true for all possible truth values
    - i.e. it is called tautology
  - false for all possible truth values
    - i.e. it is called contradiction

- Propositions are sentences expressed in some language and can be expressed in a canocinal form
    - x is P
    - x is the symbol of subject
    - P is the predicate
    - e.g. «Indianapolis is in Indiana»

- Logic operations are functions of two propositions and are defined via thruth tables
  - $\neg$, $\wedge$, $\vee$, $\rightarrow$

Set operators

- Let X be universe of discourse (universal set)

$$A \cap B = \{x \in X \mid x \in A \wedge x \in B\}$$

$$A \cup B = \{x \in X \mid x \in A \vee x \in B\}$$

$$A^c = \{x \in X \mid x \notin A\} = \{x \in X \mid \neg(x \in A)\}$$

$$A \subseteq B \text{ if and only if } (x \in A) \to (x \in B) \text{ for all } x \in X$$

Inference rules

- Various forms of tautologies exist to perform deductive inference (inference rules)

$$(a \wedge (a \to b)) \to b \qquad \textit{(modus ponens)}$$

$$(\neg b \wedge (a \to b)) \to \neg a \qquad \textit{(modus tollens)}$$

$$((a \to b) \wedge (b \to c)) \to (a \to c) \qquad \textit{(hypothetical syllogism)}$$

- Fuzzy Logic is an extension of bivalence logic

Basic principles

The Principle of Bivalence:

"Every proposition is either true or false."

It has been formally developed by Tarski.

Alfred Tarski (1902-1983)

Łukasiewicz suggested to replace it by
The Principle of Valence:

"Every proposition has a truth value."

Propositions can have intermediate truth value, expressed by a number from the unit interval $[0, 1]$.

Jan Łukasiewicz (1878-1956)

The multi-valued logic is to fuzzy set theory what classical logic is to set theory

Fuzzy Logic

- In 1965, Zadeh proposed a logic with values in [0, 1]

$$\neg a = 1 - a,$$
$$a \wedge b = \min(a, b),$$
$$a \vee b = \max(a, b).$$

- Membership operators

$$\neg\mu : X \to X, \neg\mu(x) = 1 - \mu(x),$$
$$\mu \wedge \mu' : X \to X(\mu \wedge \mu')(x) = \min\{\mu(x), \mu'(x)\},$$
$$\mu \vee \mu' : X \to X(\mu \vee \mu')(x) = \max\{\mu(x), \mu'(x)\}.$$

Lattice

Definition

We define the following algebraicoperators on $\mathcal{F}(X)$:

$$(\mu \wedge \mu')(x) \stackrel{\text{def}}{=} \min\{\mu(x), \mu'(x)\} \qquad \text{intersection ("AND")},$$

$$(\mu \vee \mu')(x) \stackrel{\text{def}}{=} \max\{\mu(x), \mu'(x)\} \qquad \text{union ("OR")},$$

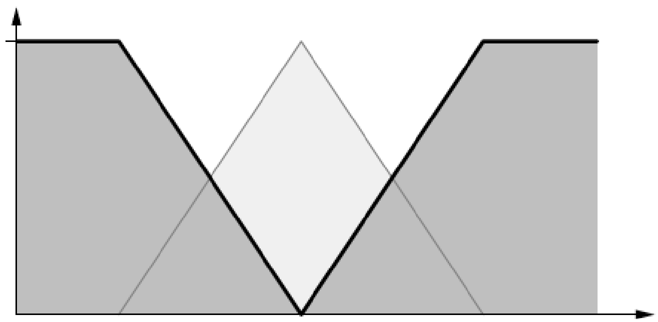
$$\neg\mu(x) \stackrel{\text{def}}{=} 1 - \mu(x) \qquad \text{complement ("NOT")}.$$

μ is subset of μ' if and only if $\mu \leq \mu'$.

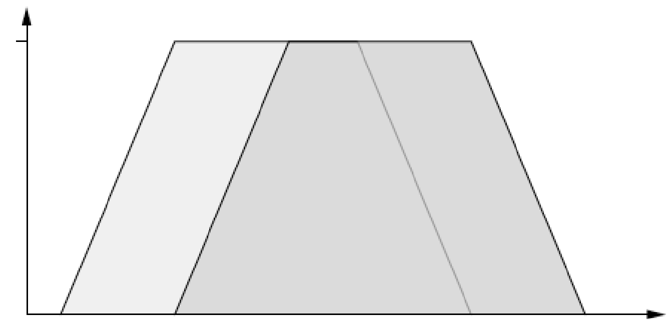
Theorem

$(\mathcal{F}(X), \wedge, \vee, \neg)$ *is a complete distributive lattice but no boolean algebra.*

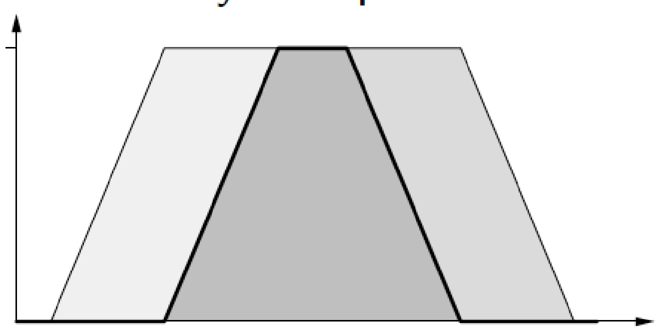
Fuzzy Set operators



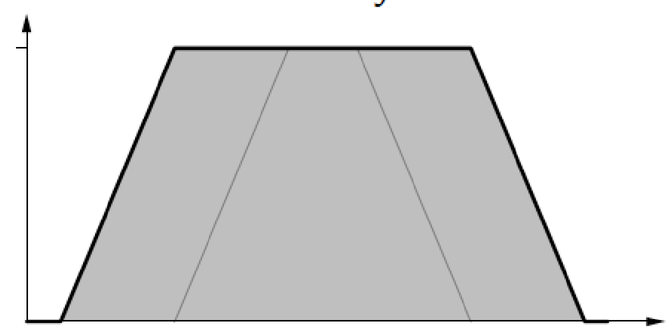
fuzzy complement



two fuzzy sets



fuzzy intersection



fuzzy union

T-norm and t-conorm

- Let *A*, *B* be Fuzzy subsets of *X*, i.e. A, B $\in F(X)$
- Their intersection and union can be defined pointwise using

$$(A \cap B)(x) = \top(A(x), B(x)) \quad \text{where} \quad \top : [0,1]^2 \to [0,1]$$
$$(A \cup B)(x) = \perp(A(x), B(x)) \quad \text{where} \quad \perp : [0,1]^2 \to [0,1].$$

$\top$ is a *triangular norm (t-norm)* $\iff$ $\top$ satisfies conditions T1-T4

$\perp$ is a *triangular conorm (t-conorm)* $\iff$ $\perp$ satisfies C1-C4

T-norm and t-conorm

Identity Law for all $x, y, z \in [0, 1]$, the following laws hold

T1: $\top(x, 1) = x \quad (A \cap X = A)$
C1: $\bot(x, 0) = x \quad (A \cup \emptyset = A)$.

Commutativity

T2: $\top(x, y) = \top(y, x) \quad (A \cap B = B \cap A)$,
C2: $\bot(x, y) = \bot(y, x) \quad (A \cup B = B \cup A)$.

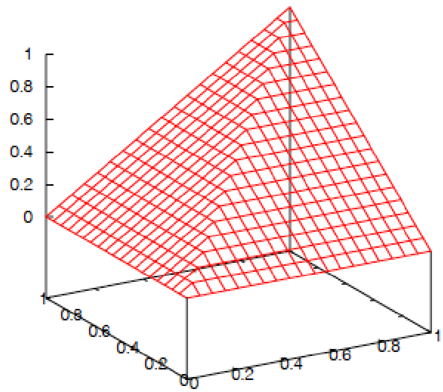
Associativity

T3: $\top(x, \top(y, z)) = \top(\top(x, y), z) \quad (A \cap (B \cap C)) = ((A \cap B) \cap C)$,
C3: $\bot(x, \bot(y, z)) = \bot(\bot(x, y), z) \quad (A \cup (B \cup C)) = ((A \cup B) \cup C)$.

Monotonicity

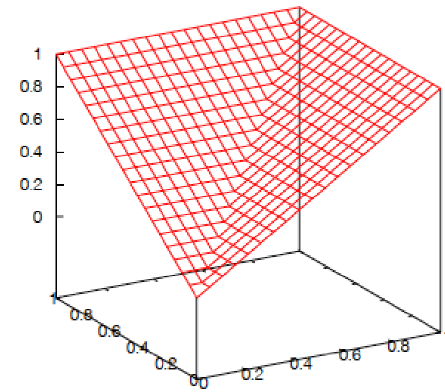
$y \leq z$ implies
T4: $\top(x, y) \leq \top(x, z)$
C4: $\bot(x, y) \leq \bot(x, z)$.

T-norm and t-conorm



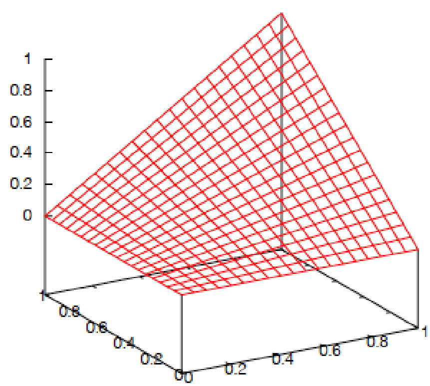
$\top_{\mathrm{min}}$

$$\top_{\mathrm{min}}(x, y) = \min(x, y)$$



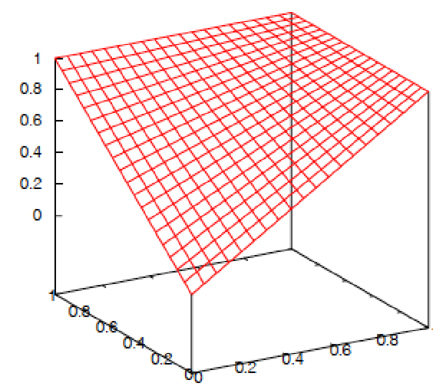
$\perp_{\mathrm{max}}$

$$\perp_{\mathrm{max}}(x, y) = \max(x, y)$$



$\top_{\mathrm{prod}}$

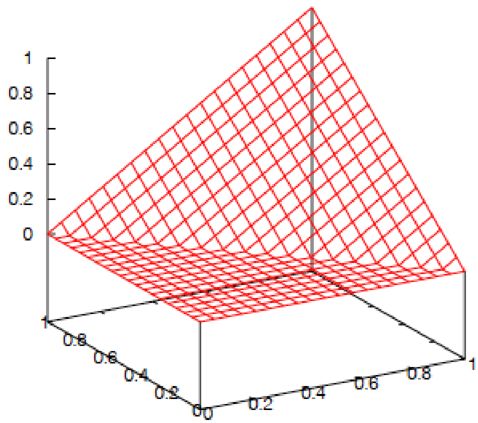
$$\top_{\mathrm{prod}}(x, y) = x \cdot y$$



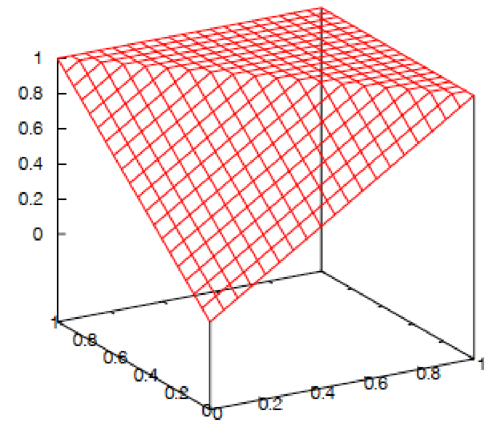
$\perp_{\mathrm{sum}}$

$$\perp_{\mathrm{sum}}(x, y) = x + y - x \cdot y$$

T-norm and t-conorm



$\top_{\text{Łuka}}$

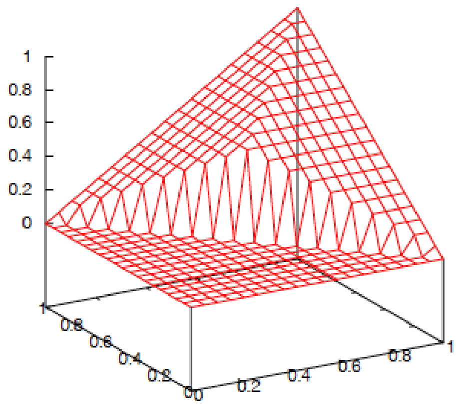


$\bot_{\text{Łuka}}$

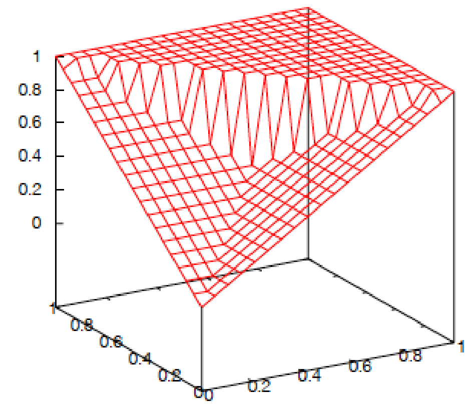
$$\top_{\text{Łuka}}(x, y) = \max\{0,\ x + y - 1\}$$

$$\bot_{\text{Łuka}}(x, y) = \min\{1,\ x + y\}$$

T-norm and t-conorm



$\top_{\min_0}$

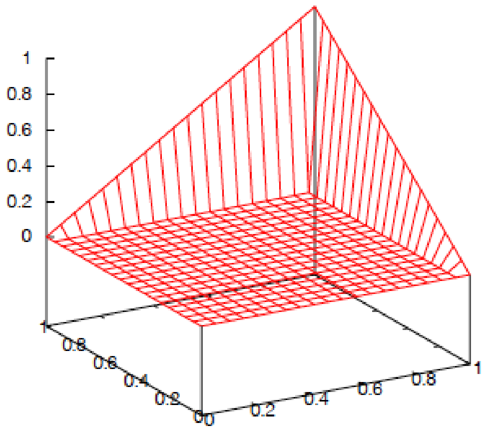


$\perp_{\max_1}$

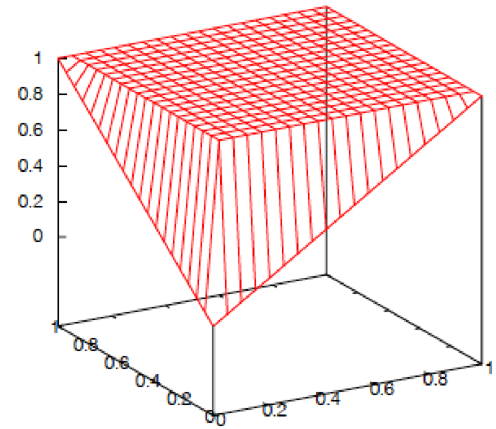
$$\top_{\min_0}(x,y) = \begin{cases} \min(x,y) & \text{if } x+y>1 \\ 0 & \text{otherwise} \end{cases} \qquad \perp_{\max_1}(x,y) = \begin{cases} \max(x,y) & \text{if } x+y<1 \\ 1 & \text{otherwise} \end{cases}$$

Nilpotent Minimum and Maximum

T-norm and t-conorm



$\top_{-1}$



$\perp_{-1}$

$$\top_{-1}(x,y) = \begin{cases} \min(x,y) & \text{if } \max(x,y) = 1 \\ 0 & \text{otherwise} \end{cases}$$

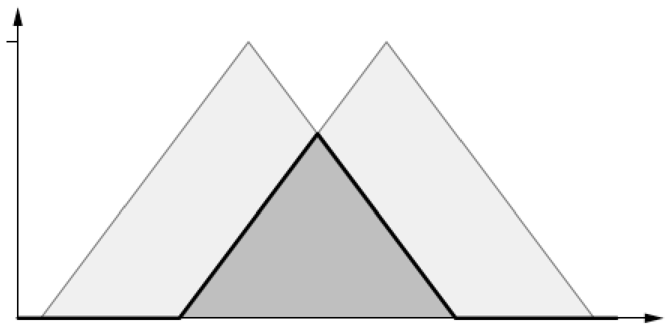
$$\perp_{-1}(x,y) = \begin{cases} \max(x,y) & \text{if } \min(x,y) = 0 \\ 1 & \text{otherwise} \end{cases}$$

Drastic Product and Sum

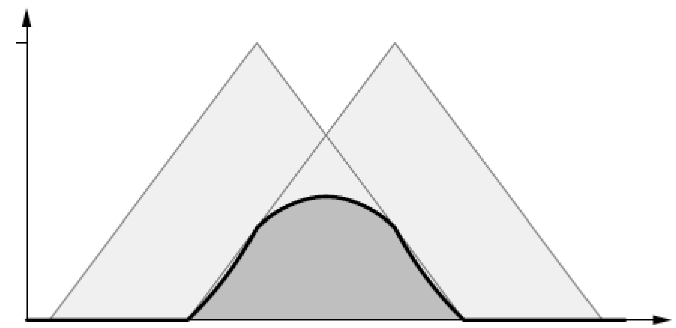
$$\top_{-1} \leq \top \leq \top_{\min}, \quad \perp_{\max} \leq \perp \leq \perp_{-1} \text{ for any } \top \text{ and } \perp$$

Fuzzy Logic

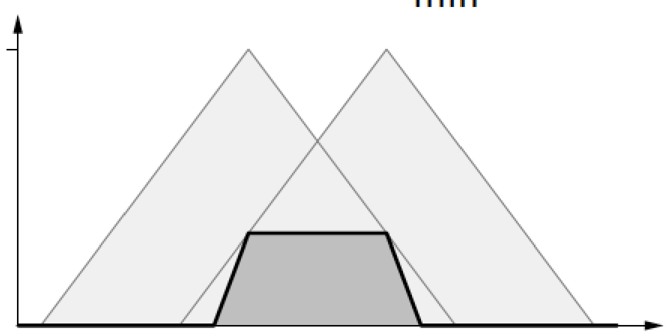
T-norm and t-conorm



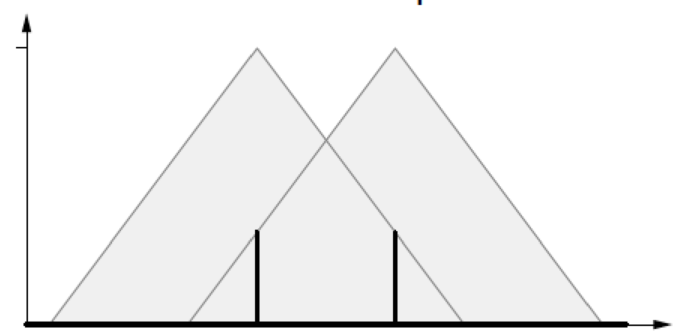
t-norm $\top_{\text{min}}$



t-norm $\top_{\text{prod}}$

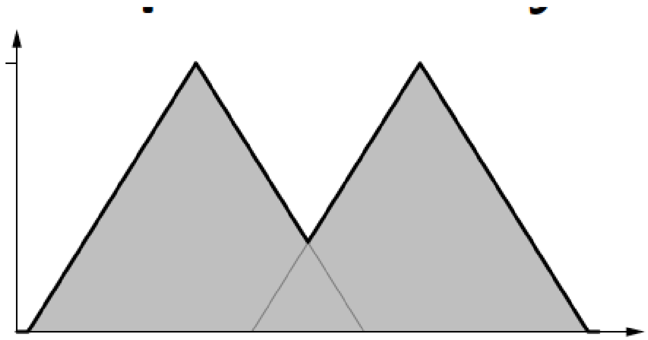


t-norm $\top_{\text{Łuka}}$

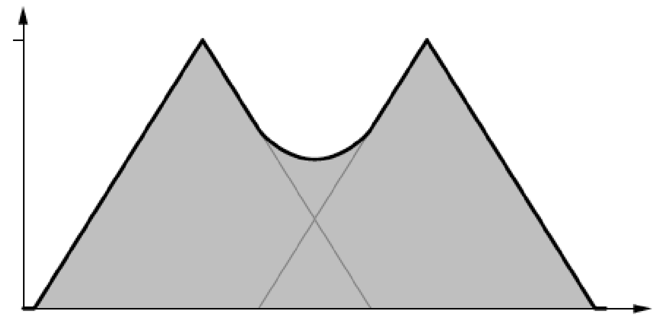


t-norm $\top_{-1}$

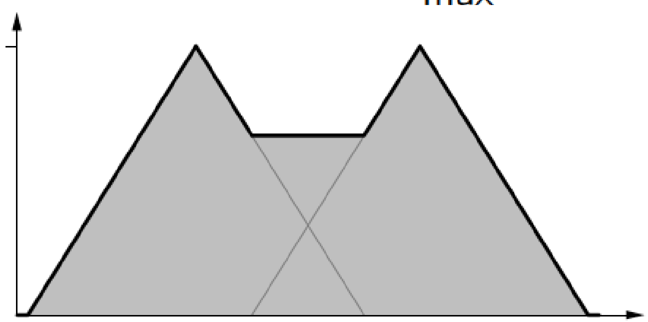
T-norm and t-conorm



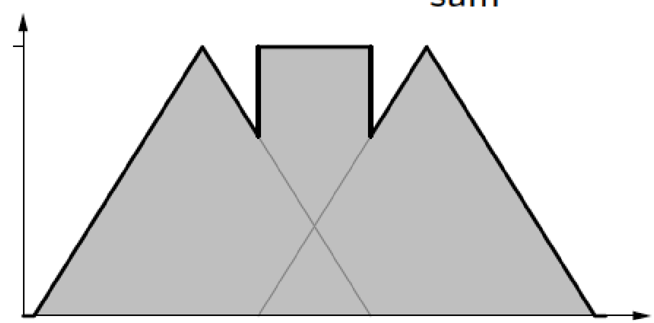
t-conorm $\perp_{\text{max}}$



t-conorm $\perp_{\text{sum}}$



t-conorm $\perp_{\text{Łuka}}$



t-conorm $\perp_{-1}$

Genaralization

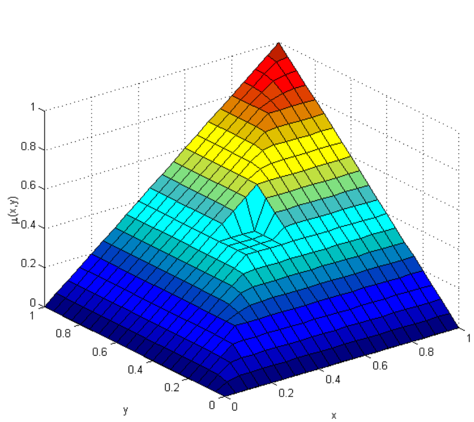
For $0 < p < \infty$ and $x, y \in [0, 1]$, define

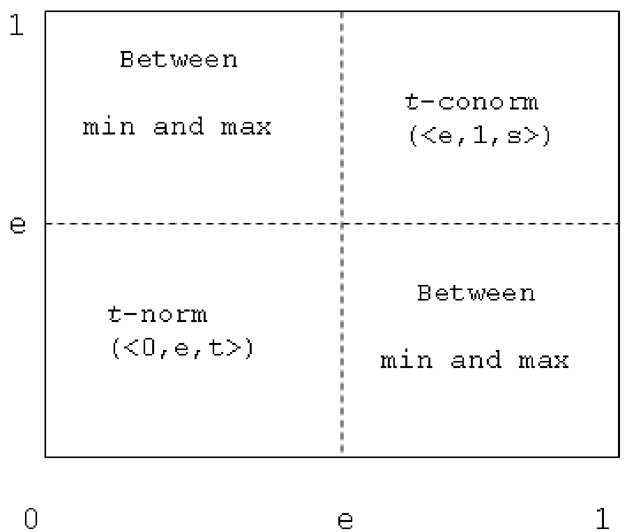
$$\top_p(x,y) = \max\left\{1 - ((1-x)^p + (1-y)^p)^{1/p},\ 0\right\},$$

$$\perp_p(x,y) = \min\left\{(x^p + y^p)^{1/p},\ 1\right\}.$$

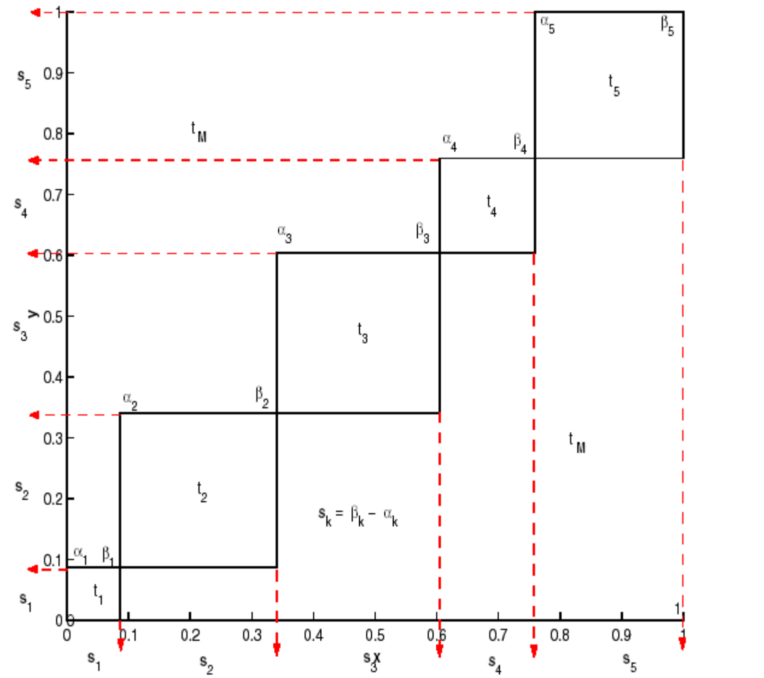
Yager norm

Genalization of norms





Uninorm

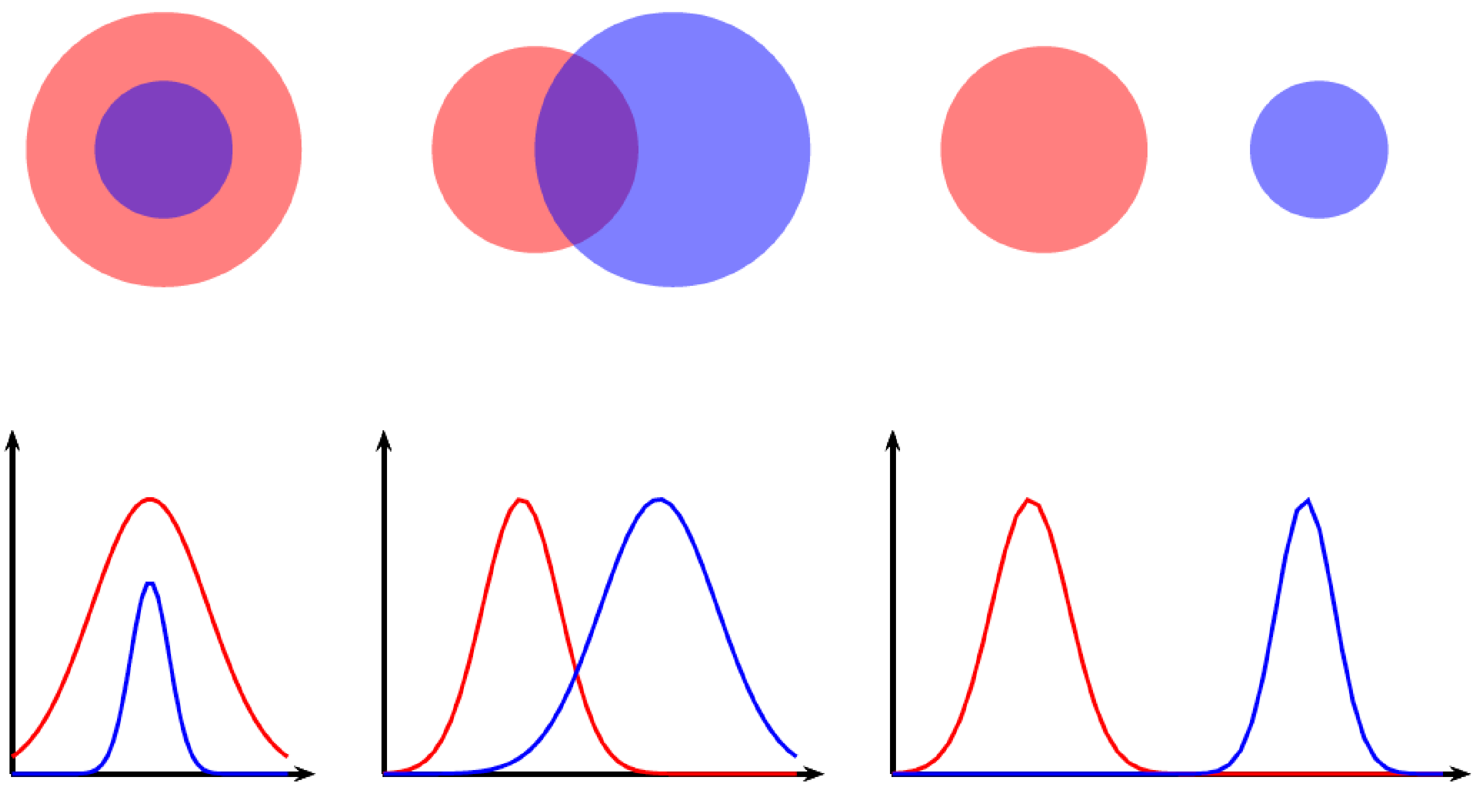


Parameters of Ordinal Sums

$$T(x,y)=\begin{cases} a_i + (b_i - a_i)\cdot T_i\left(\frac{x-a_i}{b_i-a_i}, \frac{y-a_i}{b_i-a_i}\right) & if (x,y) \in \,]a_i, b_i]^2, \\ \min(x,y) & otherwise\,. \end{cases}$$

Ordinal sums

Fuzzy implications ($a \rightarrow b$)



crisp: $\bar{x} \in A \Rightarrow \bar{x} \in B$, **fuzzy:** $x \in \mu \Rightarrow x \in \mu'$

Fuzzy implications ($a \to b$)

$[\![\varphi]\!]$	$[\![\psi]\!]$	$[\![\varphi \to \psi]\!]$
1	1	1
1	0	0
0	1	1
0	0	1

crisp: $x \in A \Rightarrow x \in B$, **fuzzy:** $x \in \mu \Rightarrow x \in \mu'$

$$I(a, b) = \neg a \vee b \qquad I(a, b) = \bot(\sim a, b)$$

$$I(a, b) = \max\{x \in \{0, 1\} \mid a \wedge x \leq b\} \qquad I(a, b) = \sup\{x \in [0, 1] \mid \top(a, x) \leq b\}$$

Fuzzy implications ($a \to b$)

$I(a, b) = \perp(\sim a, b)$ may also be written as either

$$I(a, b) = \neg a \vee (a \wedge b) \qquad \text{or}$$
$$I(a, b) = (\neg a \wedge \neg b) \vee b.$$

S-Implications

Fuzzy logical extensions are thus, respectively,

$$I(a, b) = \perp(\sim a, \top(a, b)),$$
$$I(a, b) = \perp(\top(\sim a, \sim b), b)$$

R-Implications

S-Implications ($a \to b$)

Name	$I(a,b)$	$\perp(a,b)$
Kleene-Dienes	$I_{\max}(a,b) = \max(1-a, b)$	$\max(a,b)$
Reichenbach	$I_{\mathrm{sum}}(a,b) = 1 - a + ab$	$a + b - ab$
Łukasiewicz	$I_{Ł}(a,b) = \min(1,\ 1-a+b)$	$\min(1,\ a+b)$
largest	$I_{-1}(a,b) = \begin{cases} b, & \text{if } a = 1 \\ 1-a, & \text{if } b = 0 \\ 1, & \text{otherwise} \end{cases}$	$\begin{cases} b, & \text{if } a = 0 \\ a, & \text{if } b = 0 \\ 1, & \text{otherwise} \end{cases}$

R-Implications ($a \to b$)

Name	Formula	$\top(a,b) =$
Gödel	$I_{\min}(a,b) = \begin{cases} 1, & \text{if } a \leq b \\ b, & \text{if } a > b \end{cases}$	$\min(a,b)$
Goguen	$I_{\text{prod}}(a,b) = \begin{cases} 1, & \text{if } a \leq b \\ b/a, & \text{if } a > b \end{cases}$	ab
Łukasiewicz	$I_{Ł}(a,b) = \min(1,\ 1-a+b)$	$\max(0,\ a+b-1)$
largest	$I_{\mathrm{L}}(a,b) = \begin{cases} b, & \text{if } a = 1 \\ 1, & \text{otherwise} \end{cases}$	not defined

Linguistic variables

- Each linguistic variable is defined by quintuple

$$(x, T(x), X, G, M)$$

  - name x of the variable
  - set T(x) of linguistic terms of x
  - base variable $X \subseteq IR$
  - syntactic rule G (grammar) for generating linguistic terms
  - semantic rule M that assigns meaning M(t) to every $t \in T$, i.e. $m : T \longrightarrow F(X)$

Example of linguistic variables

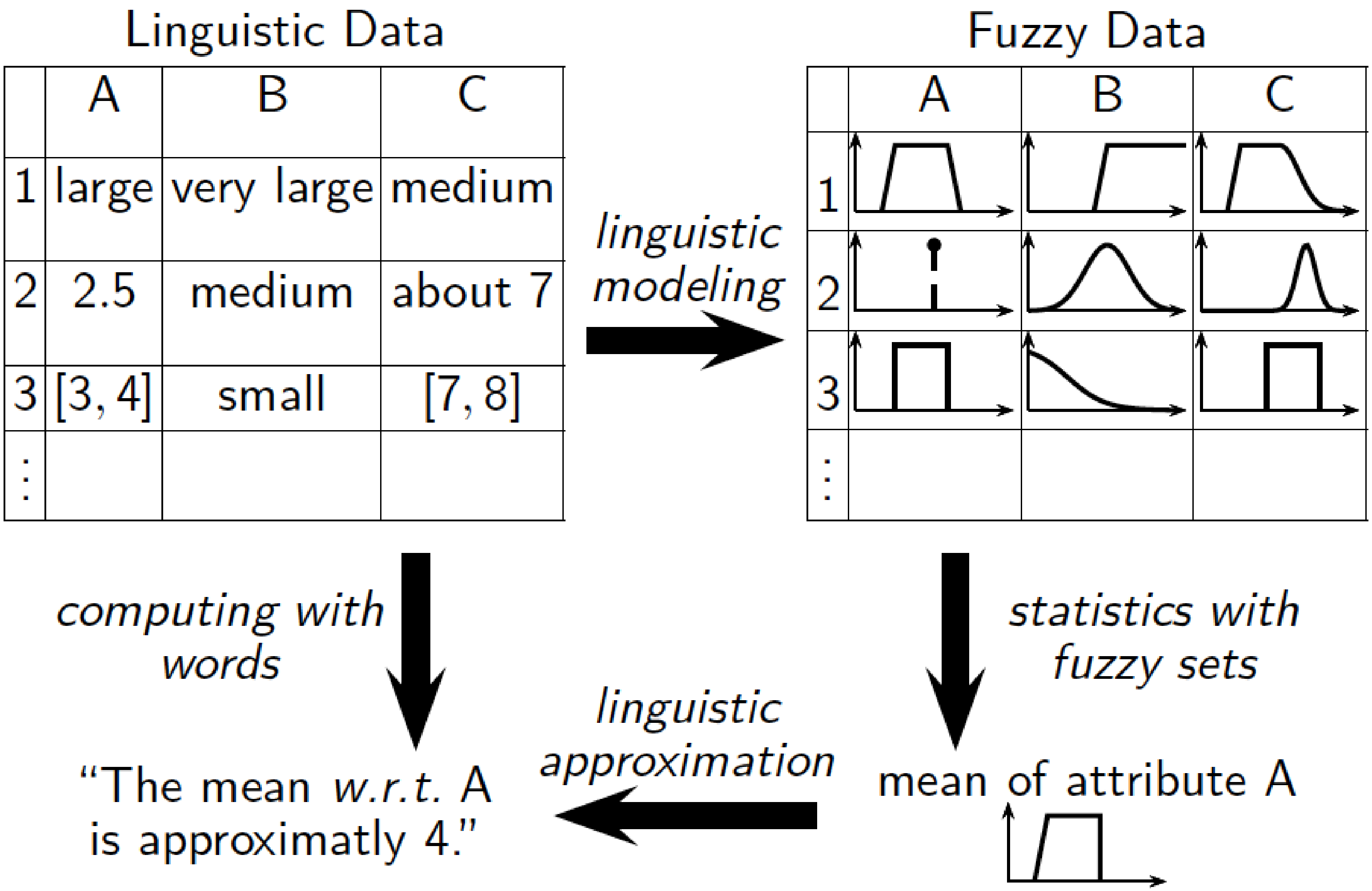
- *x = speed* a linguistic variable with $X = [0, 100]$

- Term set $T(x)$ of x
  - T(speed) = {VERY Slow, Slow, Moderate, Fast}

- G of generating the names (or the labels) of the elements in *T*(*speed*) is quite intuitive

- Semantic rulse M
  - M(slow) = the fuzzy set for «a speed below about 40 miles per hour (mph)» with membership function μ_{Slow}
  - M(moderate) = the fuzzy set for a «speed close to 55 mph» with memberhip function $\mu_{Moderate}$
  - M(Fast) = the fuzzy set for a «a speed above about 70 mph» with memberhip function μ_{Fast}

Linguistic Data



Fuzzy Logic

Example of application

- Consider the problem to model the climatic conditions of several towns
  - A tourist may want information about tourist attractions
  - Assume that linguistic random samples are based on subjective observations of selected people, e.g.
    - climatic attribute clouding
    - linguistic values cloudless, clear, fair, cloudy, . . .

Example of application

- Linguisti Modeling by an expert (Fuzzification)

The attribute *clouding* is modeled by elementary linguistic values, *e.g.*

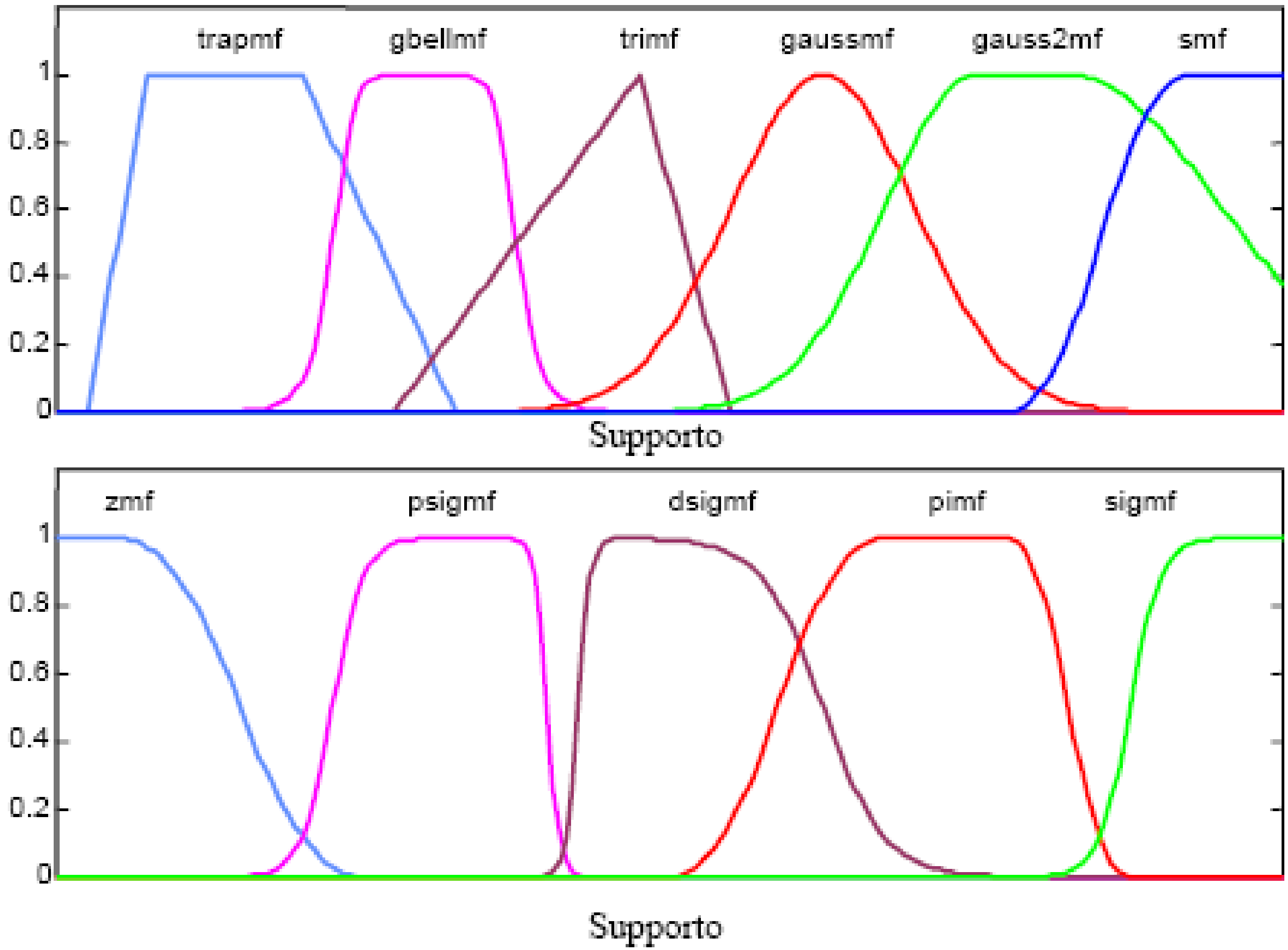
$$
\begin{aligned}
\text{cloudless} &\mapsto \text{sigmoid}(0, -0.07) \\
\text{clear} &\mapsto \text{Gauss}(25, 15) \\
\text{fair} &\mapsto \text{Gauss}(50, 20) \\
\text{cloudy} &\mapsto \text{Gauss}(75, 15) \\
\text{overcast} &\mapsto \text{sigmoid}(100, 0.07) \\
\text{exactly})(x) &\mapsto \text{exact}(x) \\
\text{approx})(x) &\mapsto \text{Gauss}(x, 3) \\
\text{between}(x, y) &\mapsto \text{rectangle}(x, y) \\
\text{approx\_between}(x, y) &\mapsto \text{trapezoid}(x - 20, x, y, y + 20)
\end{aligned}
$$

where $x, y \in [0, 100] \subset \mathbb{R}$.

Fuzzy Logic

Example of application

- Linguistic Modeling by an expert (Fuzzification)



Linguistic hedge

- Some popular linguistic hadge (modifier)
  - VERY in «VERY Young»
  - Concentration

$$\mu_{CON(A)}(x) = \left(\mu_A(x)\right)^2$$

  - Dilation

$$\mu_{DIL(A)}(x) = \left(\mu_A(x)\right)^{1/2}$$

  - Intensification

$$\mu_{INT(A)}(x) = \begin{cases} \{2\left(\mu_A(x)\right)^2 & \mu_A(x) \in [0, 0.5] \\ 1 - 2\left(1 - \mu_A(x\ )\right)^2. & otherwise \end{cases}$$

Fuzzy Relations

- Crisp relation

$$R(x_1, \ldots, x_n) = \begin{cases} 1, & \text{if and only if } (x_1, \ldots, x_n) \in R, \\ 0, & \text{otherwise.} \end{cases}$$

- Fuzzy Relation
  - is a fuzzy set defined on tuples $(x_1, \ldots, x_n)$ that may have varying degrees of membership within the relation
  - The membership grade indicates strength of the present relation between elements of the tuple

Fuzzy Relations

- Cartesian Product

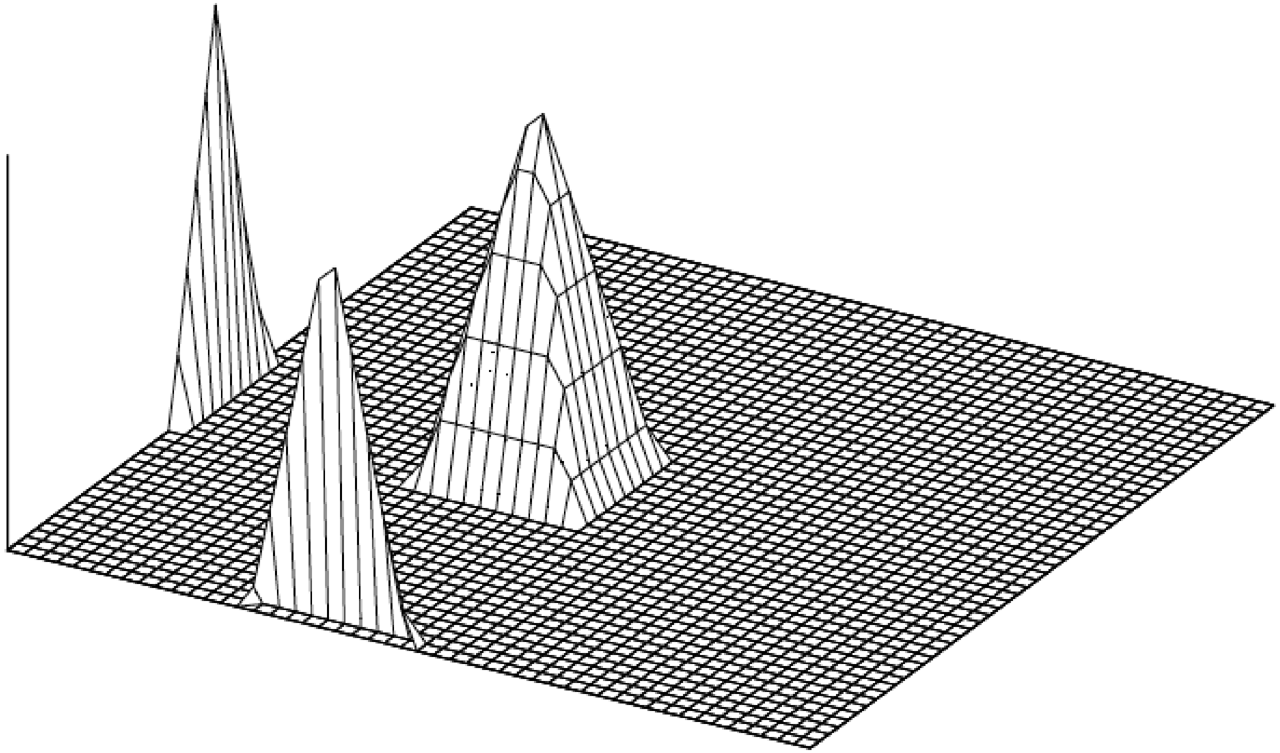
$$\mu_{A_1\times\ldots\times A_n}(x_1,\ldots,x_n) = \top\left(\mu_{A_1}(x_1),\ldots,\mu_{A_n}(x_n)\right)$$

Membership function

- 2D

$$\mu_{A\times B}(x,y) = \top\left[\mu_A(x),\ \mu_B(y)\right], \quad \forall x \in X,\ \forall y \in Y.$$

Fuzzy Relations



Cartesian product *F*(X x Y) with t-norm = min

Fuzzy Relations

- Projection

$$[R \downarrow \mathcal{Y}](\boldsymbol{y}) = \max_{\boldsymbol{x} \succ \boldsymbol{y}} R(\boldsymbol{x})$$

$(x_1,$	$x_2,$	$x_3)$	$R(x_1, x_2, x_3)$	$R_{12}(x_1, x_2)$	$R_{13}(x_1, x_3)$	$R_{23}(x_2, x_3)$	$R_1(x_1)$	$R_2(x_2)$	$R_3(x_3)$
0	0	0	0.4	0.9	1.0	0.5	1.0	0.9	1.0
0	0	1	0.9	0.9	0.9	0.9	1.0	0.9	0.9
0	0	2	0.2	0.9	0.8	0.2	1.0	0.9	1.0
0	1	0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
0	1	1	0.0	1.0	0.9	0.5	1.0	1.0	0.9
0	1	2	0.8	1.0	0.8	1.0	1.0	1.0	1.0
1	0	0	0.5	0.5	0.5	0.5	1.0	0.9	1.0
1	0	1	0.3	0.5	0.5	0.9	1.0	0.9	0.9
1	0	2	0.1	0.5	1.0	0.2	1.0	0.9	1.0
1	1	0	0.0	1.0	0.5	1.0	1.0	1.0	1.0
1	1	1	0.5	1.0	0.5	0.5	1.0	1.0	0.9
1	1	2	1.0	1.0	1.0	1.0	1.0	1.0	1.0

All possible projections

Fuzzy Relations

Projection

$(x_1,$	$x_2,$	$x_3)$	$R(x_1, x_2, x_3)$	$R_{12}(x_1, x_2)$
0	0	0	0.4	
0	0	1	0.9	$\max[R(0,0,0), R(0,0,1), R(0,0,2)] = 0.9$
0	0	2	0.2	
0	1	0	1.0	
0	1	1	0.0	$\max[R(0,1,0), R(0,1,1), R(0,1,2)] = 1.0$
0	1	2	0.8	
1	0	0	0.5	
1	0	1	0.3	$\max[R(1,0,0), R(1,0,1), R(1,0,2)] = 0.5$
1	0	2	0.1	
1	1	0	0.0	
1	1	1	0.5	$\max[R(1,1,0), R(1,1,1), R(1,1,2)] = 1.0$
1	1	2	1.0	

Projection R_{12}

Binary Fuzzy Relations

- Binary Fuzzy Relations
  - $R(X,Y)$
- Standard composition

$$[P \circ Q](x,z) = \sup_{y \in Y} \min\{P(x,y),\ Q(y,z)\},\ \ \forall x \in X,\ \forall z \in Z$$

If Y is finite supe operator is replaced by max (max-min composition)

- Associativity

$$[P(X,Y)] \circ Q(Y,Z)] \circ R(Z,W) = P(X,Y) \circ [Q(Y,Z) \circ R(Z,W)]$$

Example of composition

$$\boldsymbol{P} \circ \boldsymbol{Q} = \boldsymbol{R}$$

$$\begin{bmatrix} .3 & .5 & .8 \\ 0 & .7 & 1 \\ .4 & .6 & .5 \end{bmatrix} \circ \begin{bmatrix} .9 & .5 & .7 & .7 \\ .3 & .2 & 0 & .9 \\ 1 & 0 & .5 & .5 \end{bmatrix} = \begin{bmatrix} .8 & .3 & .5 & .5 \\ 1 & .2 & 5 & .7 \\ .5 & .4 & .5 & .5 \end{bmatrix}$$

max-min composition

Example of composition

- Problem
  - The three students are planning to take one of the courses based on their different preferences.
  - By fuzzy relations we help them in their decision making

- Serts
  - X = {Peter, Mary, John}
  - Y = {Theory, Application, Hardware, Programming}
  - Z = {Fuzzy Theory, Fuzzy Controls, Neural Networks, Expert Systems}

Example of composition

	Theory	Applica tion	Hardwa rer	Progra mming
Peter	0.2	1	0.8	0.1
Mary	1	0.1	0	0.5
John	0.5	0.9	0.5	1

P(X,Y)

	FT	FC	NN	ES
Theory	1	0.5	0.6	0.1
Application	0.2	1	0.8	0.8
Hardware	0	0.3	0.7	0
Programming	0.1	0.5	0.8	1

P(Y,Z)

Example of composition

	FT	FC	NN	ES
Peter	0.2	1	0.8	0.8
Mary	1	0.5	0.6	0.5
John	0.5	0.9	0.8	1

P(X,Y) o P(Y,Z)

Example of composition

Consider the following fuzzy relations for airplanes:

- relation A between maximal speed and maximal height,
- relation B between maximal height and the type.

A	h_1	h_2	h_3
s_1	1	.2	0
s_2	.1	1	0
s_3	0	1	1
s_4	0	.3	1

B	t_1	t_2
h_1	1	0
h_2	.9	1
h_3	0	.9

Types of Airplanes (Speed, Height, Type)

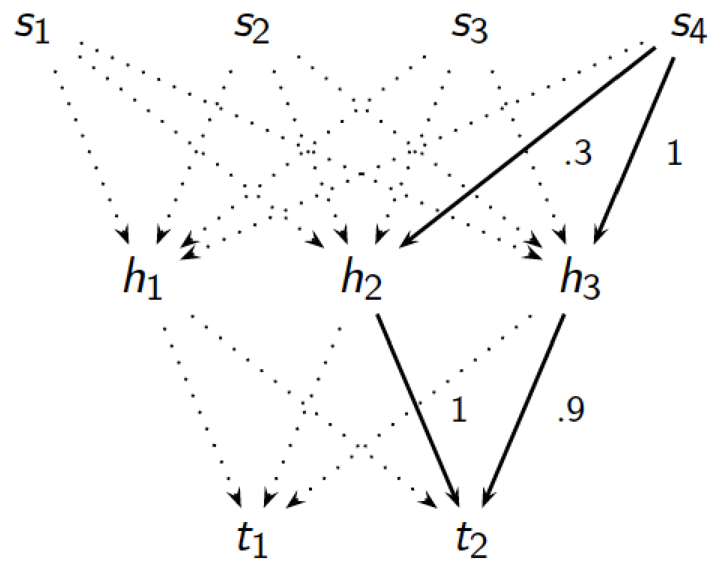
Example of composition

matrix multiplication scheme

			1	0
A	$\circ$	***B***	.9	1
			0	.9
1	.2	0	1	.2
.1	1	0	.9	1
0	1	1	.9	1
0	.3	1	.3	.9

$A \circ B$ speed-type relation

flow scheme



$$
\begin{aligned}
(A \circ B)(s_4, t_2) &= \max\{\min\{.3,\ 1\},\ \min\{1,\ .9\}\} \\
&= .9
\end{aligned}
$$

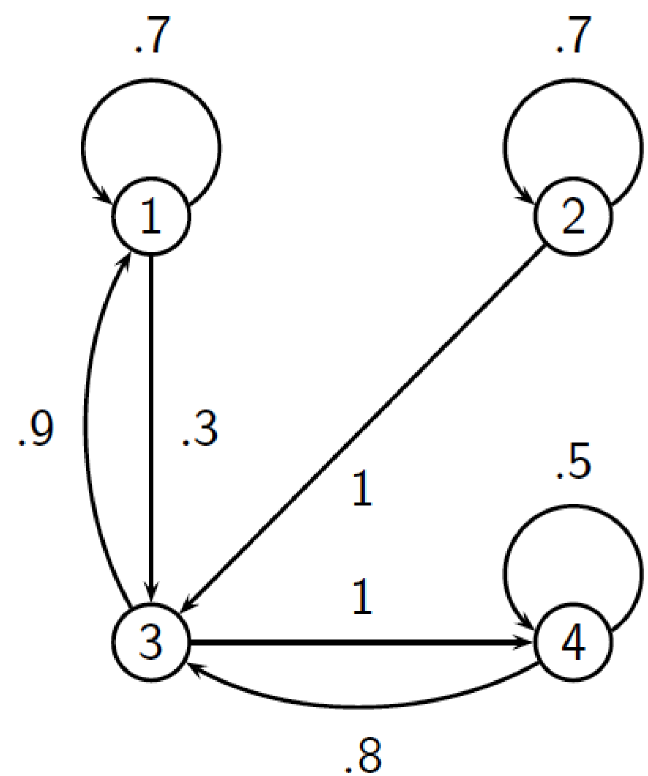
Types of Airplanes (Speed, Height, Type)

BR on single set

An example of $R(X, X)$ defined on $X = \{1, 2, 3, 4\}$.

Two different representation are shown below.

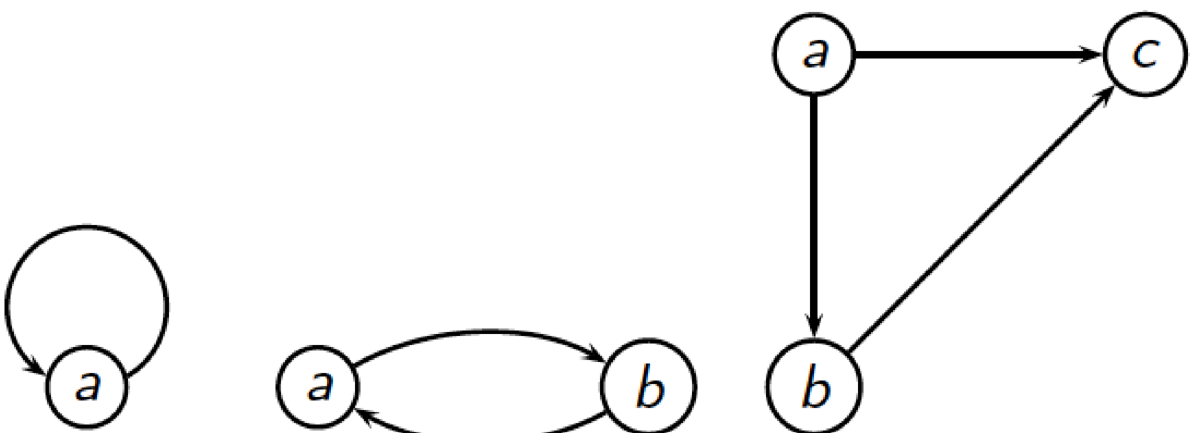
	1	2	3	4
1	.7	0	.3	0
2	0	.7	1	0
3	.9	0	0	1
3	0	0	.8	.5



BR on single set

A crisp relation $R(X, X)$ is called

- *reflexive* if and only if $\forall x \in X : (x, x) \in R$,
- *symmetric* if and only if $\forall x, y \in X : (x, y) \in R \leftrightarrow (y, x) \in R$,
- *transitive* if and only if $(x, z) \in R$ whenever both $(x, y) \in R$ and $(y, z) \in R$ for at least one $y \in X$.



reflexivity symmetry transitivity

Properties of Fuzzy Relations

A fuzzy relation $R(X, X)$ is called

- *reflexive* if and only if $\forall x \in X : R(x, x) = 1$,
- *symmetric* if and only if $\forall x, y \in X : R(x, y) = R(y, x)$,
- *transitive* if it satisfies

$$R(x, z) \geq \max_{y \in Y} \min\{R(x, y),\ R(y, z)\}, \quad \forall (x, z) \in X^2$$

A FB Relation that is reflexive, symmetric and transitive is called
Fuzzy Equivalence Relation

Approximate Reasoning

- Rules of inference govern the deduction of a proposition q from a set of premises $\{p_1, p_2, \ldots, p_n\}$

- Four principal modes of fuzzy reasoning
  - Categorical reasoning
  - Qualitative reasoning
  - Syllogistic reasoning
  - Dispositional reasoning

Qualitative Reasoning

- IF-THEN rules in which the pre-conditions and consequences involve fuzzy or linguistic variables

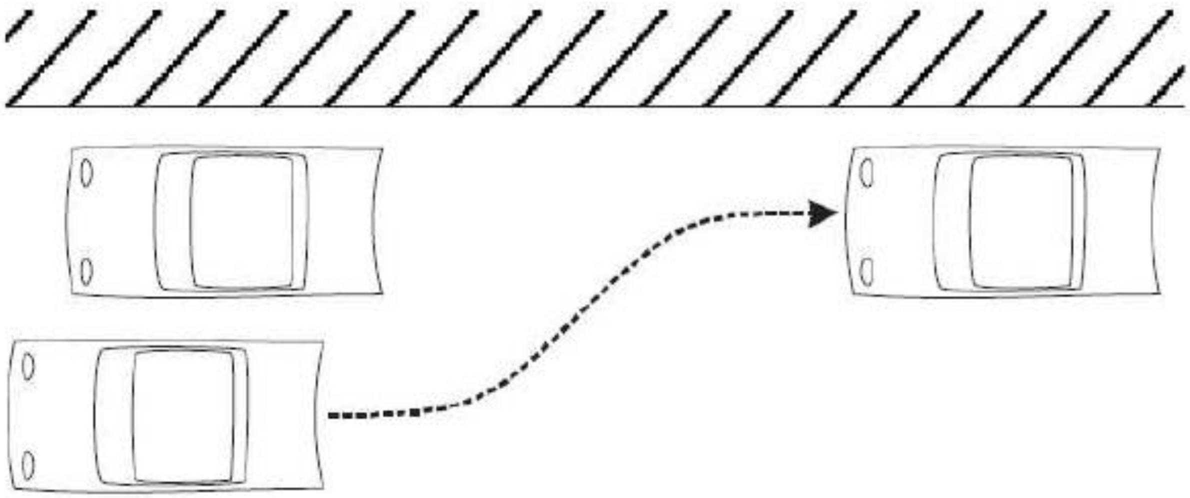
R_1: IF x is A_1 and y is B_1 THEN z is C_1
R_2: IF x is A_2 and y is B_2 THEN z is C_2

. . .

R_1: IF x is A_n and y is B_n THEN z is C_n

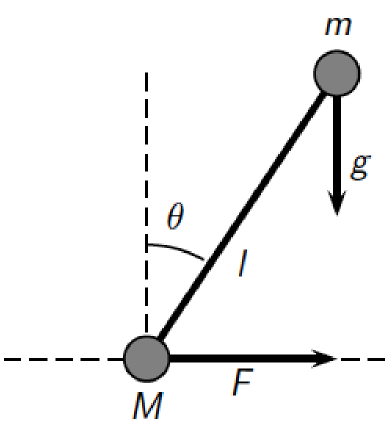
Fuzzy Control Systems

- Basic idea of FLC
  - incorporate the «expert experience» by fuzzy control rules rather than a complicated dynamic model



Parking a car backwards

Example of Fuzzy Controls



Balance an upright standing pole by moving its foot.

Lower end of pole can be moved unrestrained along horizontal axis.

Mass m at foot and mass M at head.

Influence of mass of shaft itself is negligible.

Determine force F (control variable) that is necessary to balance pole standing upright.

That is measurement of following output variables:

- angle θ of pole in relation to vertical axis,
- change of angle, *i.e.* triangular velocity $\dot{\theta} = \frac{d\theta}{dt}$.

Both should converge to zero.

Cartpole problem

Classical solution

Differential equation of cartpole problem:

$$(M+m)\sin^2\theta \cdot l \cdot \ddot{\theta} + m \cdot l \cdot \sin\theta\cos\theta \cdot \dot{\theta}^2 - (M+m) \cdot g \cdot \sin\theta = -F \cdot \cos\theta$$

Compute $F(t)$ such that $\theta(t)$ and $\dot{\theta}(t)$ converge towards zero quickly.

Physical analysis demands knowledge about physical process.

Cartpole problem

Fuzzy partitioning

left fuzzy set:

$$\mu_1^{(1)} : [a, b] \to [0, 1]$$
$$x \mapsto \begin{cases} 1, & \text{if } x \leq x_1 \\ 1 - \min\{\varepsilon \cdot (x - x_1),\ 1\} & \text{otherwise} \end{cases}$$

right fuzzy set:

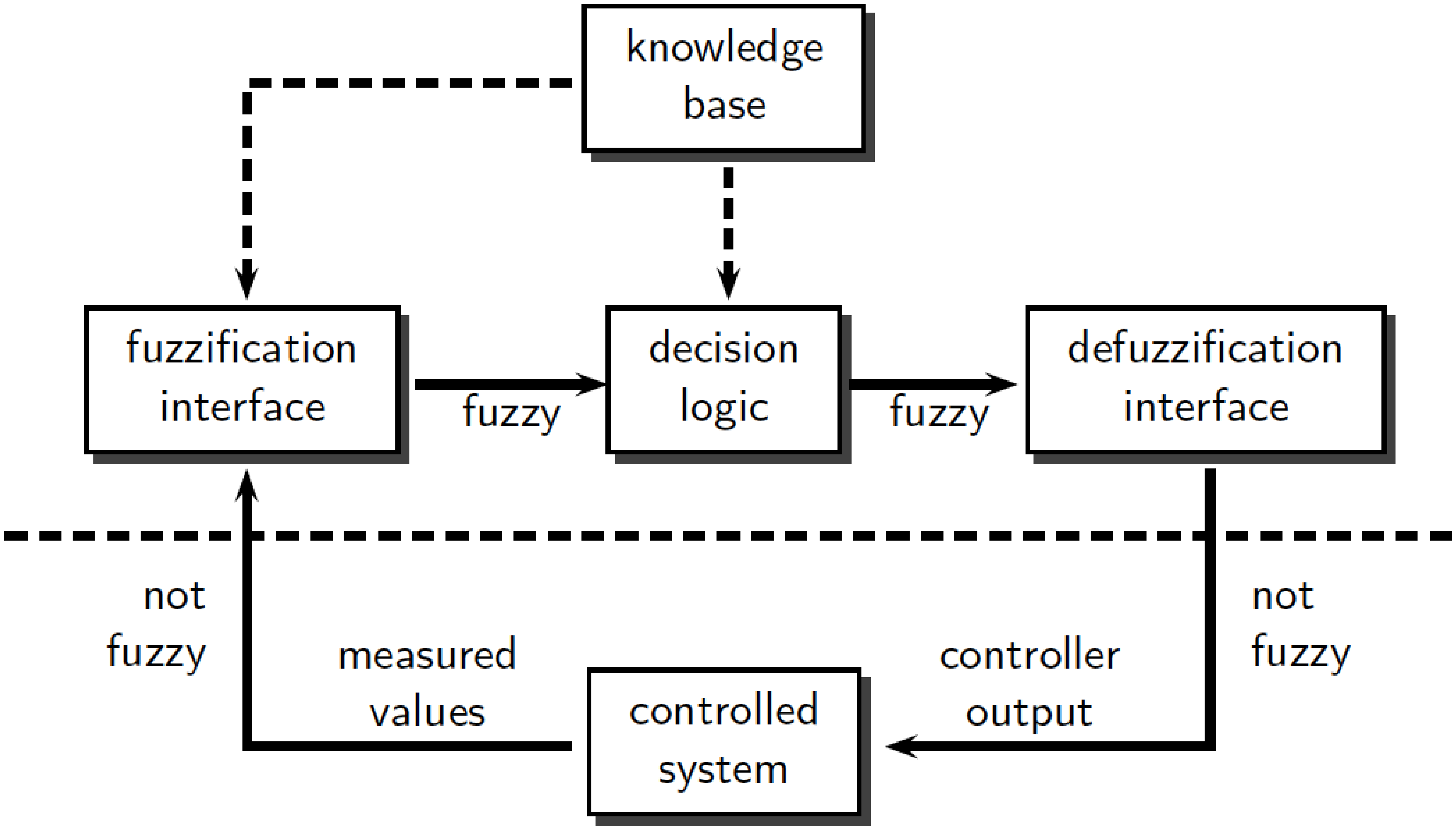
$$\mu_{p_1}^{(1)} : [a, b] \to [0, 1]$$
$$x \mapsto \begin{cases} 1, & \text{if } x_{p_1} \leq x \\ 1 - \min\{\varepsilon \cdot (x_{p_1} - x),\ 1\} & \text{otherwise} \end{cases}$$

Fuzzy partitioning

$\dot{\theta}$ \ θ	nb	nm	ns	az	ps	pm	pb
nb			ps	pb			
nm				pm			
ns	nm		ns	ps			
az	nb	nm	ns	az	ps	pm	pb
ps				ns	ps		pm
pm				nm			
pb				nb	ns		

19 rules for cartpole problem, often not necessary to determine all table entries *e.g.*

Fuzzy Controller



Fuzzy Controller

Fuzzification interface

- receives current input value (eventually maps it to suitable domain),
- converts input value into linguistic term or into fuzzy set.

Knowledge base (consists of **data base** and **rule base**)

- Data base contains information about boundaries, possible domain transformations, and fuzzy sets with corresponding linguistic terms.
- Rule base contains linguistic control rules.

Decision logic (represents processing unit)

- computes output from measured input accord. to knowledge base.

Defuzzification interface (represents processing unit)

- determines crisp output value
  (and eventually maps it back to appropriate domain).

Fuzzy knowledge base

- General schema

$$
\begin{array}{ll}
\textit{Rule 1:} & \textbf{if } \mathrm{X} \text{ is } M_1, \textbf{ then } \mathrm{Y} \text{ is } N_1 \\
\textit{Rule 2:} & \textbf{if } \mathrm{X} \text{ is } M_2, \textbf{ then } \mathrm{Y} \text{ is } N_2 \\
\vdots & \vdots \\
\textit{Rule } r: & \textbf{if } \mathrm{X} \text{ is } M_r, \textbf{ then } \mathrm{Y} \text{ is } N_r \\
\textit{Fact:} & \mathrm{X} \text{ is } M' \\
\hline
\textit{Conclusion:} & \mathrm{Y} \text{ is } N'
\end{array}
$$

Inference engine

- This is the kernel of the FLC in modelling human decision making within the conceptual framework of FL and approximate resoning

- Generalized modus ponens (categorical resoning)

Premise 1: IF x is A THEN y is B
Premise 2: x is A'

Conclusion: y is B'

Inference engine

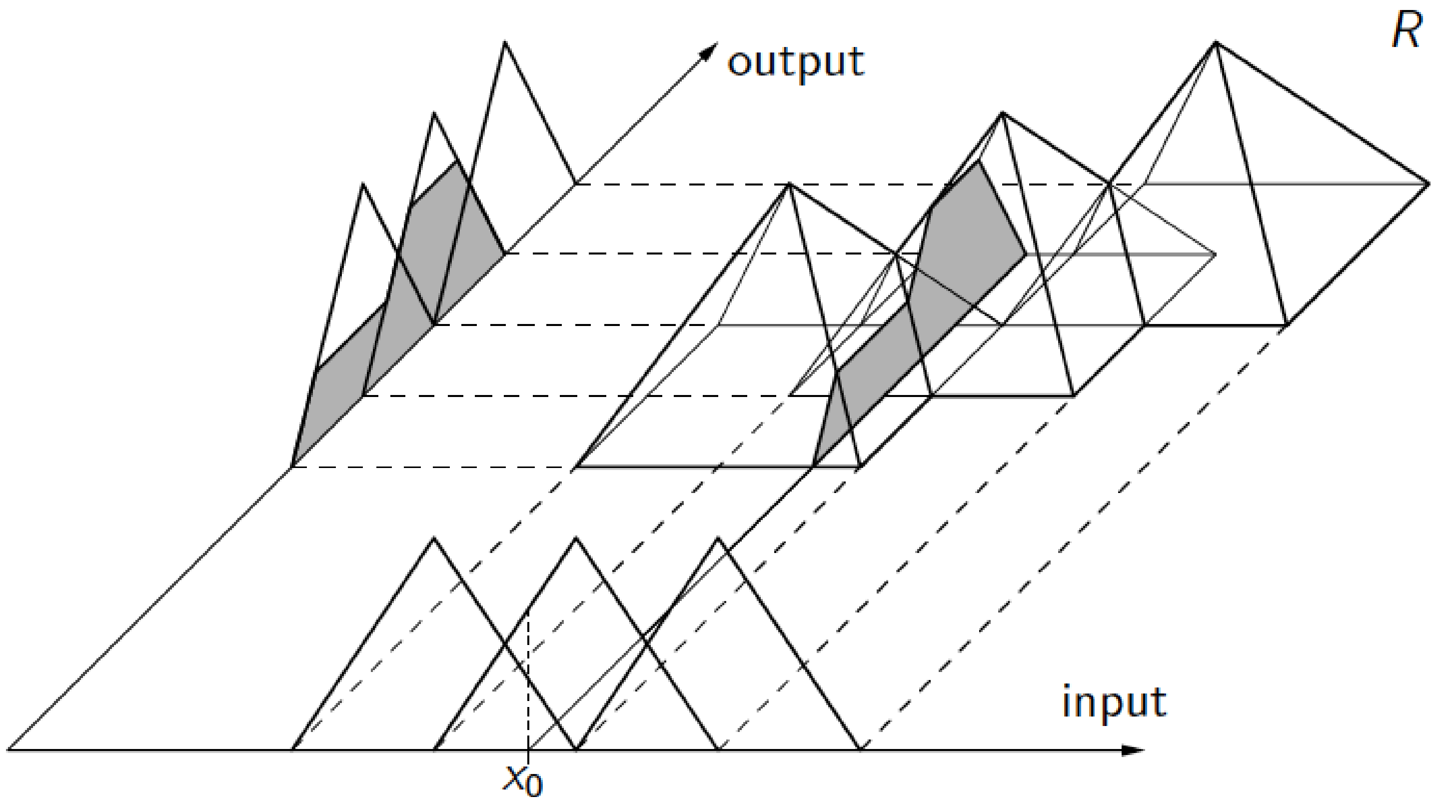
- In general a fuzzy control rule (Premise 1) is a fuzzy relation which is expressed as a fuzzy implication

$$R = A \to B$$

- conclusion can be obtained as

$$B' = A' \circ (A \to B)$$

Approximate Reasoning

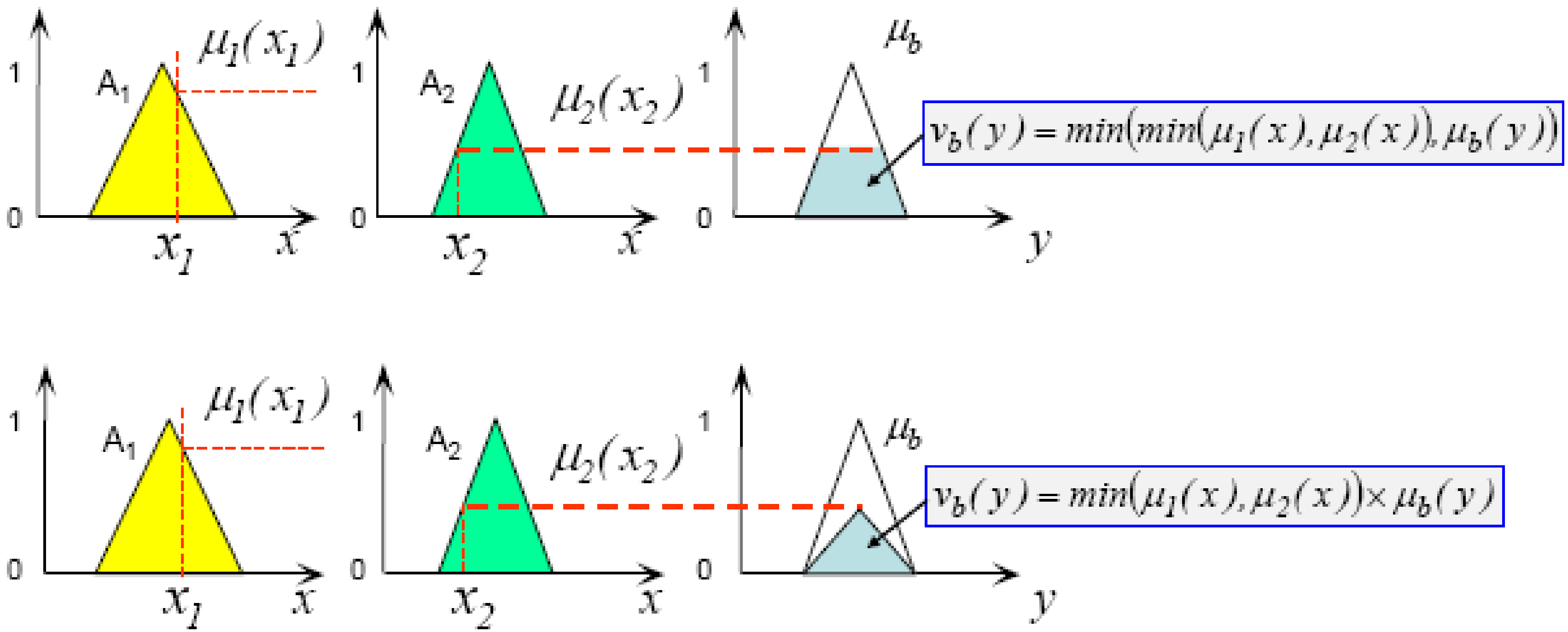


3 Fuzzy Rules

Inference (Mamdani)

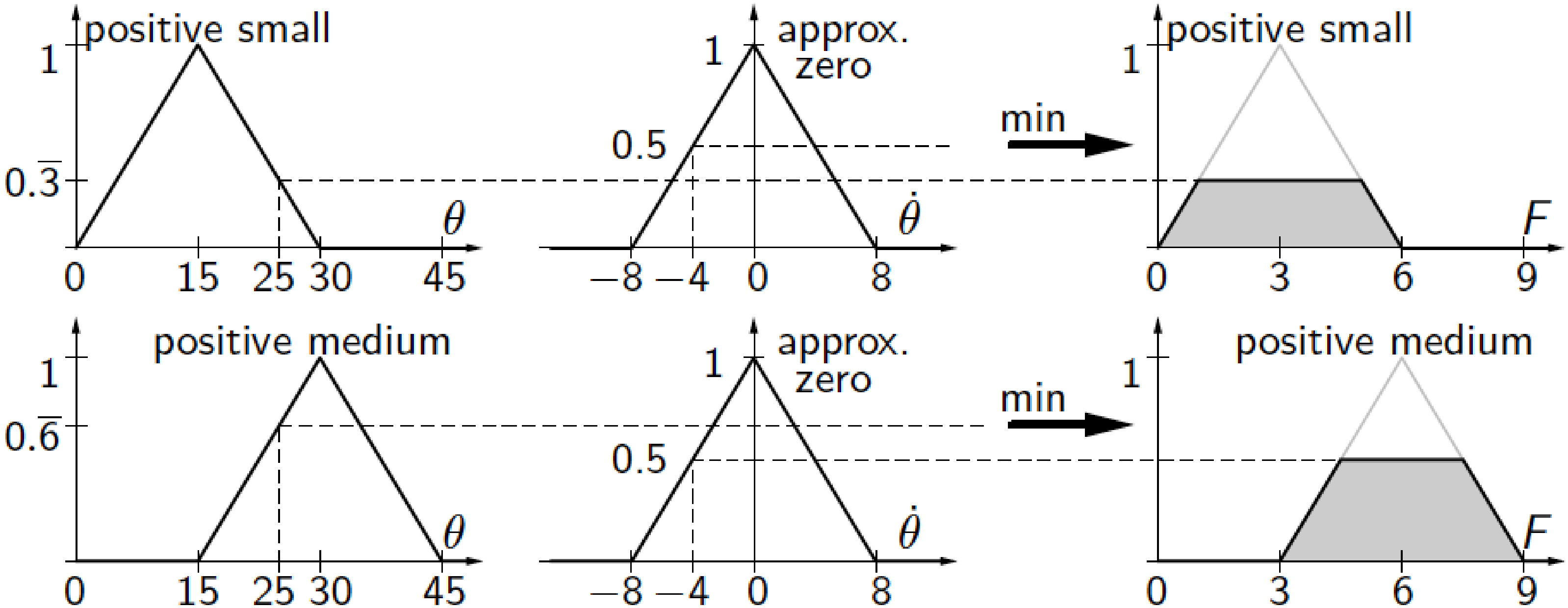
$$IF\ (x_1\ is\ A_1)\ AND\ (x_2\ is\ A_2)\ THEN\ y\ is\ B$$

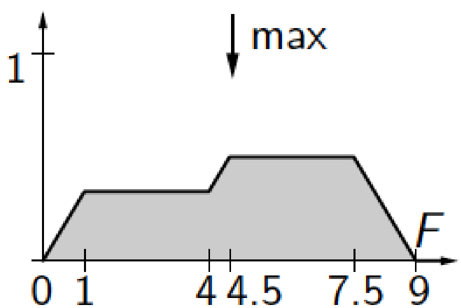
$$\mu_b(y) = \left(\mu_1(x^*) \wedge \mu_2(x^*)\right) \wedge \mu_b(y)$$



Mamdani based inference

Approximate Reasoning





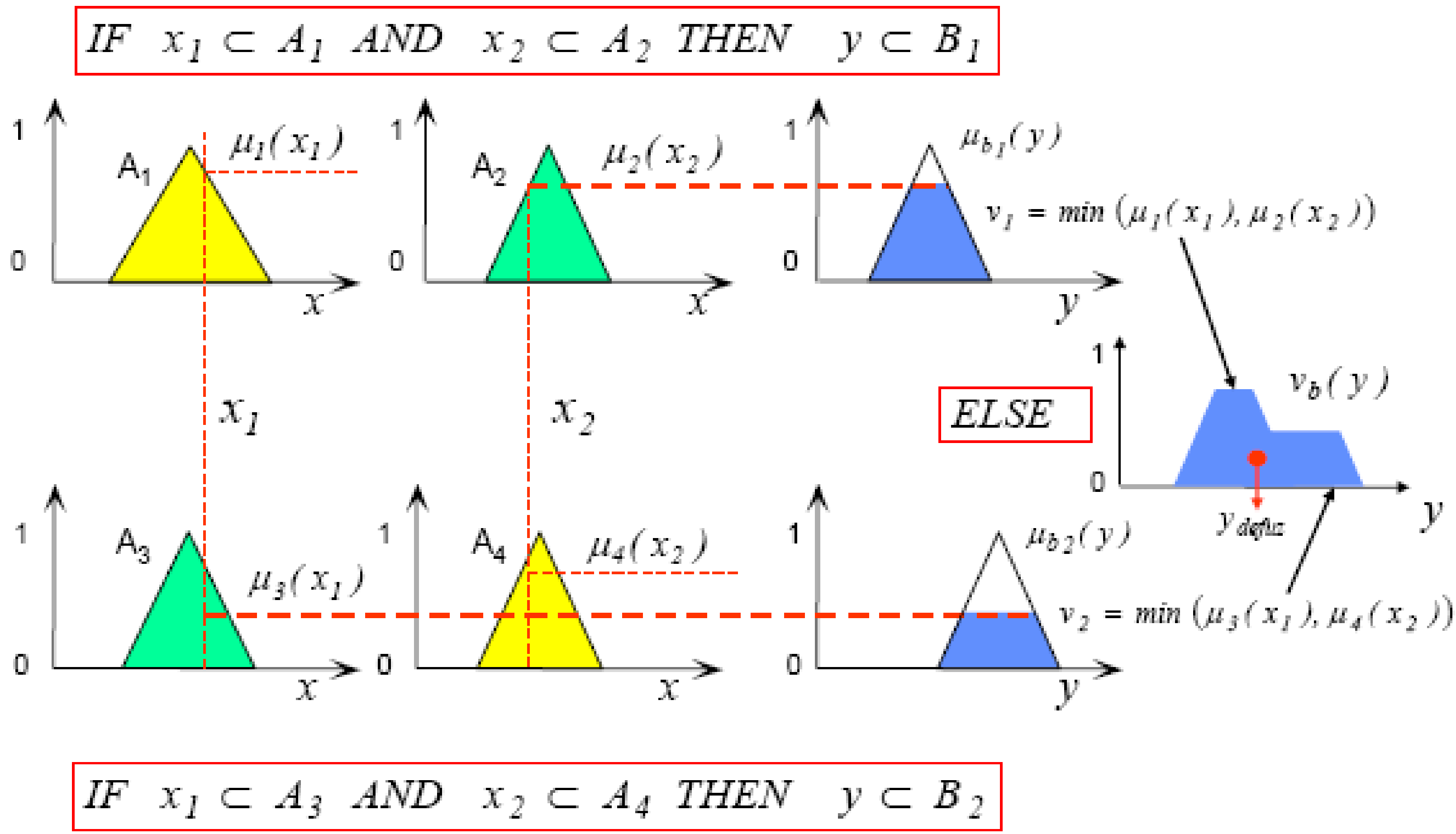
Rule evaluation for Mamdani-Assilian controller.

Input tuple $(25, -4)$ leads to fuzzy output.

Crisp output is determined by defuzzification.

Mamdani controller

Defuzzification



Inference and defuzzification

Defuzzification - Max Criterion Method

Choose an arbitrary $y \in Y$ for which $\mu^{\text{output}}_{x_1,\ldots,x_n}$ reaches the maximum membership value.

Advantages:

- Applicable for arbitrary fuzzy sets.
- Applicable for arbitrary domain Y (even for $Y \neq \mathbb{R}$).

Defuzzification – Mean of Maxima (MOM)

Preconditions:

(i) Y is interval

(ii) $Y_{\text{Max}} = \{y \in Y \mid \forall y' \in Y : \mu^{\text{output}}_{x_1,\ldots,x_n}(y') \leq \mu^{\text{output}}_{x_1,\ldots,x_n}(y)\}$ is non-empty and measurable

(iii) Y_{Max} is set of all $y \in Y$ such that $\mu^{\text{output}}_{x_1,\ldots,x_n}$ is maximal

Crisp output value $=$ mean value of Y_{Max}.

if Y_{Max} is finite:

$$\eta = \frac{1}{|Y_{\text{Max}}|} \sum_{y_i \in Y_{\text{Max}}} y_i$$

if Y_{Max} is infinite:

$$\eta = \frac{\int_{y \in Y_{\text{Max}}} y \, dy}{\int_{y \in Y_{\text{Max}}} dy}$$

MOM can lead to discontinuous control actions.

Defuzzification – Center of Gravity (COG)

Same preconditions as MOM method.

$\eta =$ center of gravity/area of $\mu^{\text{output}}_{x_1,\ldots,x_n}$

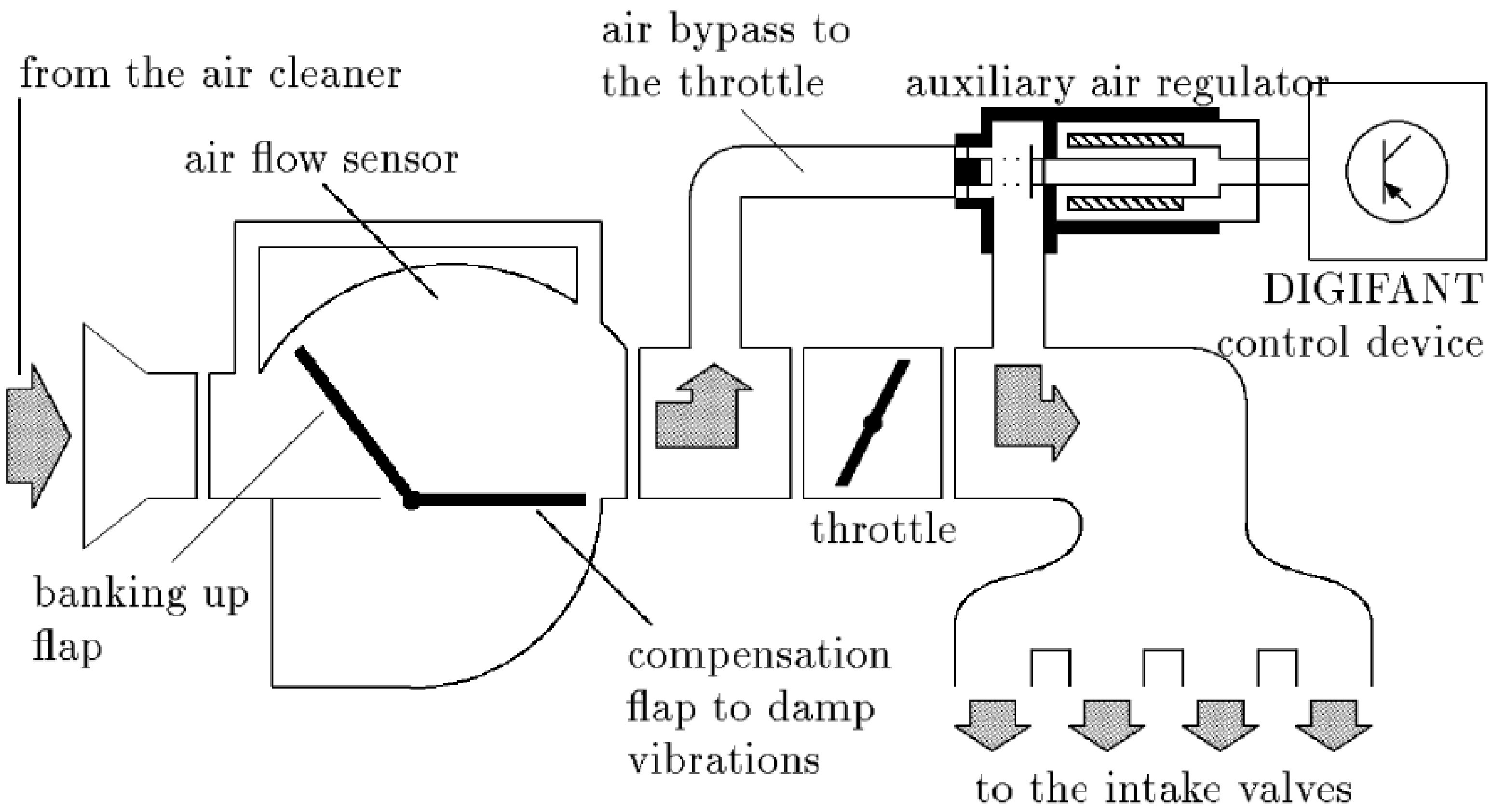
If Y is finite, then

$$\eta = \frac{\sum_{y_i \in Y} y_i \cdot \mu^{\text{output}}_{x_1,\ldots,x_n}(y_i)}{\sum_{y_i \in Y} \mu^{\text{output}}_{x_1,\ldots,x_n}(y_i)}.$$

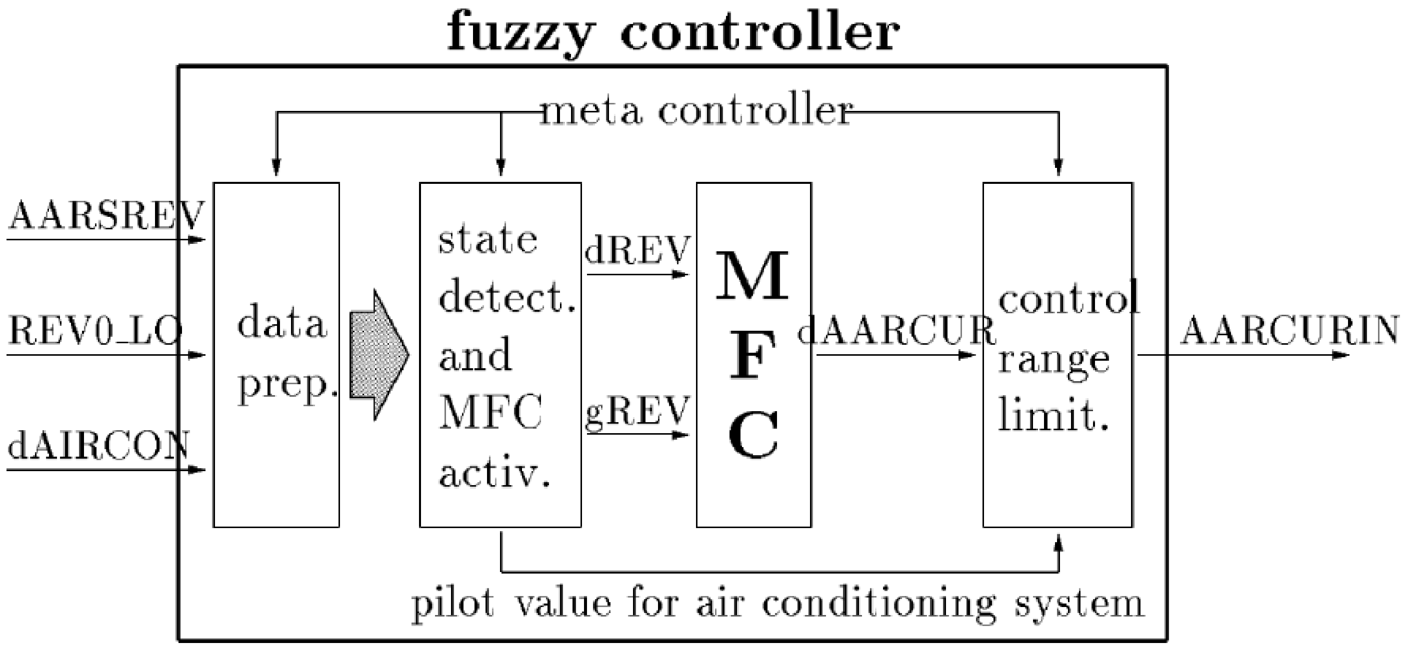
If Y is infinite, then

$$\eta = \frac{\int_{y \in Y} y \cdot \mu^{\text{output}}_{x_1,\ldots,x_n}(y)\, dy}{\int_{y \in Y} \mu^{\text{output}}_{x_1,\ldots,x_n}(y)\, dy}.$$

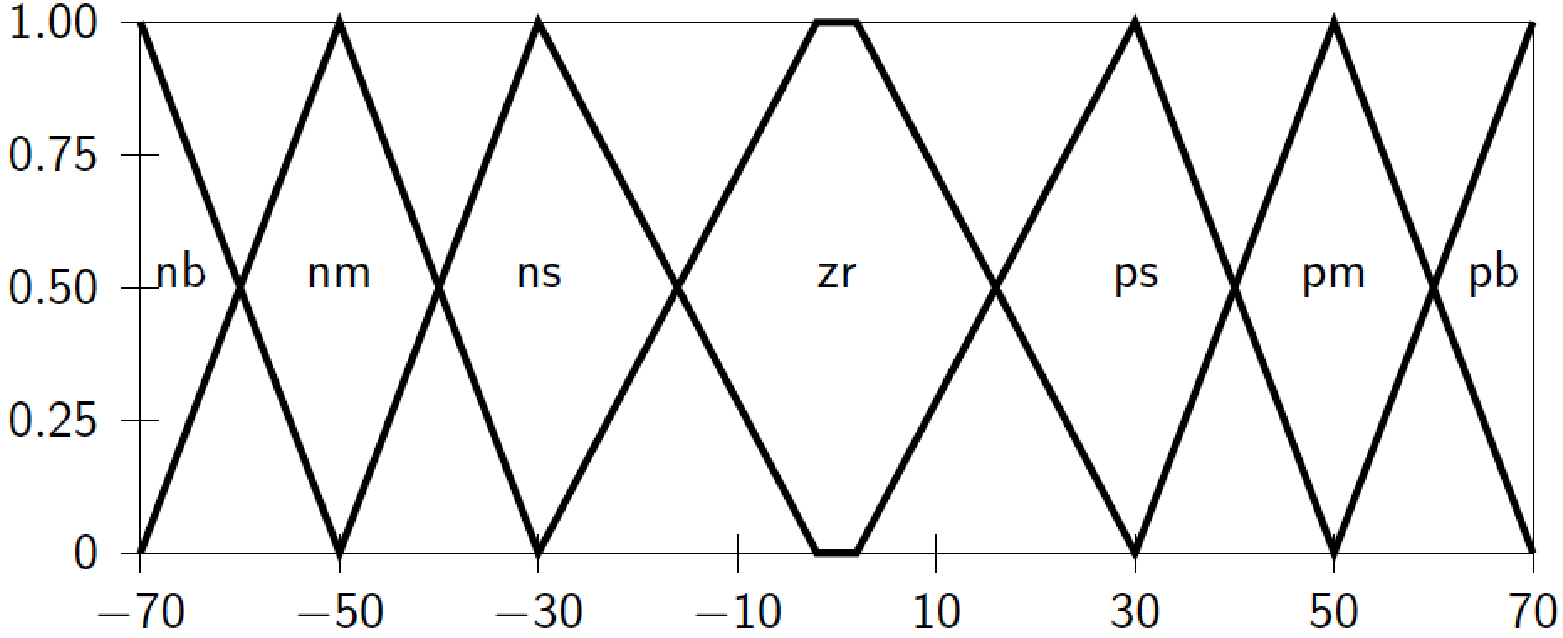
Example - Engine Idle Speed Control



Example - Engine Idle Speed Control

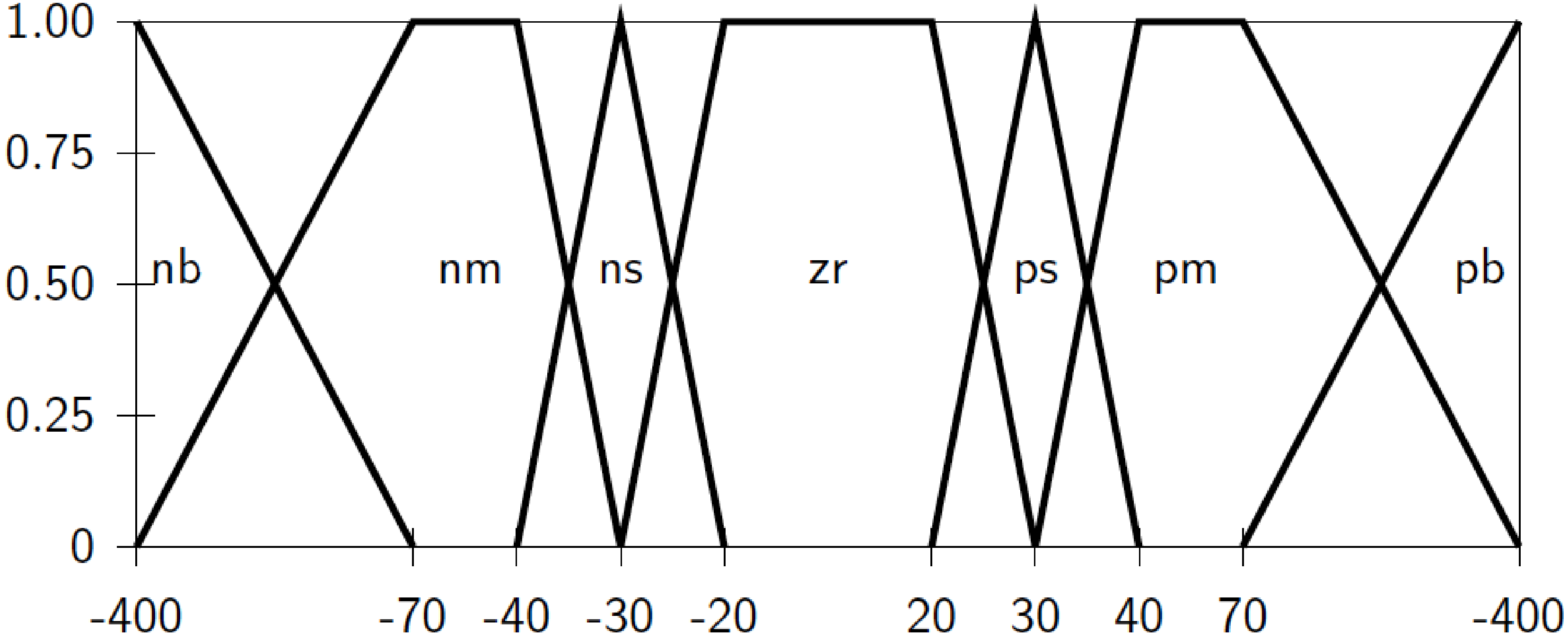


Example - Engine Idle Speed Control



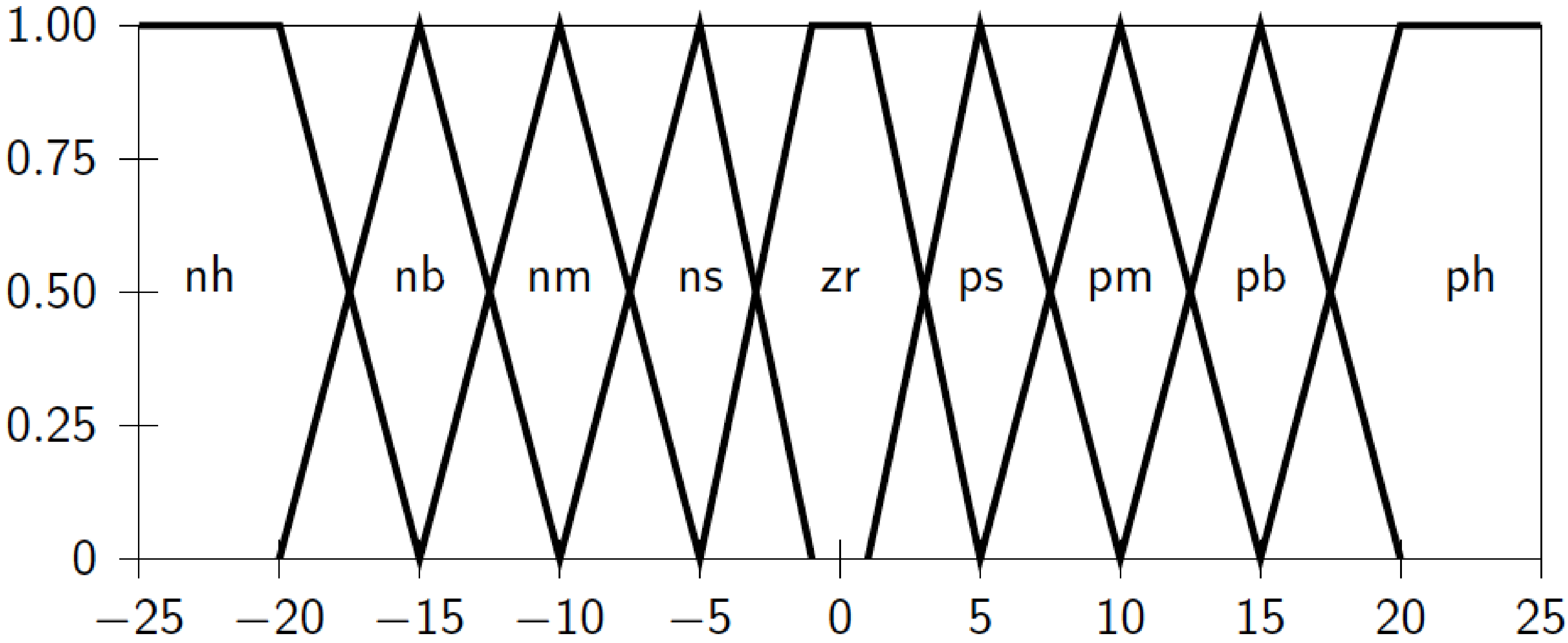
Number of revolutions - dREV

Example – Engine Idle Speed Control



Number of revolutions - gREV

Example - Engine Idle Speed Control



Change of Current for Auxiliary Air Regulator - dAARCUR

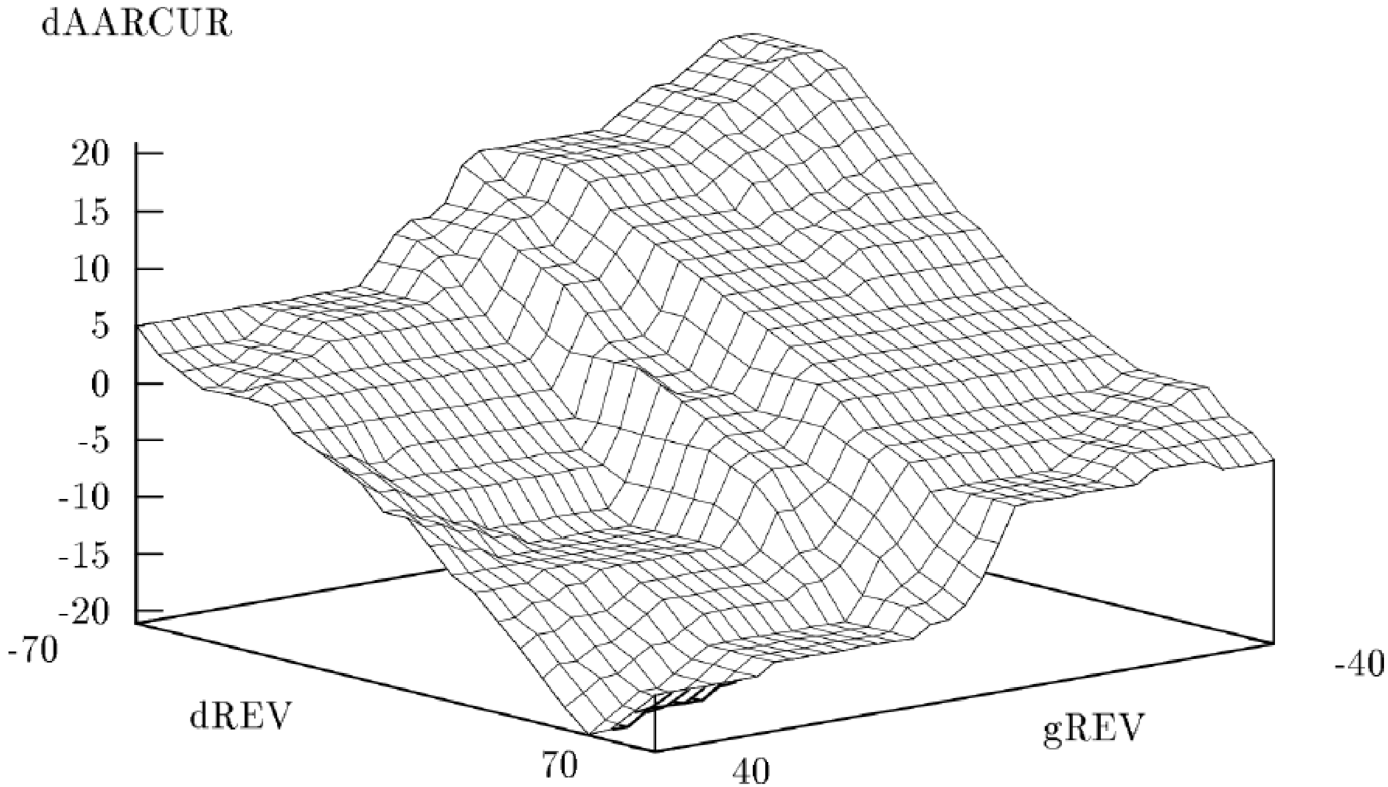
Example - Engine Idle Speed Control

		gREV						
		nb	nm	ns	az	ps	pm	pb
	nb	ph	pb	pb	pm	pm	ps	ps
	nm	ph	pb	pm	pm	ps	ps	az
	ns	pb	pm	ps	ps	az	az	az
dREV	az	ps	ps	az	az	az	nm	ns
	ps	az	az	az	ns	ns	nm	nb
	pm	az	ns	ns	ns	nb	nb	nh
	pb	ns	ns	nm	nb	nb	nb	nh

Rule Base

If the deviation from the desired number of revolutions is negative small **and** the gradient is negative medium,
then the change of the current for the auxiliary air regulation should be positive medium.

Example - Engine Idle Speed Control



Performance Characteristics

Fuzzy Control

Difference to Mamdani controller:

- no fuzzy partition of output domain Y,
- controller rules $R_1, \ldots, R_k$ are given by

$$R_r : \textbf{if } \xi_1 \text{ is } A^{(1)}_{i_{1,r}} \text{ and } \ldots \text{ and } \xi_n \text{ is } A^{(n)}_{i_{n,r}}$$
$$\textbf{then } \eta_r = f_r(\xi_1, \ldots, \xi_n),$$

$$f_r : X_1 \times \ldots \times X_n \to Y.$$

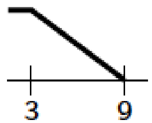
- Generally, f_r is linear, *i.e.* $f_r(x_1, \ldots, x_n) = a_0^{(r)} + \sum_{i=1}^{n} a_i^{(r)} x_i$

Takagi-Sugeno controller

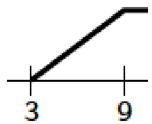
$$\eta = \frac{\sum_{r=1}^{k} \alpha_r \cdot f_r(x_1, \ldots, x_n)}{\sum_{r=1}^{k} \alpha_r}$$

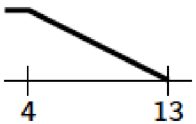
Crisp output

Example

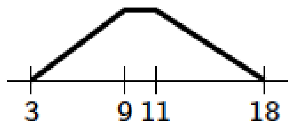
R_1 : if ξ_1 is 

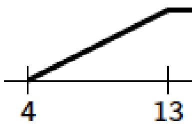
then $\eta_1 = 1 \cdot \xi_1 + 0.5 \cdot \xi_2 + 1$

R_2 : if ξ_1 is 

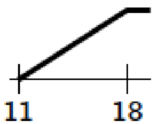
and ξ_2 is 

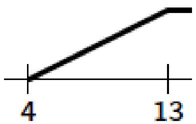
then $\eta_2 = -0.1 \cdot \xi_1 + 4 \cdot \xi_2 + 1.2$

R_3 : if ξ_1 is 

and ξ_2 is 

then $\eta_3 = 0.9 \cdot \xi_1 + 0.7 \cdot \xi_2 + 9$

R_4 : if ξ_1 is 

and ξ_2 is 

then $\eta_4 = 0.2 \cdot \xi_1 + 0.1 \cdot \xi_2 + 0.2$

Takagi-Sugeno controller

Fuzzy data analysis

- FUZZY Data Analysis
  - Fuzzy Techniques for the analysis of (crisp) data
    - Fuzzy Clustering

- FUZZY DATA Analysis
  - Analysis of Data in Form of Fuzzy Sets
  - Random Sets, Fuzzy Random Variables

Fuzzy Clustering

- Unsupervised learning task
  - goal is to divide the dataset such that both constraints hold
    - objects belonging to same cluster: as similar as possible
  - objects belonging to different clusters: as dissimilar as possible

- The similarity is measured in terms of a distance function
  - The smaller the distance, the more similar two data tuples

Distance

Distance d

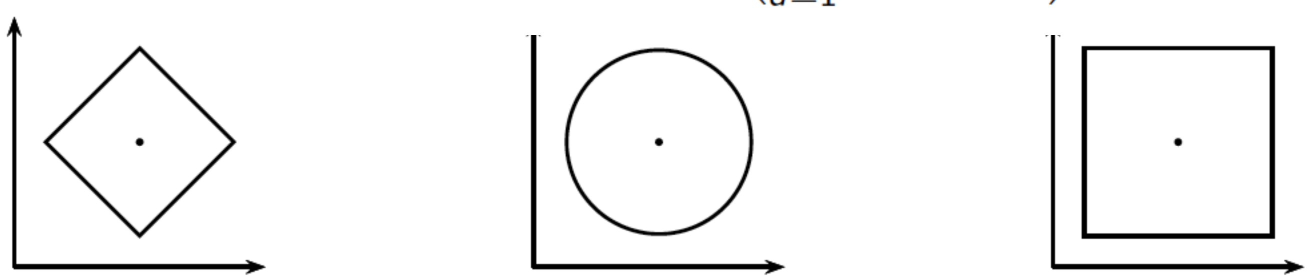
Definition

$d : \mathbb{R}^p \times \mathbb{R}^p \to [0, \infty)$ is a distance function if $\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^p$:

(i)	$d(\boldsymbol{x}, \boldsymbol{y}) = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{y}$	(identity),
(ii)	$d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$	(symmetry),
(iii)	$d(\boldsymbol{x}, \boldsymbol{z}) \leq d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{z})$	(triangle inequality).

Minkowski family

$$d_k(\boldsymbol{x}, \boldsymbol{y}) = \left(\sum_{d=1}^{p} |x_d - y_d|^k \right)^{\frac{1}{k}}$$



$k = 1$ $\quad$ $k = 2$ $\quad$ $k \to \infty$

Algorithms

- Partitioning algorithms
  - given c ∈ IN, find the best partition of data into c groups
  - usually the number of (true) clusters is unknown
  - must specify a c-value
- Hierarchical techniques
  - organize data in a nested sequence of groups
  - must specify a cut threshold

Hard c-means

- Objective function

$$d_{ij} = d(\boldsymbol{c}_i, \boldsymbol{x}_j)$$

distance

$$J_h(X, U_h, C) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij} d_{ij}^2$$

$$u_{ij} = \begin{cases} 1, & \text{if } \boldsymbol{x}_j \in \Gamma_i \\ 0, & \text{otherwise.} \end{cases}$$

Partition matrix

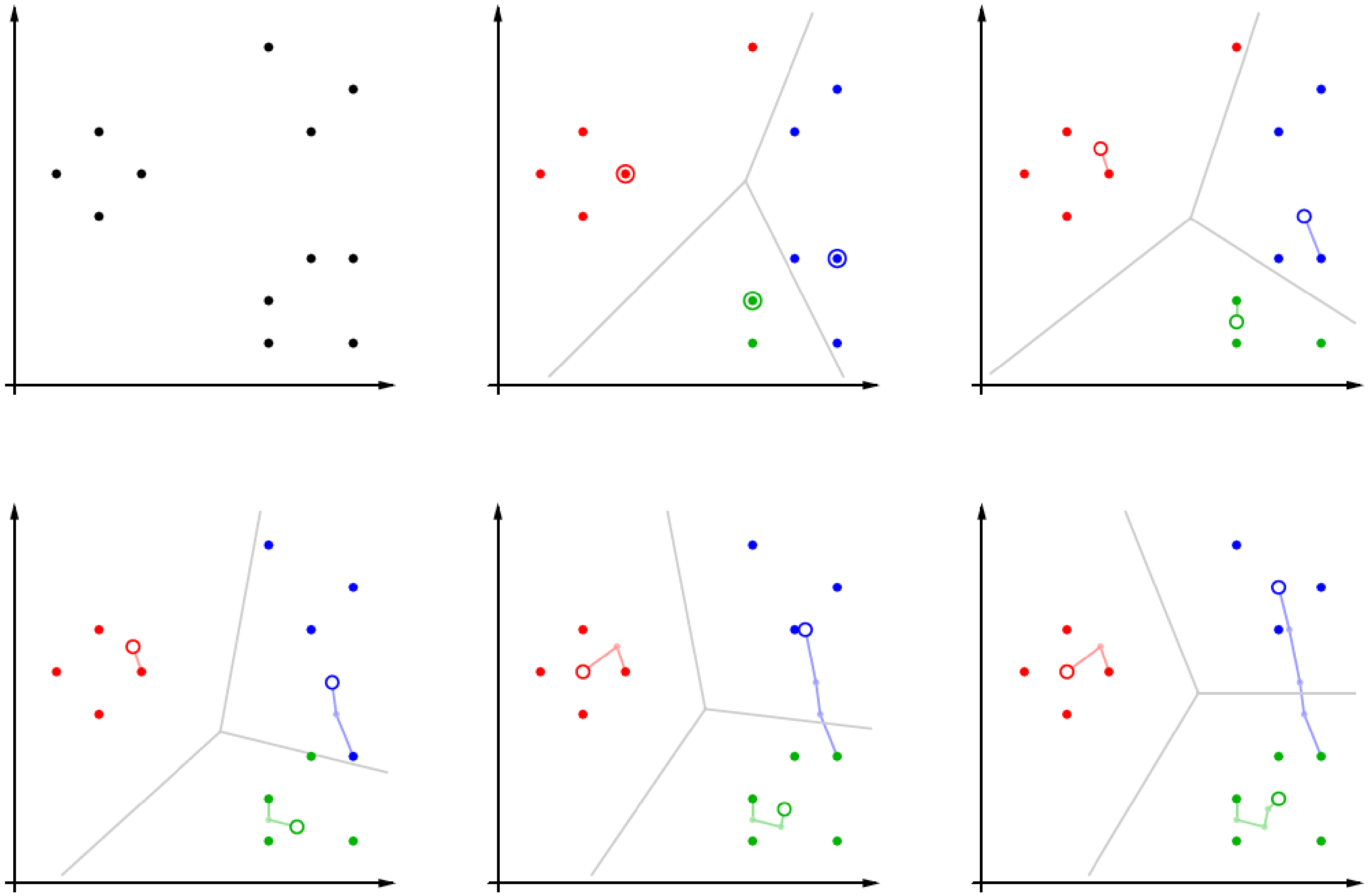
- Each data point is assegned exactly to one cluster and a every cluster must contain at least one data point

$$\forall j \in \{1, \ldots, n\} : \sum_{i=1}^{c} u_{ij} = 1 \quad \text{and} \quad \forall i \in \{1, \ldots, c\} : \sum_{j=1}^{n} u_{ij} > 0$$

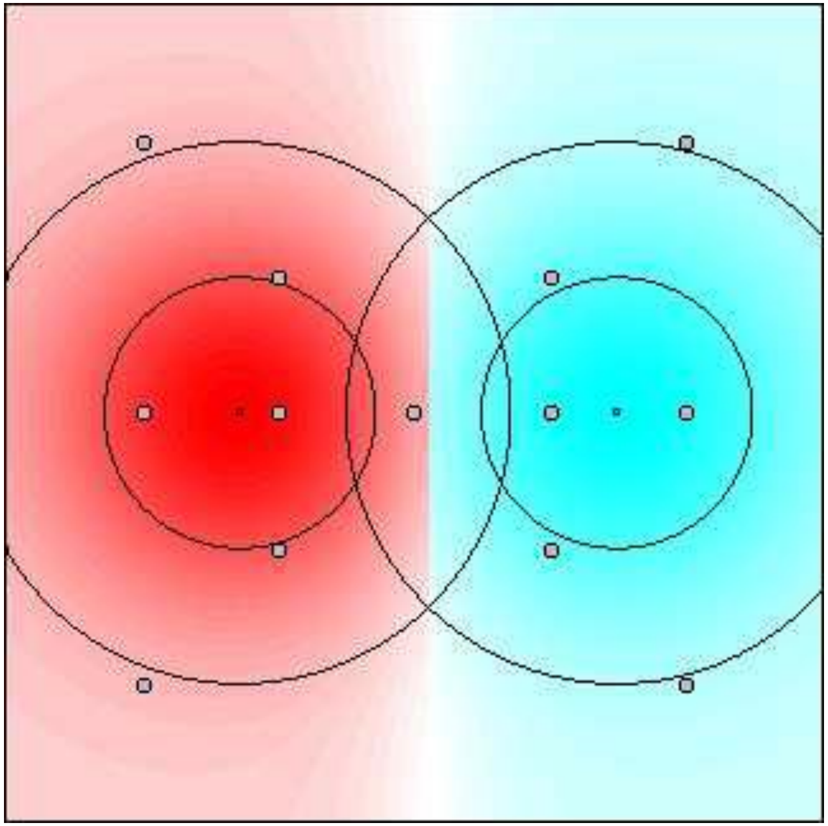
Hard c-means

- Hard c-means minimizes J_h by alternating optimization (AO)
  - The parameters to optimize are split into 2 groups
  - One group is optimized holding other one fixed (and vice versa)
  - This is an iterative update scheme: repeated until convergence

Example of c-means

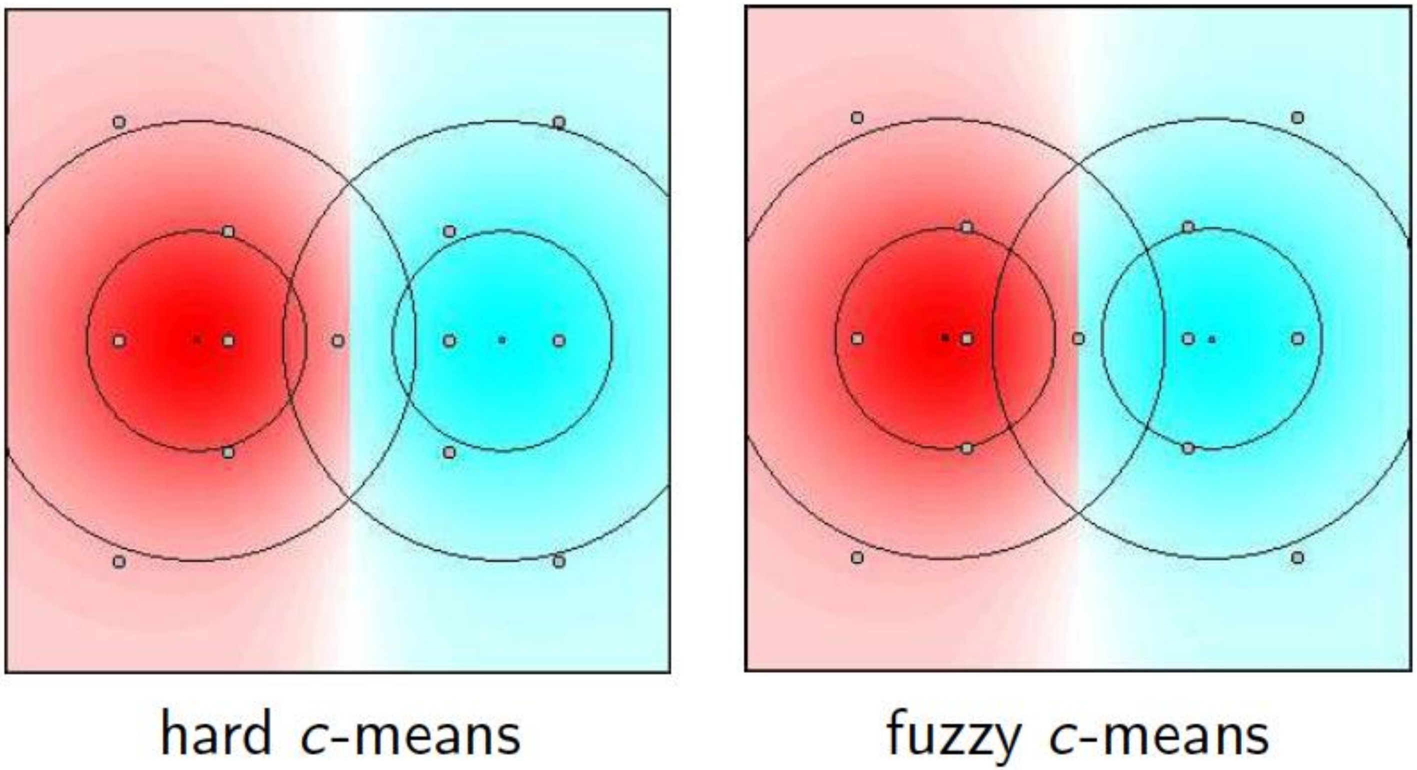


Example of c-means



Given a symmetric dataset with two clusters.
Hard c-means assigns a crisp label to the data point in the middle.

Example of c-means



In the fuzzy partition it is associated with the membership vector (0.5, 0.5)T (which expresses the ambiguity of the assignment)

Fuzzy c-means

- Objective function

$$d_{ij} = d(\mathbf{c}_i, \mathbf{x}_j)$$

distance

$$J_f(X, U_h, C) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{ij}^2$$

m > 1 is called the fuzzifier

- Subject to

$$\sum_{i=1}^{c} u_{ij} = 1, \quad \forall j \in \{1, \ldots, n\}$$

$$\sum_{j=1}^{n} u_{ij} > 0, \quad \forall i \in \{1, \ldots, c\}$$

Fuzzy Logic

Optimization

- J_f is alternately optimized
  - optimize U for a fixed cluster parameters
  - optimize C for a fixed membership degrees

Lagrange function to be minimized

$$L(X, U_f, C, \Lambda) = \underbrace{\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m d_{ij}^2}_{=J(X,U_f,C)} + \sum_{j=1}^{n} \lambda_j \left(1 - \sum_{i=1}^{c} u_{ij}\right)$$

Bezdek, 1981

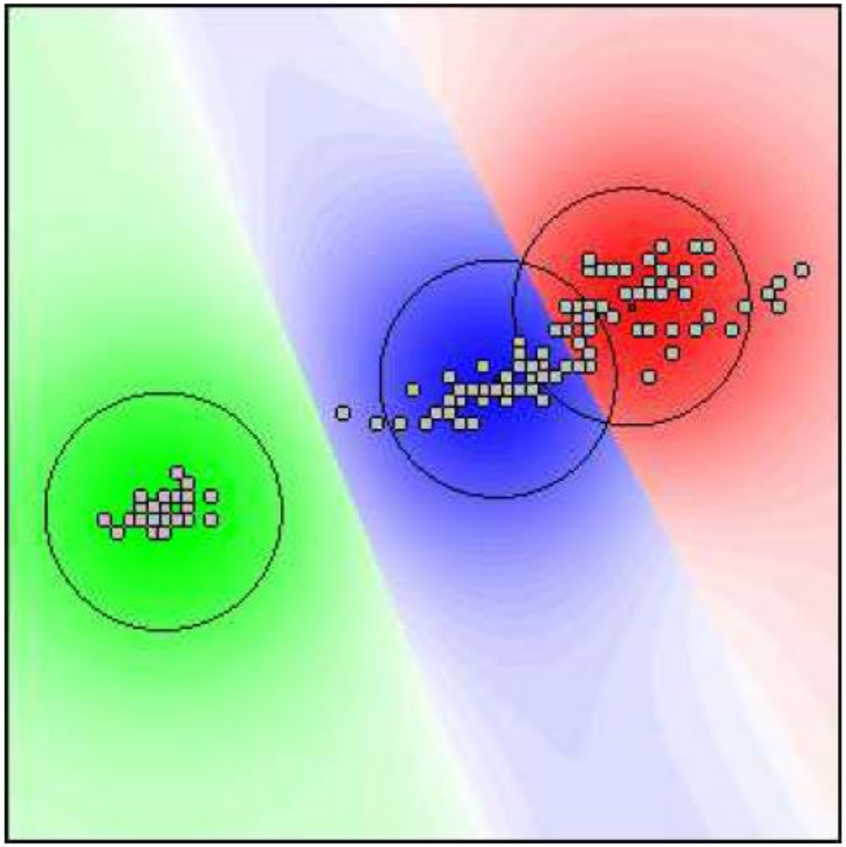
$$u_{ij} = \frac{d_{ij}^{\frac{2}{1-m}}}{\sum_{k=1}^{c} d_{kj}^{\frac{2}{1-m}}}$$

$$\boldsymbol{c}_i = \frac{\sum_{j=1}^{n} u_{ij}^m \boldsymbol{x}_j}{\sum_{j=1}^{n} u_{ij}^m}$$

First step

Second step

Example of fuzzy c-means



Fuzzy c-Means

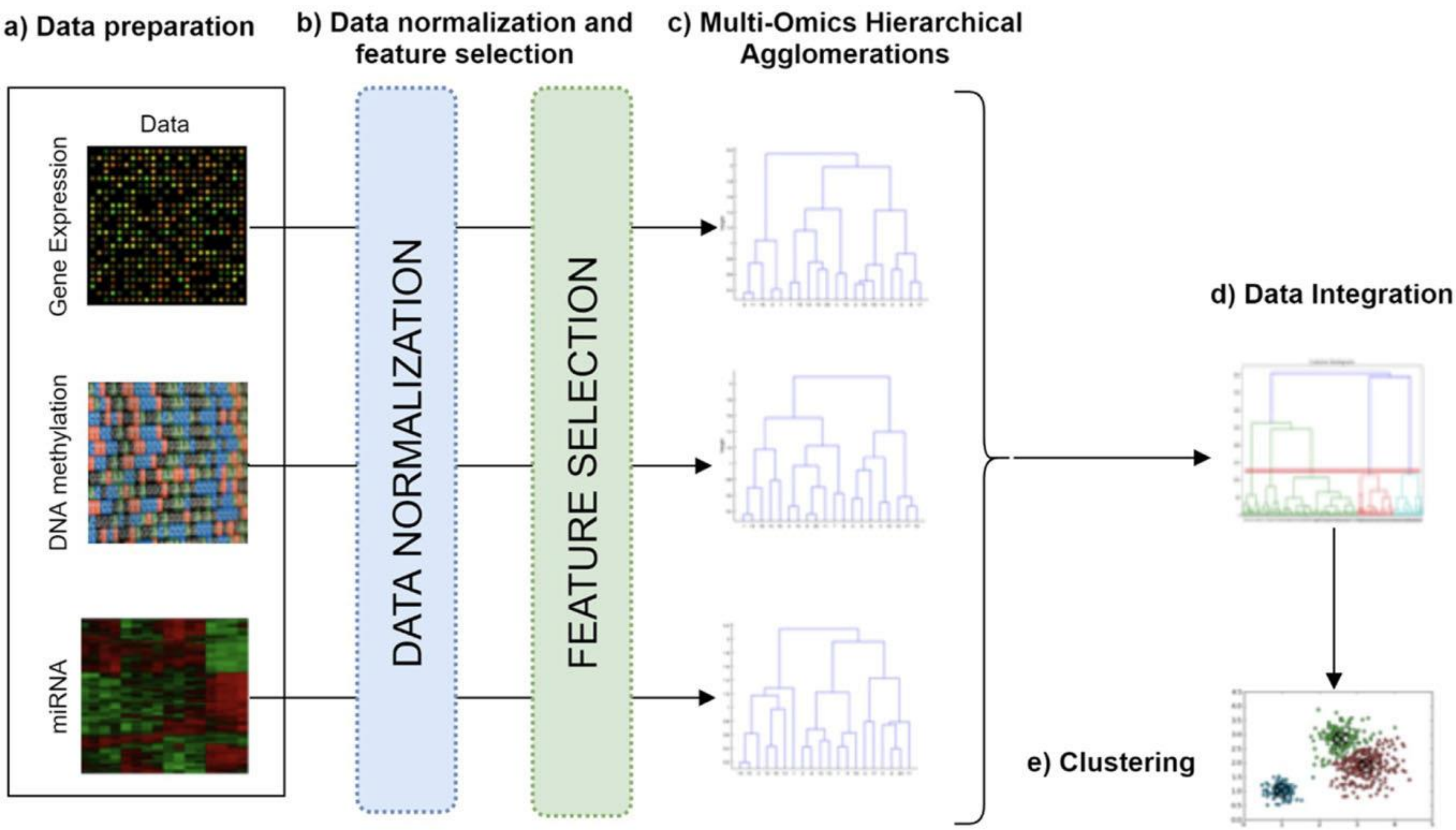
Iris data set

Data Integration

- Data integration by fuzzy similarity-based hierarchical clustering
- multi-view integration methodology for identifying patient subgroups from different omics information
  - e.g., Gene Expression, Mirna Expression, Methylation
- For each view, a dendrogram is obtained by using a hierarchical clustering based on a fuzzy equivalence relation with Łukasiewicz valued fuzzy similarity
  - https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-020-03567-6
  - Multi-Omics-Cancer-Benchmark GitHub repository

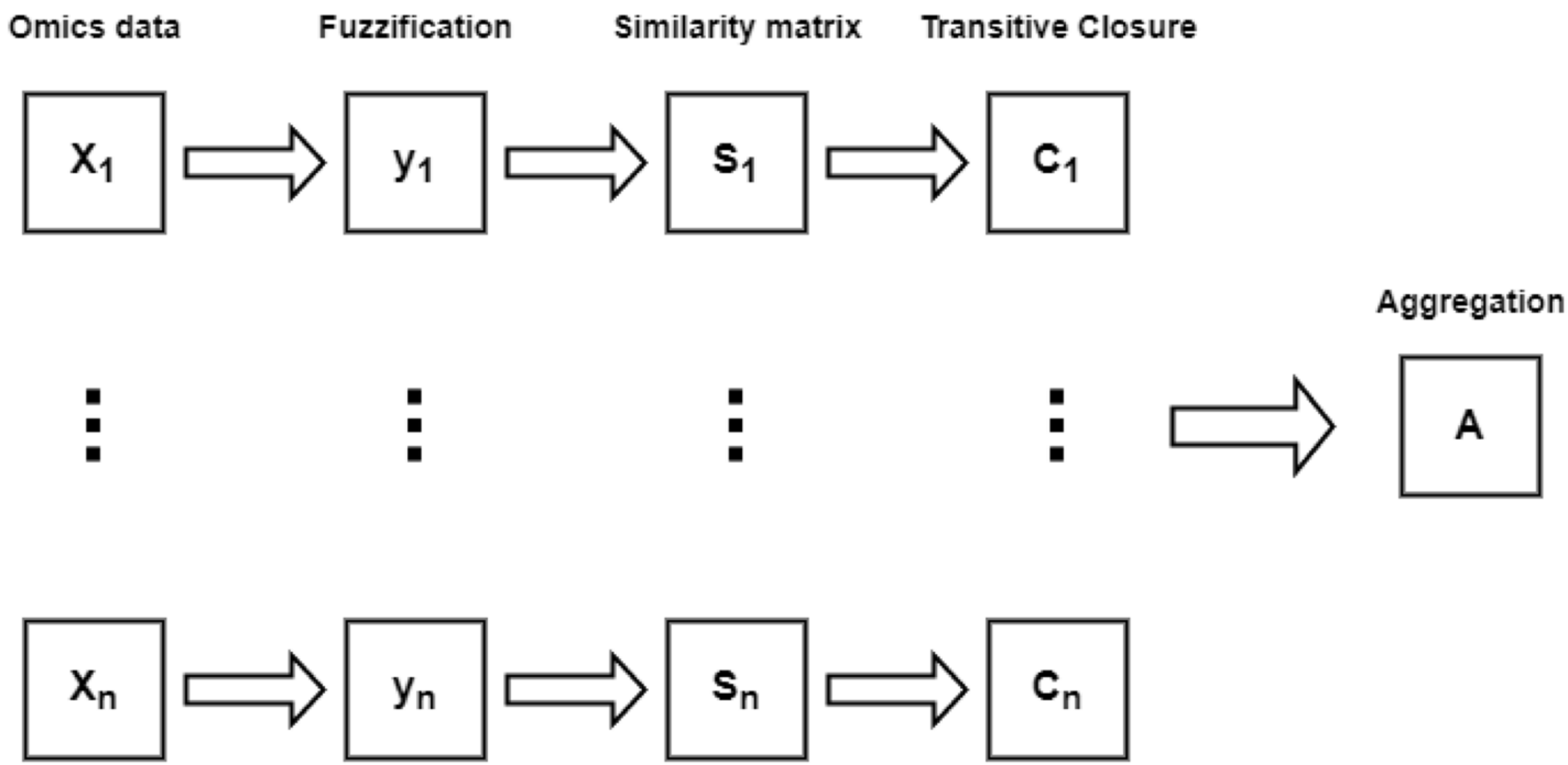
Ciaramella et al. BMC Bioinformatics 2020, 21(Suppl 10):350
https://doi.org/10.1186/s12859-020-03567-6

Data Integration



Fuzzy Logic

Data Integration



Workflow of the fuzzy based hierarchical clustering

Data Integration

Algorithm 1: Min-transitive closure

1: **Input:** relation $\mathbf{S}_i$

2: **Output:** transitive relation $\mathbf{C}_i = \mathbf{S}_i^T$

3: **Elaborate:**

1. Compute $\mathbf{S}_i^* = \mathbf{S}_i \cup (\mathbf{S}_i \circ \mathbf{S}_i)$

2. if $\mathbf{S}_i^* \neq \mathbf{S}_i$ replace $\mathbf{S}_i$ with $\mathbf{S}_i^*$ and go to step 1

else $\mathbf{C}_i = \mathbf{S}_i^T = \mathbf{S}_i^*$ and the algorithm terminates.

Algorithm 2: Combination of dendrograms

1: **Input** $\mathbf{C}_i$, $1 \leq i \leq L$ L input similarity matrices (dendrograms)

2: **Output** similarity matrix (dendrogram) **A**

1. Aggregate the similarity matrices to a final similarity matrix

$\mathbf{A} = \mathit{Aggregate}\,(\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_L)$

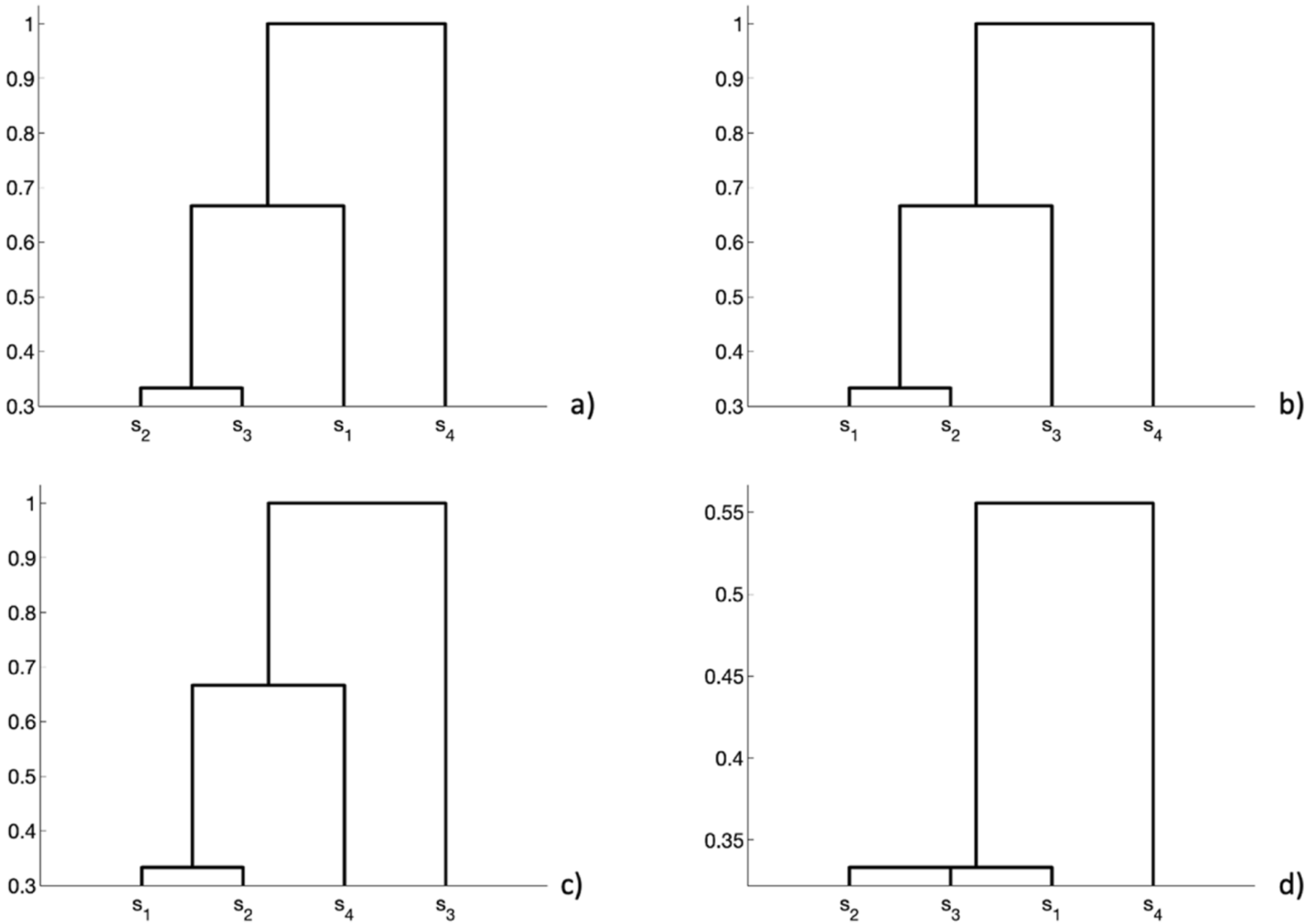
a. Let $\mathbf{A}^*$ be the identity matrix

b. For each $\mathbf{C}_i$ calculate e $\mathbf{A}^* = \mathbf{A}^* \cup (\mathbf{A}^* \circ \mathbf{C}_i)$

c. If $\mathbf{A}^*$ is not changed $\mathbf{A} = \mathbf{A}^*$ and goto step 3 else goto step 1.b

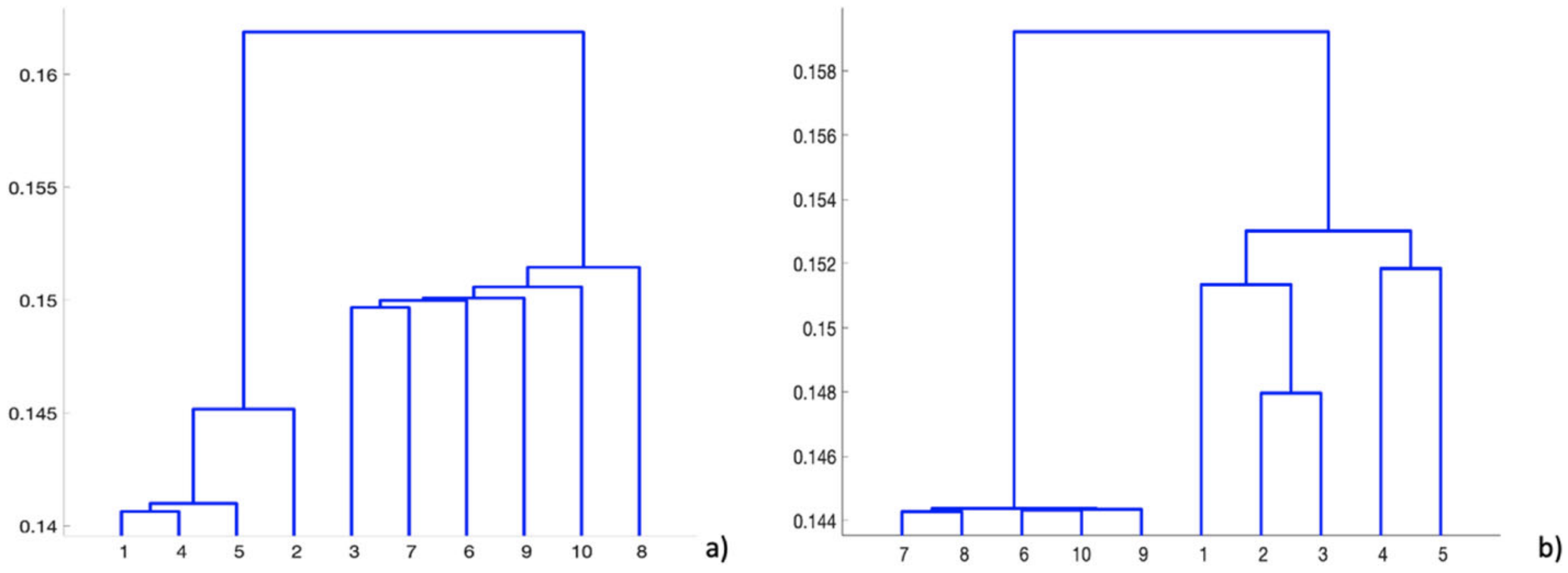
3: Create the final dendrogram from **A**

Data Integration



Data Integration approach

Data Integration



Crisp Hierarchical Clustering vs Fuzzy based Hierarchical Clustering

Data Integration

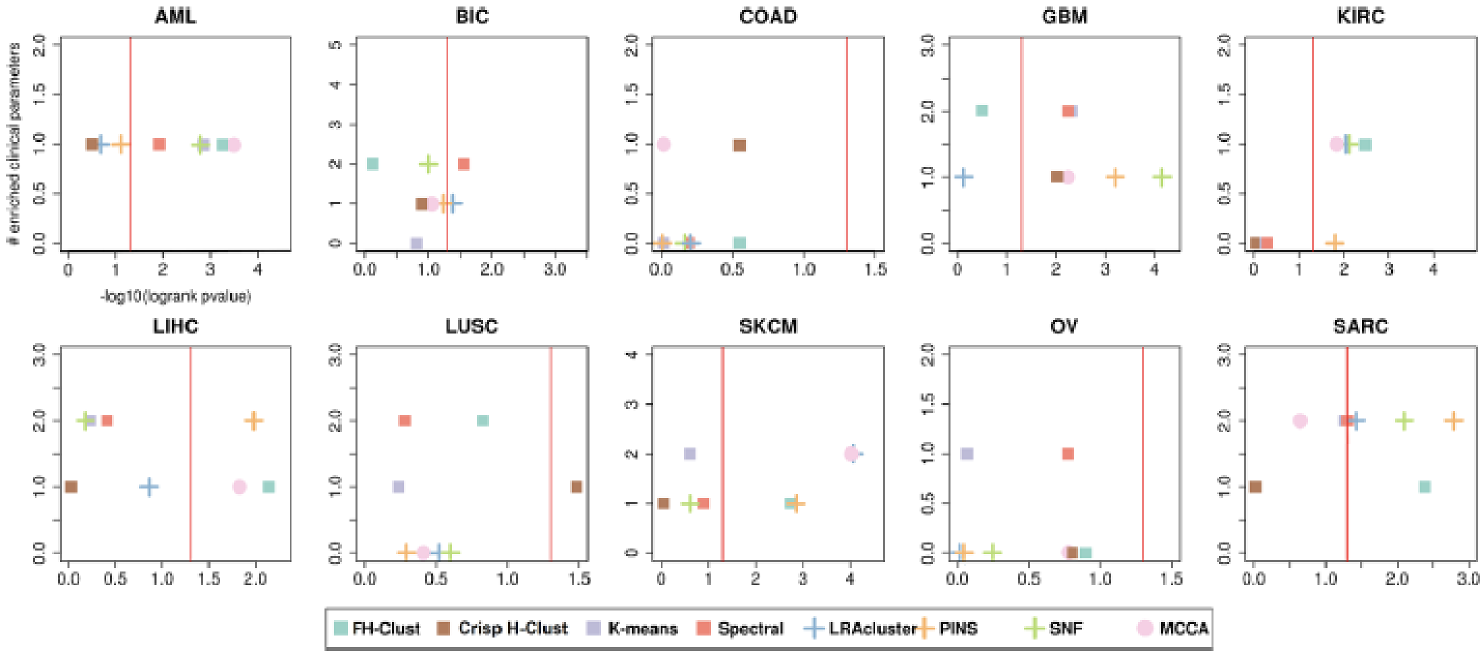


Fig. 7 Performance of the algorithms on ten multi-omics cancer datasets. For each plot, the x-axis measures the differential survival between clusters(-log10 of logrank's test *P*-value), and the y-axis is the number of clinical parameters enriched in the clusters. Red vertical lines indicate the threshold for significantly different survival (*P*-value $\leqslant 0.05$)

Data Integration

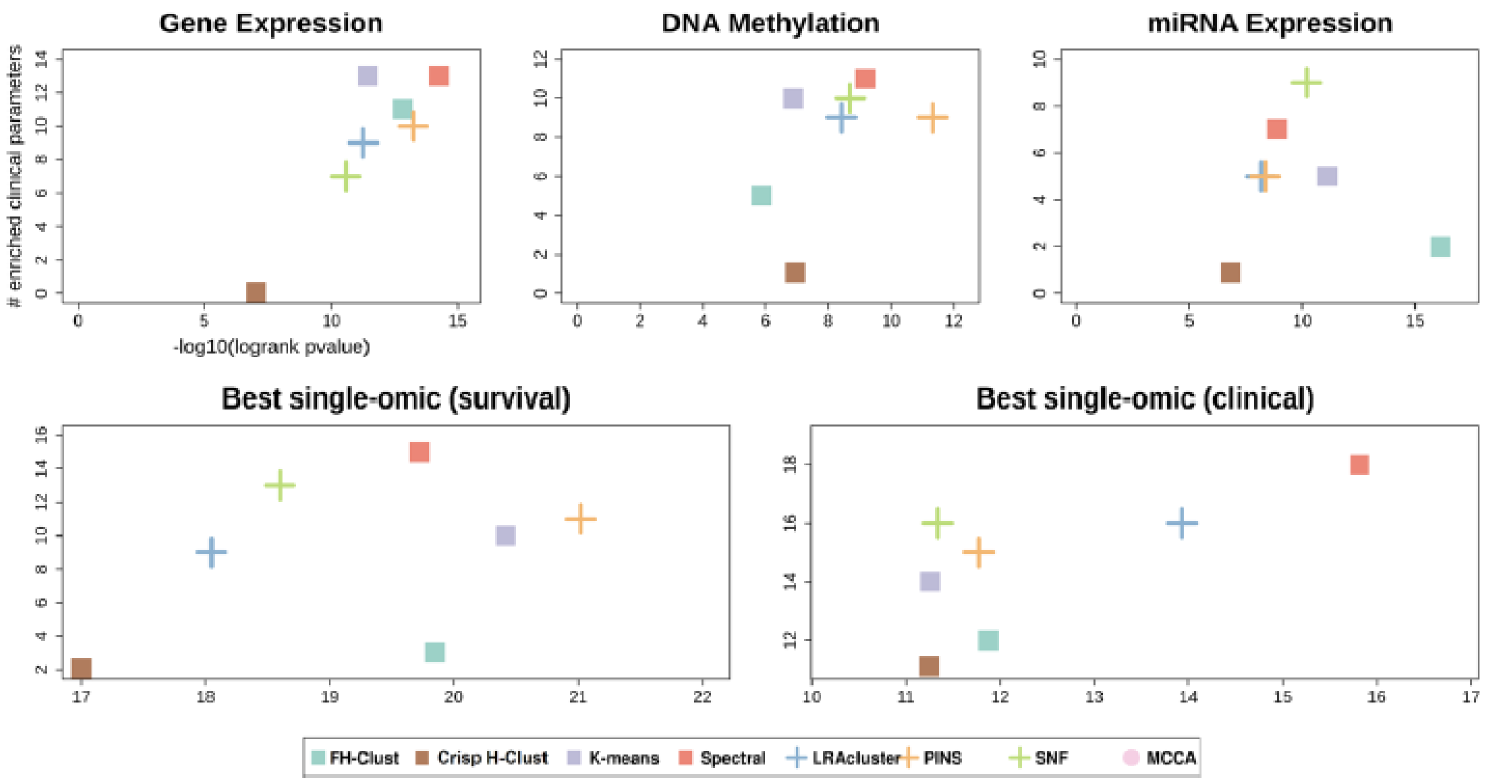
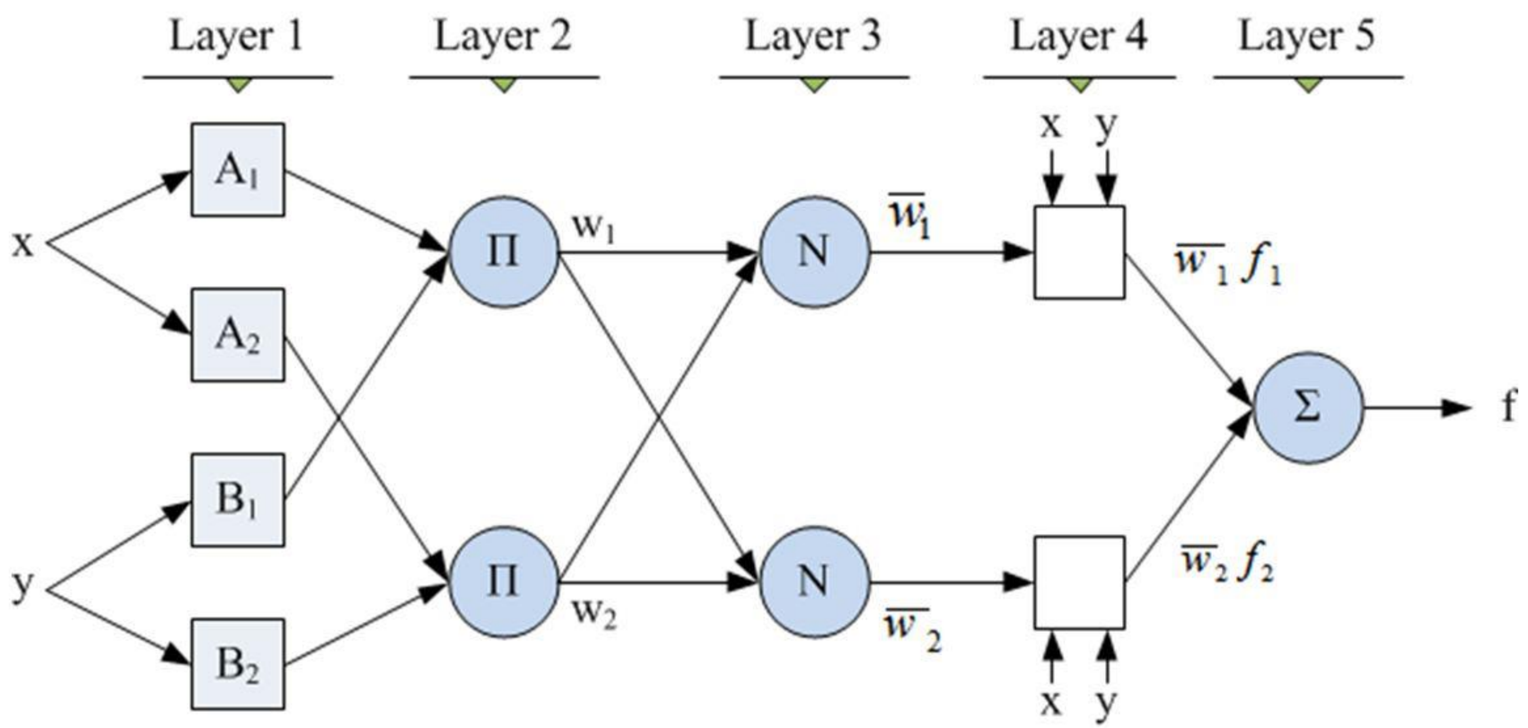


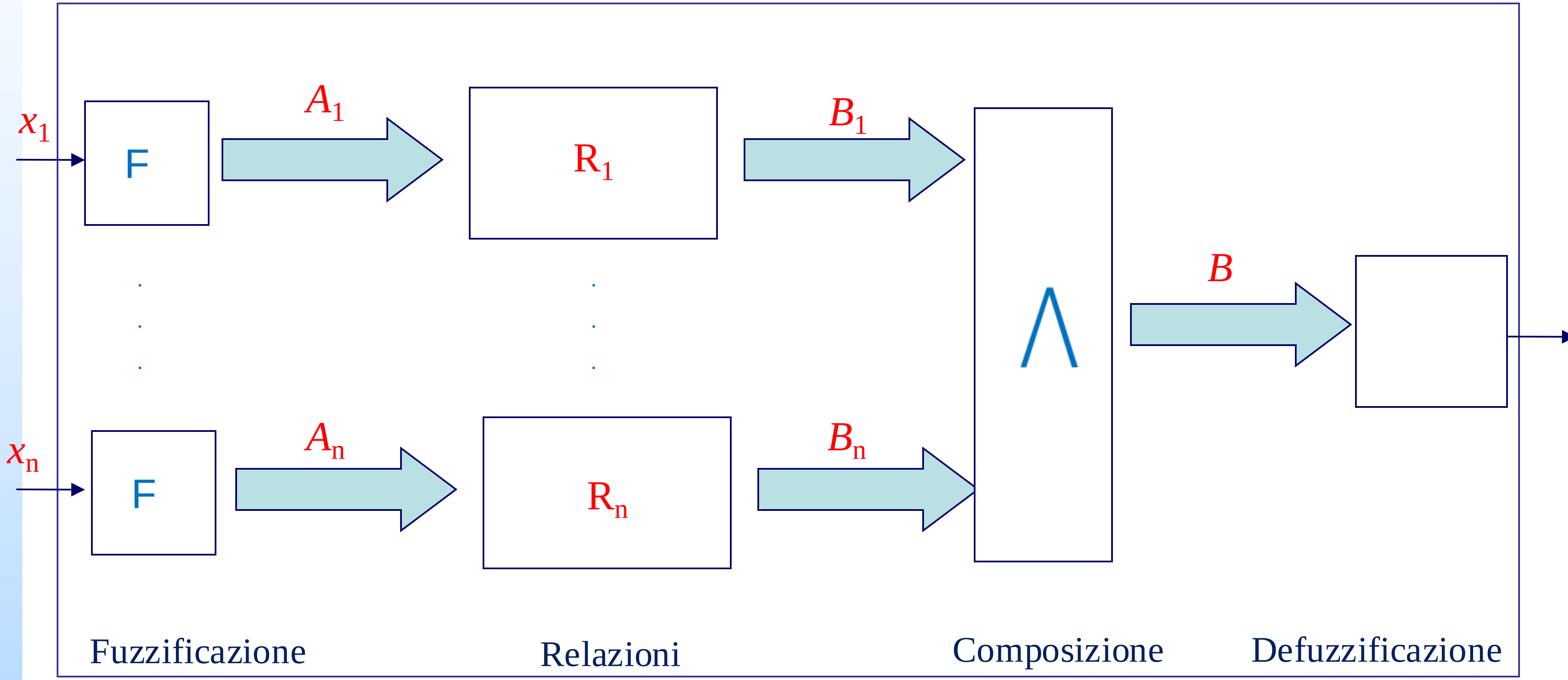
Fig. 9 Summarized performance of the algorithms across ten cancer datasets. For each plot, the x-axis measures the total differential prognosis between clusters (sum across all datasets of –log10 of logrank's test *P*-value), and the y-axis is the total number of clinical parameters enriched in the clusters across all cancer types. (**a**–**c**) Results for single-omic datasets. **d** Results when each method uses the single omic that achieves the highest significance in survival. **e** Same with respect to enrichment of clinical labels

ANFIS



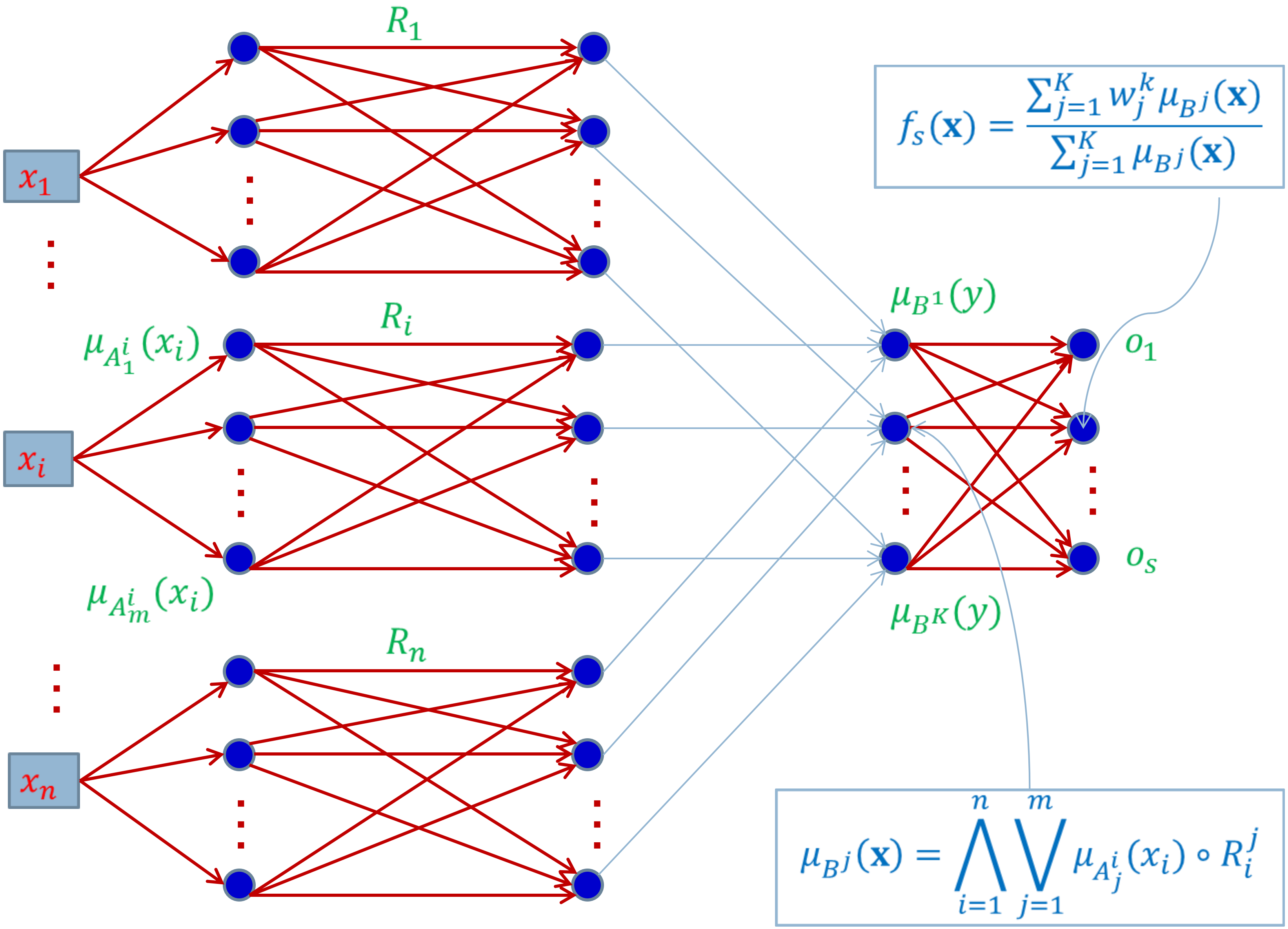
Adaptive neuro fuzzy inference system

FRNN



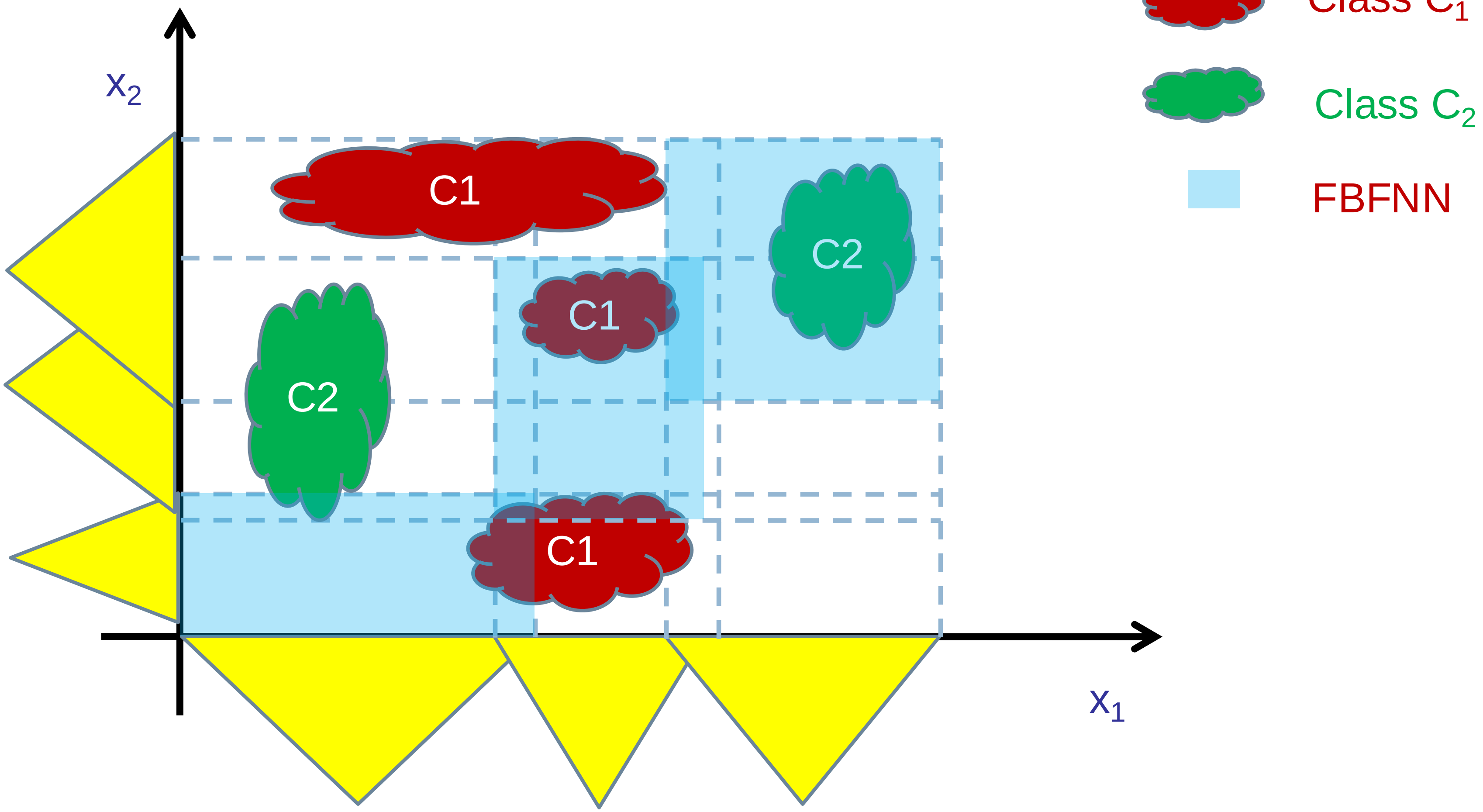
Fuzzy Relation Neural Network Model

FRNN

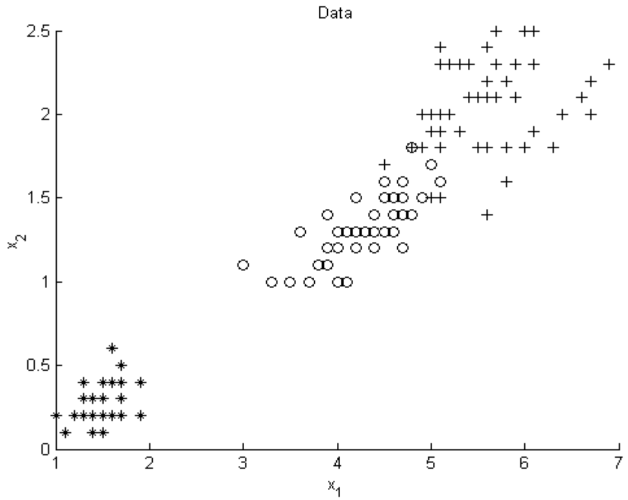


$f_s(\mathbf{x}) = \frac{\sum_{j=1}^{K} w_j^k \mu_{B^j}(\mathbf{x})}{\sum_{j=1}^{K} \mu_{B^j}(\mathbf{x})}$
R_1
x_1
R_i
$\mu_{A_1^i}(x_i)$
x_i
$\mu_{A_m^i}(x_i)$
R_n
x_n
$\mu_{B^1}(y)$
o_1
o_s
$\mu_{B^K}(y)$
$\mu_{B^j}(\mathbf{x}) = \bigwedge_{i=1}^{n} \bigvee_{j=1}^{m} \mu_{A_j^i}(x_i) \circ R_i^j$

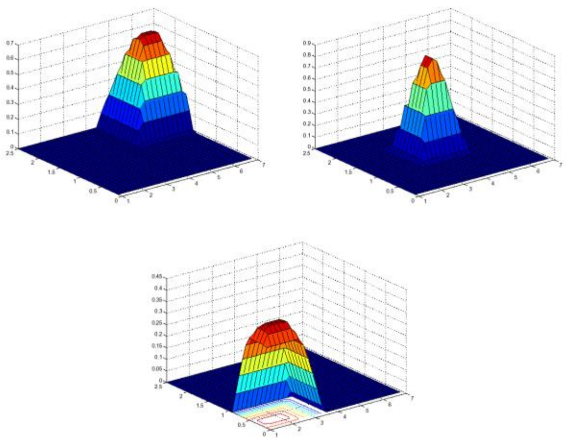
Granulation



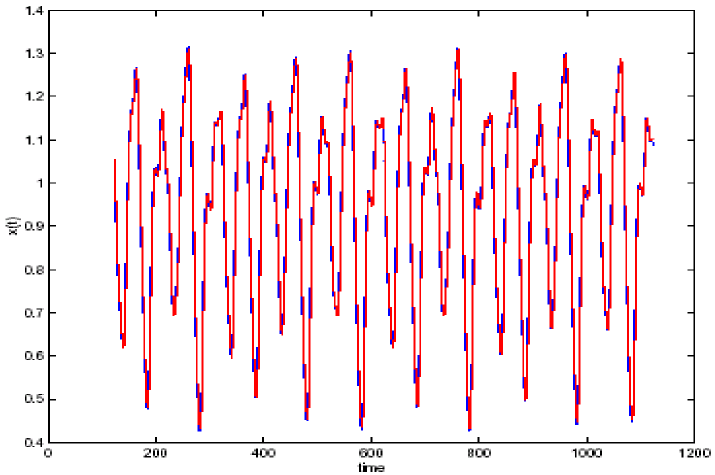
Some results



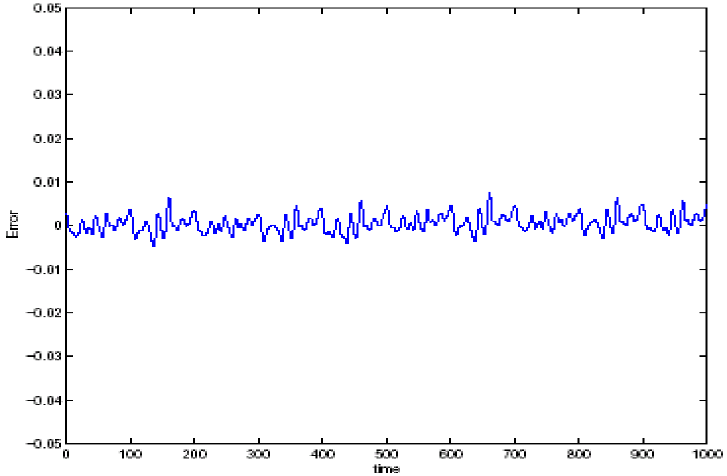
IRIS data set



Memberships

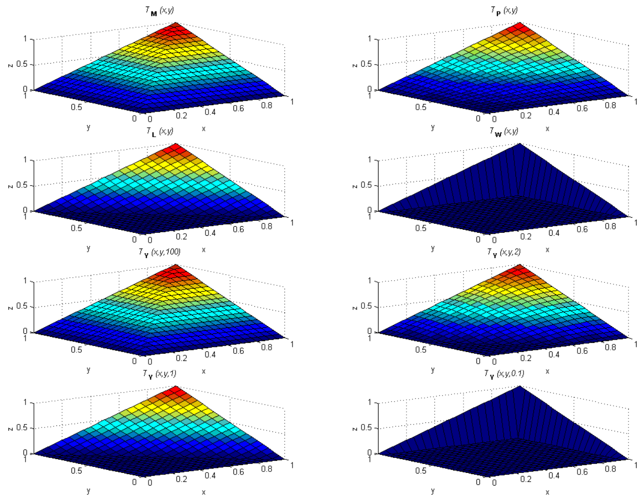


Mackey-Glass chaotic time series



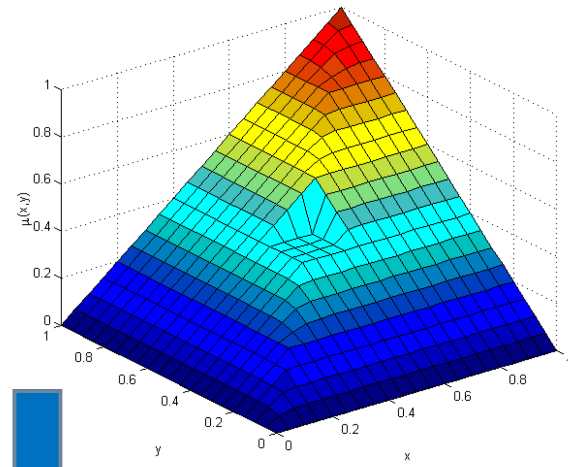
Residum

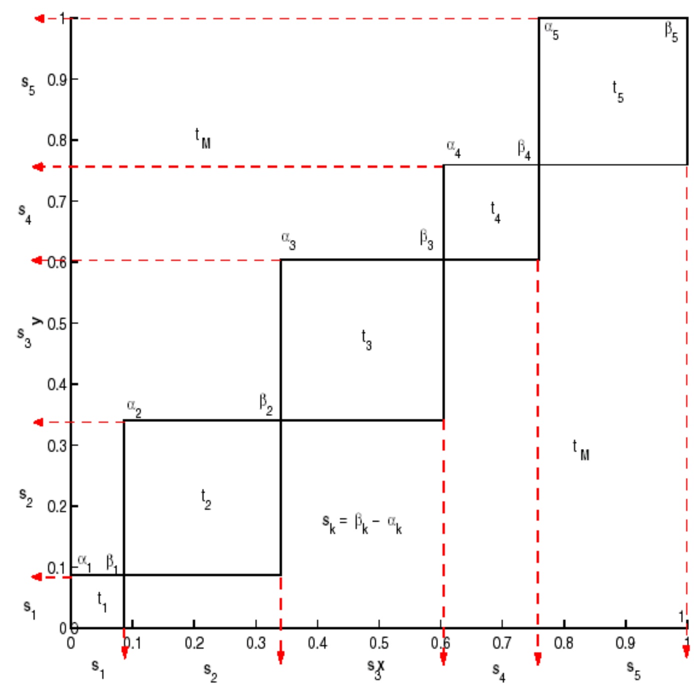
Norm generalization



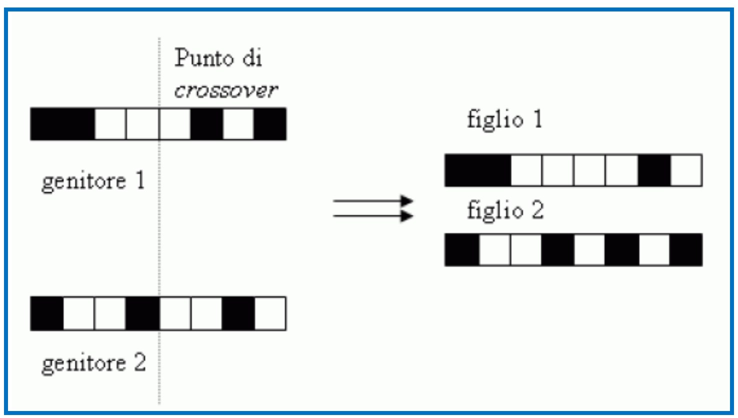
t-norms and t-conorms

Ordinal sums



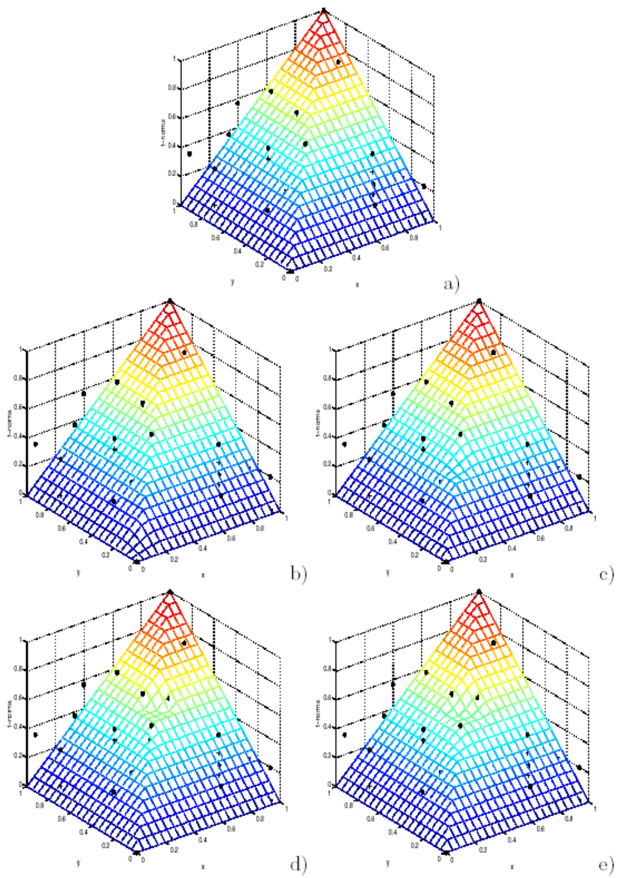


Parameters of Ordinal Sums

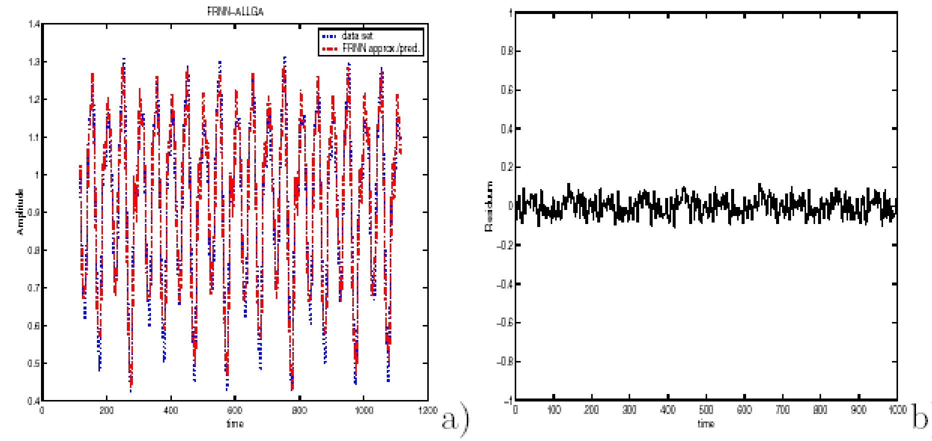


Chromosome

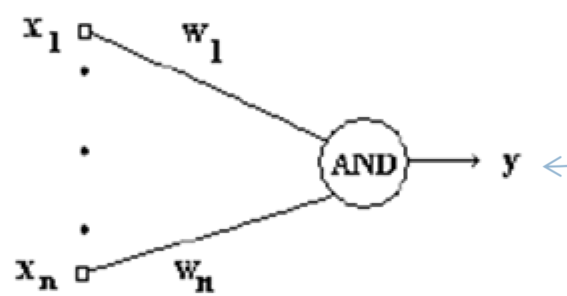
Neuron generalization



Zimmermann and Zysno data set



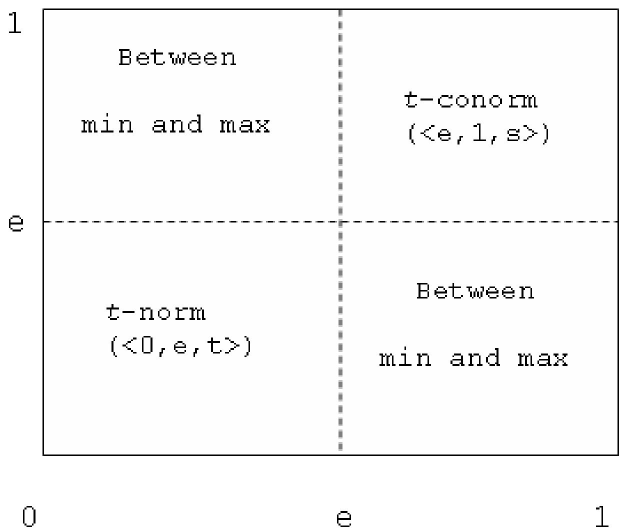
FRNN inference system



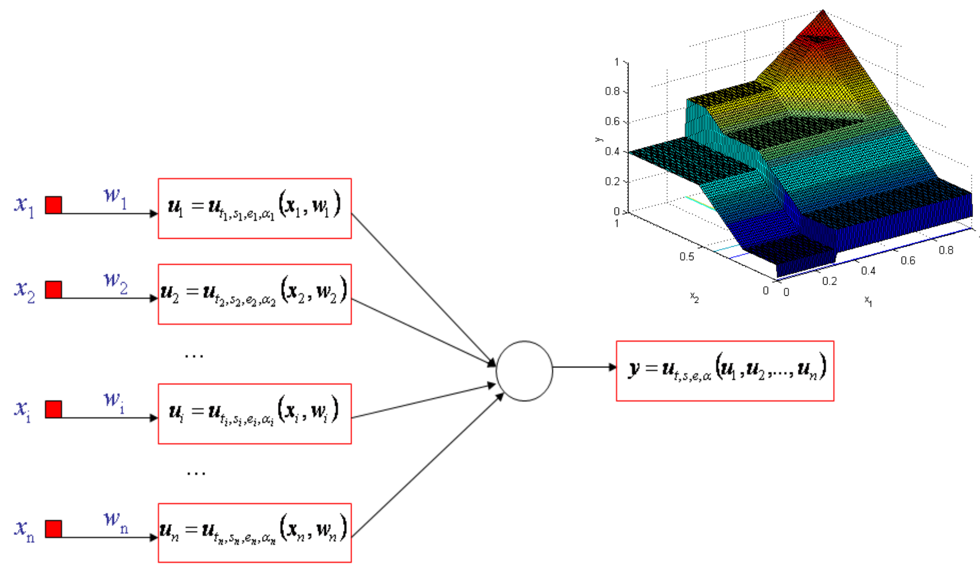
$$y = \bigwedge_{i=1}^{n} w_i \vee x_i$$

AND/OR neuron based on OS

Uninorm



Uninorm representation



Uninorm based neuron

Structured data