Advanced Fault Detection in Condition Monitoring: Combining Model-Based and Data-Driven Approaches
Part 5

Motivation for AI Models

Intelligent (AI & ML) Models

techniques motivated by human intelligence

- fuzzy systems (represent human knowledge, reasoning)
- artificial neural networks (adaptation, learning)
- genetic algorithms (optimization)
- particle swarm optimization
- etc.

Intelligent Design Tools

techniques motivated by human intelligence

- fuzzy systems (represent human knowledge, reasoning)
- artificial neural networks (adaptation, learning)
- genetic algorithms (optimization)
- particle swarm optimization
- etc.

- Fuzzy knowledge-based control
- Fuzzy data analysis, modeling, identification
- Learning and adaptive control (neural networks)
- Reinforcement learning

Fuzzy Models I

Outline

1. Fuzzy sets and set-theoretic operations
2. Fuzzy relations
3. Fuzzy systems
4. Linguistic model, approximate reasoning

Fuzzy Sets and Fuzzy Logic

Relatively new methods for representing uncertainty and reasoning under uncertainty.

Types of uncertainty:

- chance, randomness (stochastic)
- imprecision, vagueness, ambiguity (non-stochastic)

Classical Set Theory

A set is a collection of objects with a common property.

Classical Set Theory

A set is a collection of objects with a common property.

Examples:

- Set of natural numbers smaller than 5: $A = \{1,\ 2,\ 3,\ 4\}$

Classical Set Theory

A set is a collection of objects with a common property.

Examples:

- Set of natural numbers smaller than 5: $A = \{1,\ 2,\ 3,\ 4\}$
- Unit disk in the complex plane: $A = \{z | z \in \mathbb{C}, |z| \leq 1\}$

Classical Set Theory

A set is a collection of objects with a common property.

Examples:

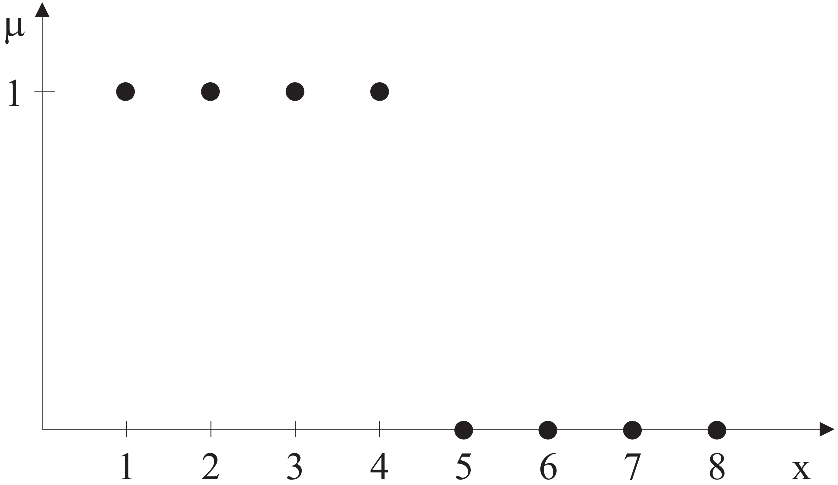
- Set of natural numbers smaller than 5: $A = \{1,\ 2,\ 3,\ 4\}$
- Unit disk in the complex plane: $A = \{z | z \in \mathbb{C}, |z| \leq 1\}$
- A line in $\mathbb{R}^2$: $A = \{(x, y) | ax + by + c = 0, (x, y, a, b, c) \in \mathbb{R}\}$

Representation of Sets

- Enumeration of elements: $A = \{x_1,\ x_2, \ldots,\ x_n\}$
- Definition by property: $A = \{x \in X | x \text{ has property } P\}$
- Characteristic function: $\mu_A(x) : X \to \{0, 1\}$

$$\mu_A(x) = \begin{cases} 1 & x \text{ is member of } A \\ 0 & x \text{ is not member of } A \end{cases}$$

Set of natural numbers smaller than 5



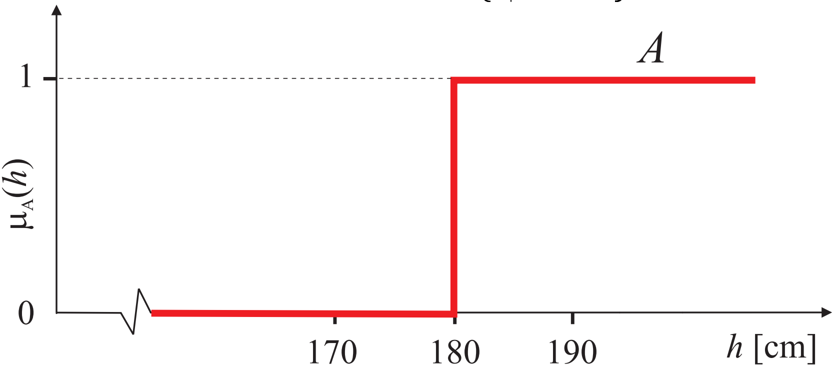
Fuzzy sets

Why Fuzzy Sets?

- Classical sets are good for well-defined concepts (maths, programs, etc.)
- Less suitable for representing commonsense knowledge in terms of vague concepts such as:
  - a tall person, slippery road, nice weather, . . .
  - want to buy a big car with moderate consumption
  - If the temperature is too low, increase heating a lot

Classical Set Approach

set of tall people $A = \{h | h \geq 180\}$

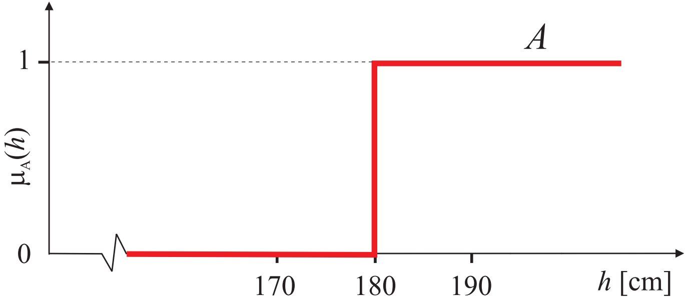


Logical Propositions

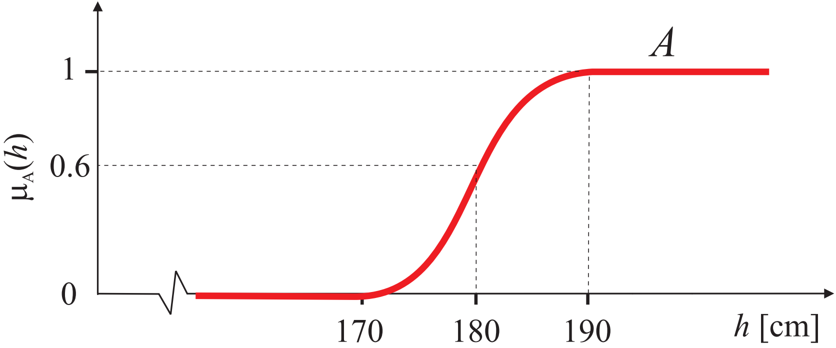
"John is tall" . . . true or false

John's height: $h_{John} = 180.0$ $\mu_A(180.0) = 1$ (true)

$h_{John} = 179.5$ $\mu_A(179.5) = 0$ (false)



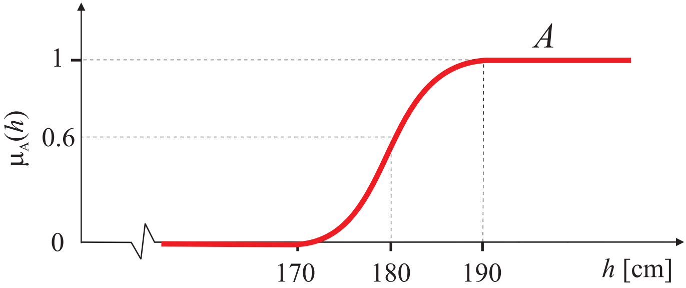
Fuzzy Set Approach



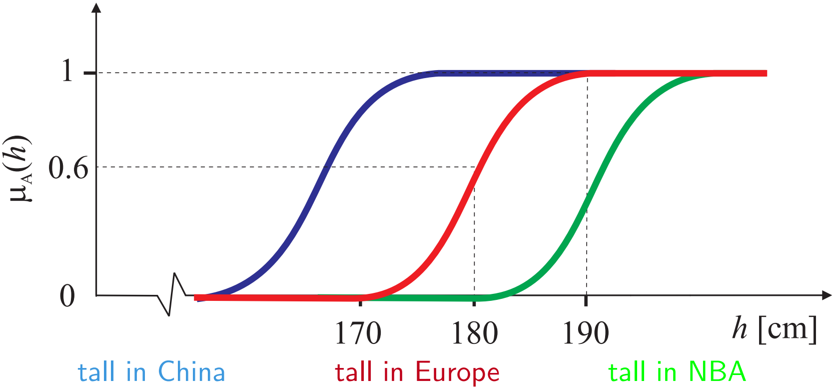
$$\mu_A(h) = \begin{cases} 1 & h \text{ is full member of } A \quad (h \geq 190) \\ (0,1) & h \text{ is partial member of } A \quad (170 < h < 190) \\ 0 & h \text{ is not member of } A \quad (h \leq 170) \end{cases}$$

Fuzzy Logic Propositions

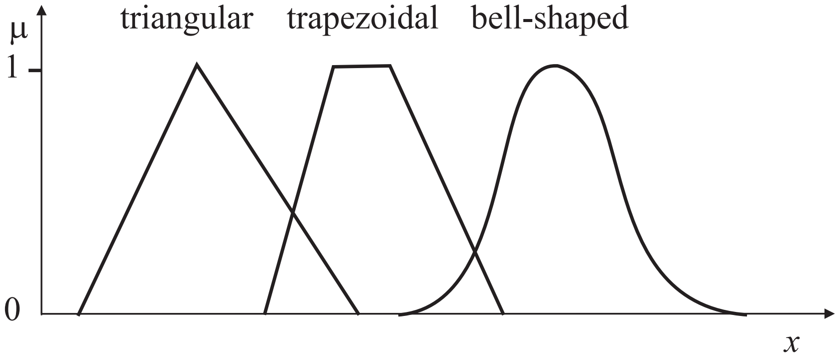
"John is tall"	... degree of truth	
John's height:	$h_{John} = 180.0$	$\mu_A(180.0) = 0.6$
	$h_{John} = 179.5$	$\mu_A(179.5) = 0.56$
	$h_{Paul} = 201.0$	$\mu_A(201.0) = 1$



Subjective and Context Dependent



Shapes of Membership Functions



Representation of Fuzzy Sets

- Pointwise as a list of membership/element pairs:

$$A = \{\mu_A(x_1)/x_1, \ldots, \mu_A(x_n)/x_n\} = \{\mu_A(x_i)/x_i | x_i \in X\}$$

- As a list of α-level/α-cut pairs:

$$A = \{\alpha_1/A_{\alpha_1},\ \alpha_2/A_{\alpha_2},\ \ldots, \alpha_n, A_{\alpha_n}\} = \{\alpha_i/A_{\alpha_i} | \alpha_i \in (0, 1)\}$$

Representation of Fuzzy Sets

- Analytical formula for the membership function:

$$\mu_A(x) = \frac{1}{1+x^2}, \quad x \in \mathbb{R}$$

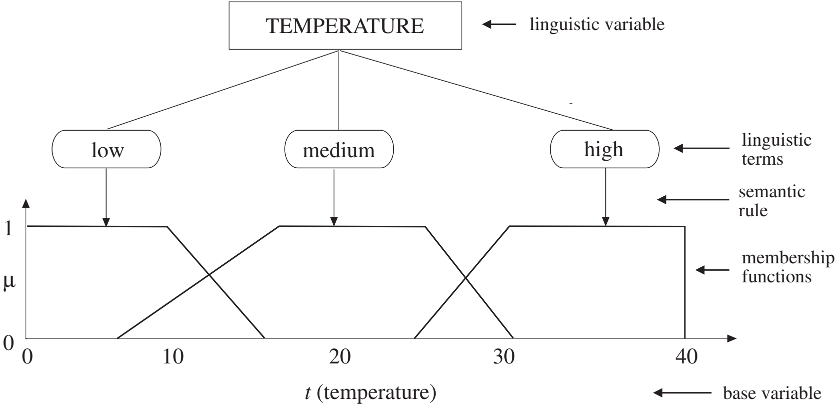
or more generally

$$\mu(x) = \frac{1}{1+d(x,v)} .$$

$d(x,v)$ ...dissimilarity measure

Various shorthand notations: $\mu_A(x) \dots A(x) \dots a$

Linguistic Variable

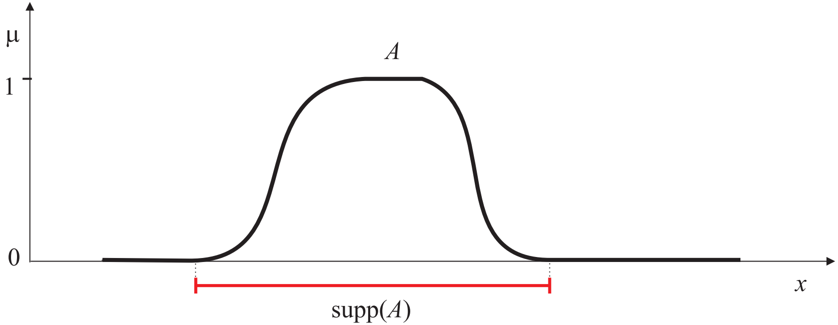


Basic requirements: coverage and semantic soundness

Properties of fuzzy sets

Support of a Fuzzy Set

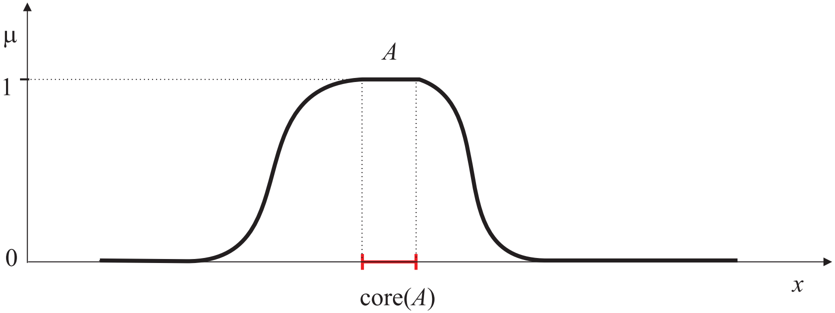
$$\mathrm{supp}(A) = \{x | \mu_A(x) > 0\}$$



support is an *ordinary set*

Core (Kernel) of a Fuzzy Set

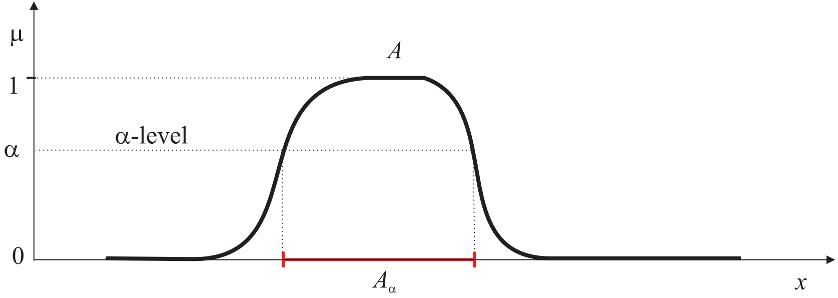
$$\text{core}(A) = \{x | \mu_A(x) = 1\}$$



core is an *ordinary set*

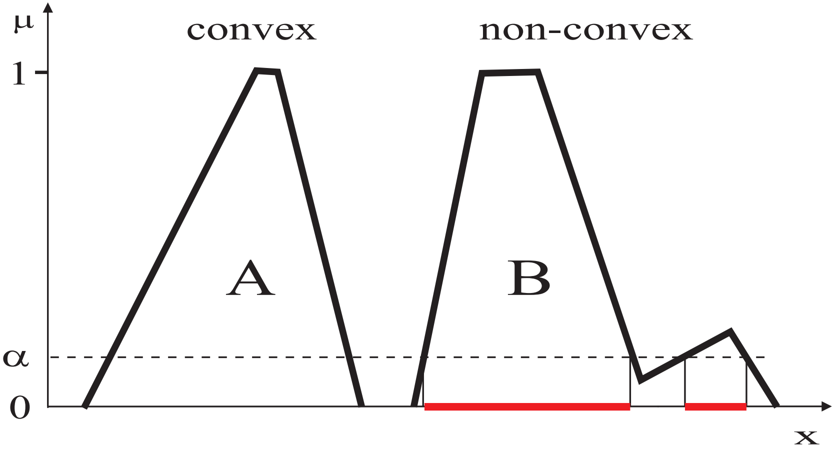
α-cut of a Fuzzy Set

$$A_\alpha = \{x | \mu_A(x) > \alpha\} \qquad \text{or} \qquad A_\alpha = \{x | \mu_A(x) \geq \alpha\}$$



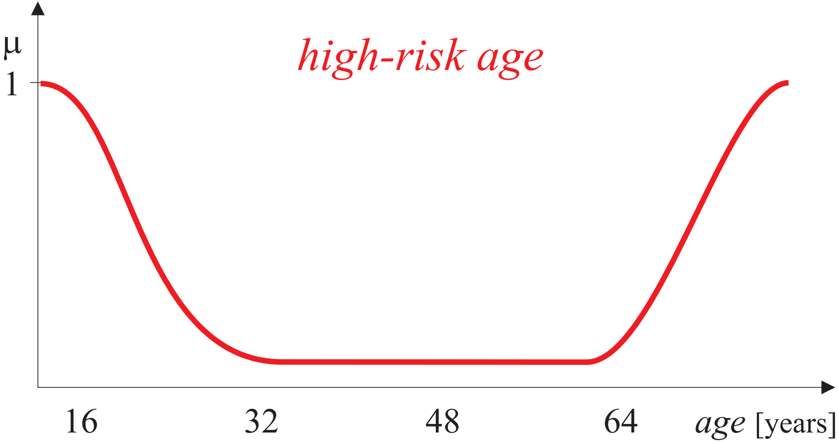
A_α is an *ordinary set*

Convex and Non-Convex Fuzzy Sets



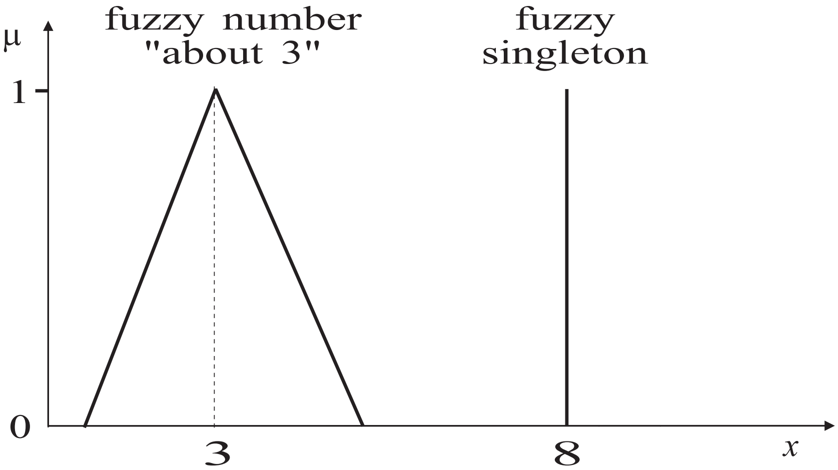
A fuzzy set is **convex** ⇔ all its α-cuts are convex sets.

Non-Convex Fuzzy Set: an Example



High-risk age for car insurance policy.

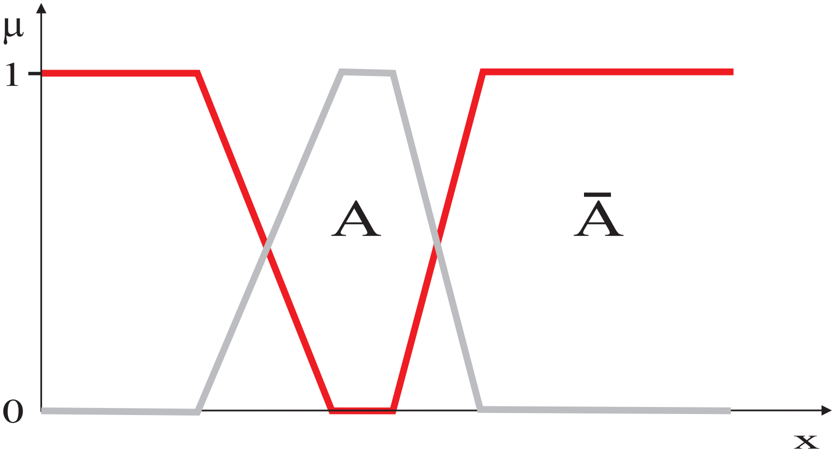
Fuzzy Numbers and Singletons



Fuzzy linear regression: $y = \tilde{3}x_1 + \tilde{5}x_2$

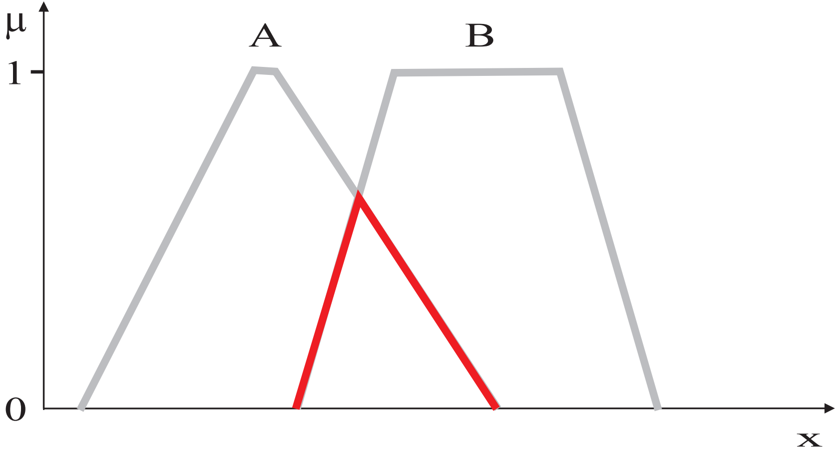
Fuzzy set-theoretic operations

Complement (Negation) of a Fuzzy Set



$$\mu_{\bar{A}}(x) = 1 - \mu_A(x)$$

Intersection (Conjunction) of Fuzzy Sets



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

Other Intersection Operators (T-norms)

Probabilistic "and" (product operator):

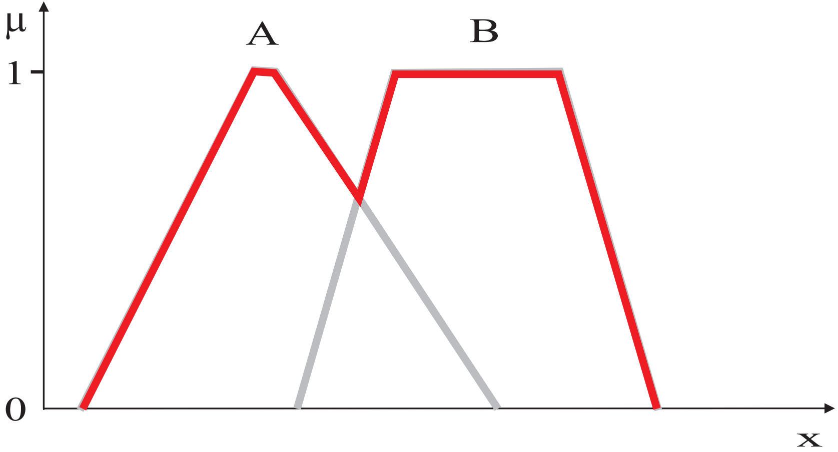
$$\mu_{A\cap B}(x) = \mu_A(x) \cdot \mu_B(x)$$

Łukasiewicz "and" (bounded difference):

$$\mu_{A\cap B}(x) = \max(0, \mu_A(x) + \mu_B(x) - 1)$$

Many other t-norms ... $[0,1] \times [0,1] \rightarrow [0,1]$

Union (Disjunction) of Fuzzy Sets



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

Other Union Operators (T-conorms)

Probabilistic "or":

$$\mu_{A\cup B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x) \cdot \mu_B(x)$$

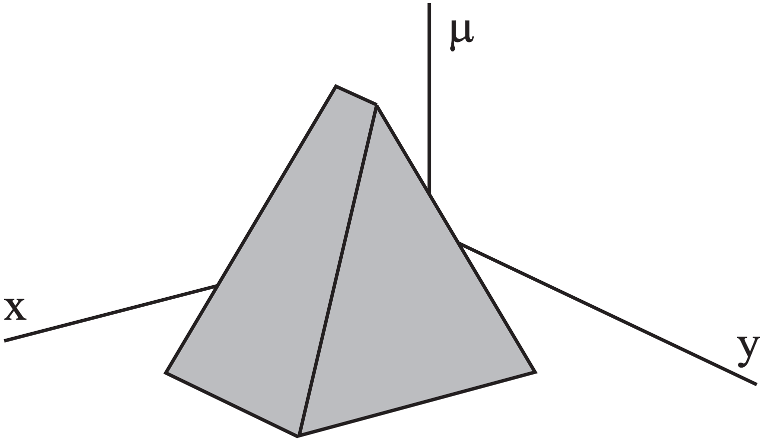
Łukasiewicz "or" (bounded sum):

$$\mu_{A\cup B}(x) = \min(1, \mu_A(x) + \mu_B(x))$$

Many other t-conorms . . . $[0, 1] \times [0, 1] \to [0, 1]$

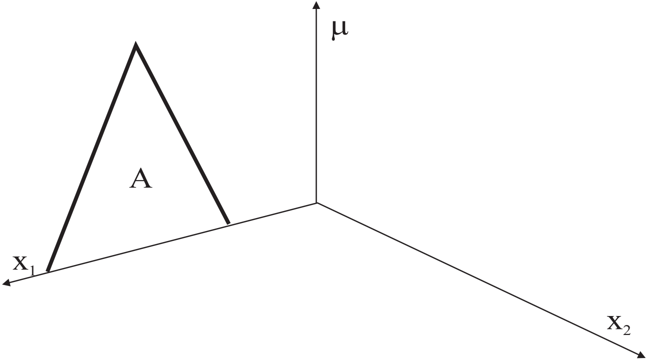
Demo of a Matlab tool

Fuzzy Set in Multidimensional Domains

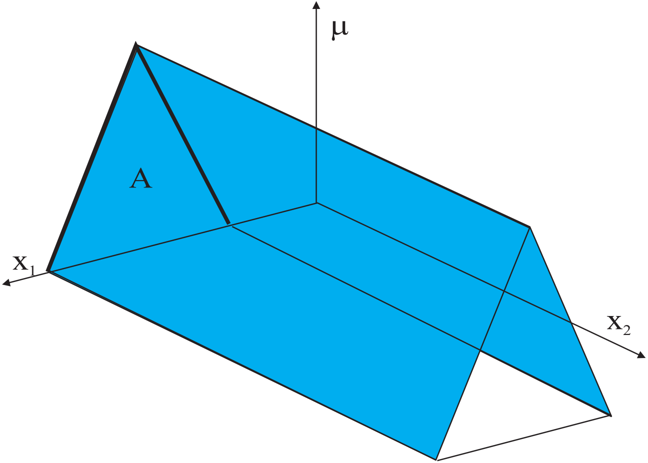


$$A = \{\mu_A(x, y)/(x, y) | (x, y) \in X \times Y\}$$

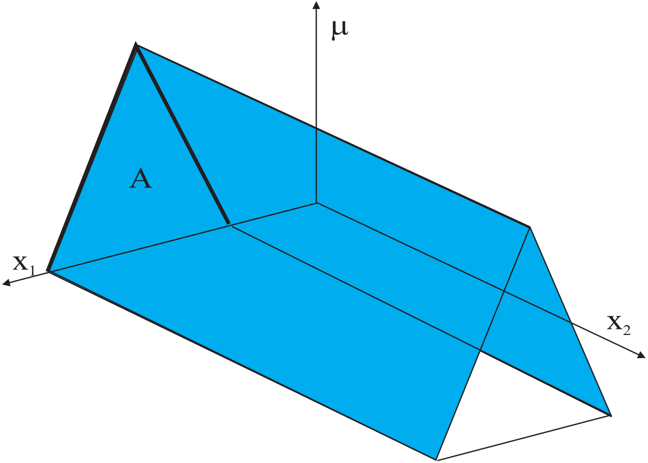
Cylindrical Extension



Cylindrical Extension

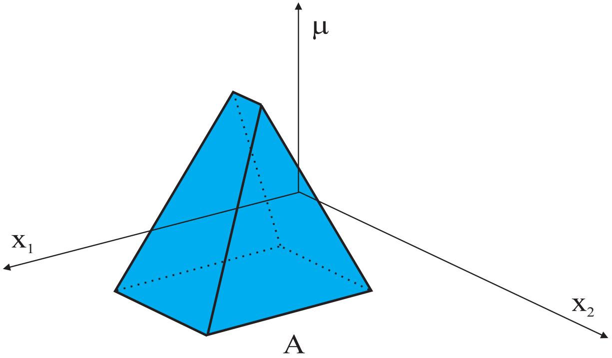


Cylindrical Extension

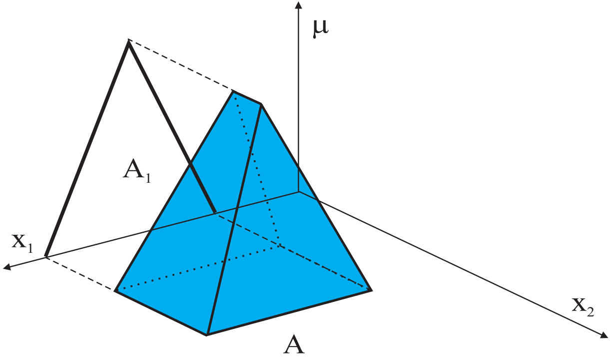


$$\text{ext}_{x_2}(A) = \{\mu_A(x_1)/(x_1, x_2) | (x_1, x_2) \in X_1 \times X_2\}$$

Projection

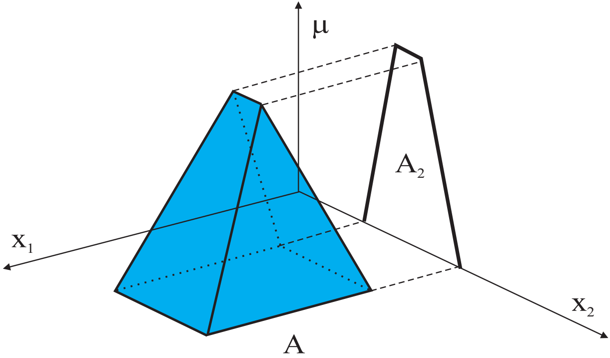


Projection onto $\mathbf{X_1}$



$$\text{proj}_{x_1}(A) = \left\{ \sup_{x_2 \in X_2} \mu_A(x_1, x_2)/x_1 | x_1 \in X_1 \right\}$$

Projection onto $\mathbf{X_2}$

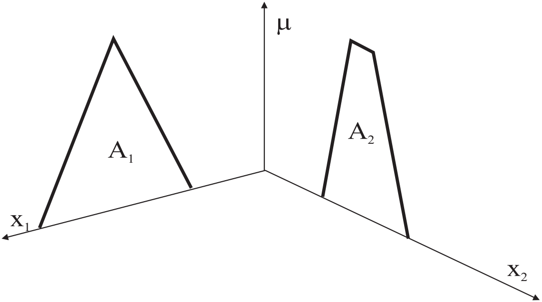


$$\text{proj}_{x_2}(A) = \left\{ \sup_{x_1 \in X_1} \mu_A(x_1, x_2)/x_2 | x_2 \in X_2 \right\}$$

Intersection on Cartesian Product Space

An operation between fuzzy sets are defined in different domains results in a multi-dimensional fuzzy set.

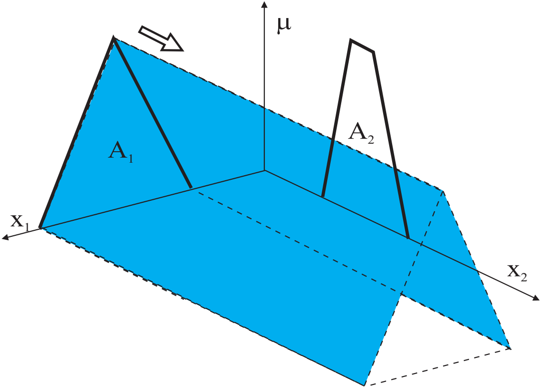
Example: $A_1 \cap A_2$ on $X_1 \times X_2$:



Intersection on Cartesian Product Space

An operation between fuzzy sets are defined in different domains results in a multi-dimensional fuzzy set.

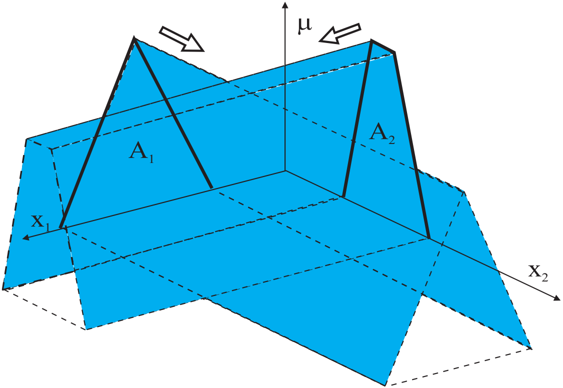
Example: $A_1 \cap A_2$ on $X_1 \times X_2$:



Intersection on Cartesian Product Space

An operation between fuzzy sets are defined in different domains results in a multi-dimensional fuzzy set.

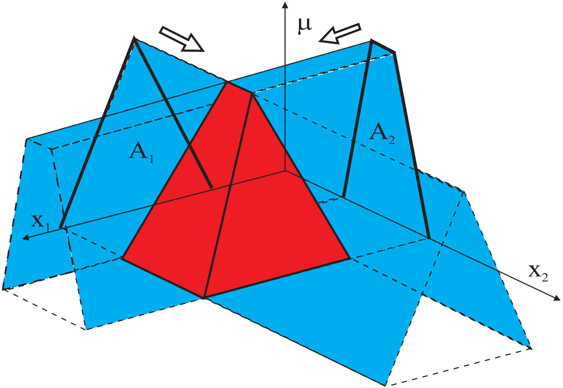
Example: $A_1 \cap A_2$ on $X_1 \times X_2$:



Intersection on Cartesian Product Space

An operation between fuzzy sets are defined in different domains results in a multi-dimensional fuzzy set.

Example: $A_1 \cap A_2$ on $X_1 \times X_2$:



Fuzzy Relations

Classical relation represents the presence or absence of interaction between the elements of two or more sets.

With fuzzy relations, the degree of association (correlation) is represented by membership grades.

An n-dimensional fuzzy relation is a mapping

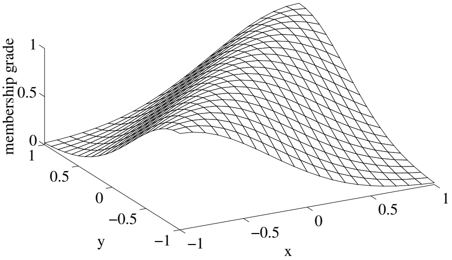
$$R : X_1 \times X_2 \times X_3 \cdots \times X_n \rightarrow [0, 1]$$

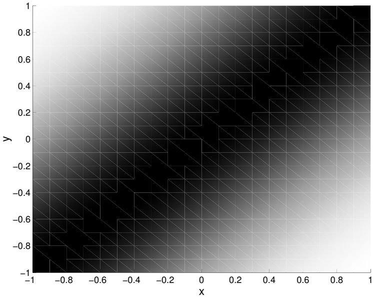
which assigns membership grades to all n-tuples $(x_1, x_2, \ldots, x_n)$ from the Cartesian product universe.

Fuzzy Relations: Example

Example: $R: \; x \approx y$ ("x is approximately equal to y")

$$\mu_R(x, y) = e^{-(x-y)^2}$$





Relational Composition

Given fuzzy relation R defined in $X \times Y$ and fuzzy set A defined in X, derive the corresponding fuzzy set B defined in Y:

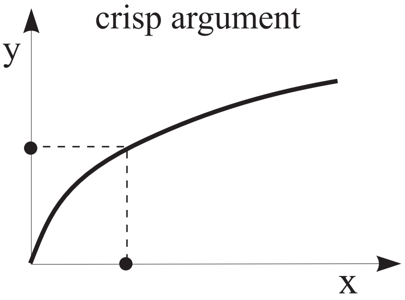
$$B = A \circ R = \mathrm{proj}_Y(\mathrm{ext}_{X \times Y}(A) \cap R)$$

max-min composition:

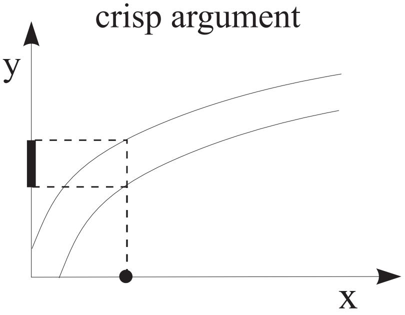
$$\mu_B(y) = \max_x \Big(\min(\mu_A(x), \mu_R(x, y)) \Big)$$

Analogous to evaluating a function.

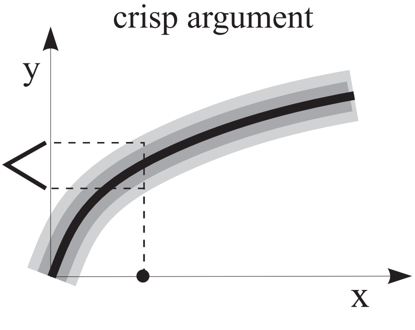
Graphical Interpretation: Crisp Function



Graphical Interpretation: Interval Function



Graphical Interpretation: Fuzzy Relation



Max-Min Composition: Example

$$\mu_B(y) = \max_x \Big(\min(\mu_A(x), \mu_R(x, y)) \Big), \quad \forall y$$

$$\begin{bmatrix} 1.0 & 0.4 & 0.1 & 0.0 & 0.0 \end{bmatrix} \circ \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.4 & 0.8 \\ 0.0 & 0.1 & 1.0 & 0.2 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.9 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.8 & 0.3 & 0.0 \end{bmatrix} =$$

Max-Min Composition: Example

$$\mu_B(y) = \max_x\Big(\min(\mu_A(x), \mu_R(x, y))\Big), \quad \forall y$$

$$\begin{bmatrix} 1.0 & 0.4 & 0.1 & 0.0 & 0.0 \end{bmatrix} \circ \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.4 & 0.8 \\ 0.0 & 0.1 & 1.0 & 0.2 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.9 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.8 & 0.3 & 0.0 \end{bmatrix} = \begin{bmatrix} 0.0 & 0.1 & 0.4 & 0.4 & 0.8 \end{bmatrix}$$

Fuzzy Systems

Fuzzy Systems

- Systems with fuzzy parameters

$$y = \tilde{3}x_1 + \tilde{5}x_2$$

- Fuzzy inputs and states

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad x(0) = \tilde{2}$$

- Rule-based systems

If *the heating power is high*
then *the temperature will increase fast*

Rule-based Fuzzy Systems

- Linguistic (Mamdani) fuzzy model

$$\textbf{If } x \textbf{ is } A \textbf{ then } y \textbf{ is } B$$

- Fuzzy relational model

$$\textbf{If } x \textbf{ is } A \textbf{ then } y \textbf{ is } B_1(0.1), B_2(0.8)$$

- Takagi–Sugeno fuzzy model

$$\textbf{If } x \textbf{ is } A \textbf{ then } y = f(x)$$

Linguistic Model

If x **is** A **then** y **is** B

x **is** A – antecedent (fuzzy proposition)

y **is** B – consequent (fuzzy proposition)

Linguistic Model

If x **is** A **then** y **is** B

x **is** A – antecedent (fuzzy proposition)

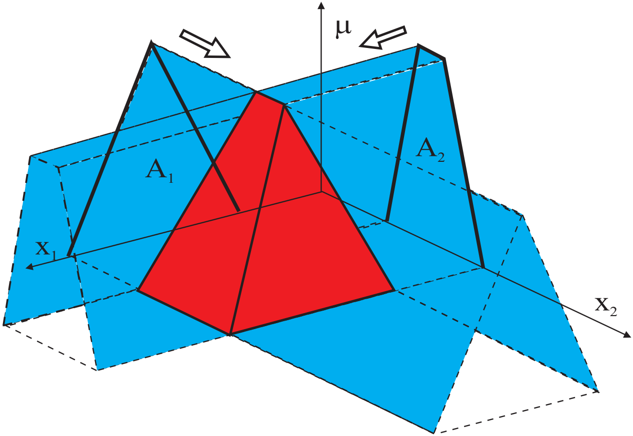
y **is** B – consequent (fuzzy proposition)

Compound propositions (logical connectives, hedges):

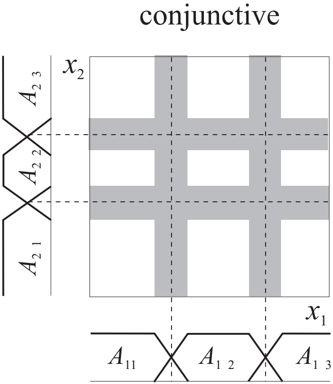
If x_1 **is** very big and x_2 **is** not small

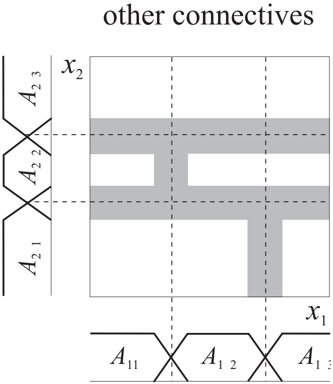
Multidimensional Antecedent Sets

$A_1 \cap A_2$ on $X_1 \times X_2$:



Partitioning of the Antecedent Space





Inference Mechanism

Given the if-then rules and an input fuzzy set, deduce the corresponding output fuzzy set.

- Formal approach based on fuzzy relations.
- Simplified approach (Mamdani inference).
- Interpolation (additive fuzzy systems).

Formal Approach

1. Represent each if–then rule as a fuzzy relation.
2. Aggregate these relations in one relation representative for the entire rule base.
3. Given an input, use *relational composition* to derive the corresponding output.

Modus Ponens Inference Rule

Classical logic

if x is A **then** y is B
x is A
y is B

Fuzzy logic

if x is A **then** y is B
x is A'
y is B'

Relational Representation of Rules

If–then rules can be represented as a *relation*, using implications or conjunctions.

Classical implication

A	B	$A \to B$ $(\neg A \vee B)$
0	0	1
0	1	1
1	0	0
1	1	1

$A \backslash B$	0	1
0	1	1
1	0	1

$$R\colon \{0, 1\} \times \{0, 1\} \to \{0, 1\}$$

Relational Representation of Rules

If–then rules can be represented as a *relation*, using implications or conjunctions.

Conjunction

A	B	$A \wedge B$
0	0	0
0	1	0
1	0	0
1	1	1

$A \backslash B$	0	1
0	0	0
1	0	1

$$R\colon \{0,1\} \times \{0,1\} \to \{0,1\}$$

Fuzzy Implications and Conjunctions

Fuzzy implication is represented by a fuzzy relation:

$$R\colon [0,1] \times [0,1] \to [0,1]$$

$$\mu_R(x,y) \quad = \quad \mathrm{I}(\mu_A(x), \mu_B(y))$$

$\mathrm{I}(a,b)$ – implication function

"classical"	Kleene–Diene	$\mathrm{I}(a,b) = \max(1-a, b)$
	Łukasiewicz	$\mathrm{I}(a,b) = \min(1, 1-a+b)$
T-norms	Mamdani	$\mathrm{I}(a,b) = \min(a,b)$
	Larsen	$\mathrm{I}(a,b) = a \cdot b$

Inference With One Rule

1. Construct implication relation:

$$\mu_R(x, y) = \mathrm{I}(\mu_A(x), \mu_B(y))$$

Inference With One Rule

1. Construct implication relation:

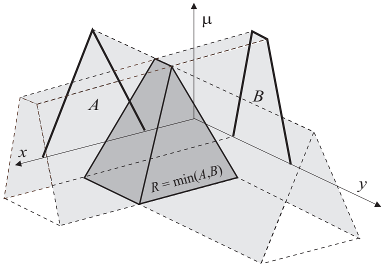
$$\mu_R(x, y) = \mathrm{I}(\mu_A(x), \mu_B(y))$$

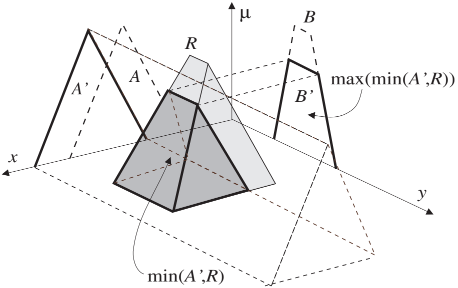
2. Use relational composition to derive B' from A':

$$B' = A' \circ R$$

Graphical Illustration

$$\mu_R(x,y) = \min\left(\mu_A(x), \mu_B(y)\right) \qquad \mu_{B'}(y) = \max_x\left(\min\left(\mu_{A'}(x), \mu_R(x,y)\right)\right)$$





Inference With Several Rules

1. Construct implication relation for each rule i:

$$\mu_{R_i}(x, y) = \mathrm{I}(\mu_{A_i}(x), \mu_{B_i}(y))$$

2. Aggregate relations R_i into one:

$$\mu_R(x, y) = \mathrm{aggr}(\mu_{R_i}(x, y))$$

   The aggr operator is the minimum for implications and the maximum for conjunctions.

3. Use relational composition to derive B' from A':

$$B' = A' \circ R$$

Example: Conjunction

1. Each rule

$$\textbf{If } \tilde{x} \textbf{ is } A_i \textbf{ then } \tilde{y} \textbf{ is } B_i$$

is represented as a fuzzy relation on $X \times Y$:

$$R_i = A_i \times B_i \qquad \mu_{R_i}(\mathbf{x}, \mathbf{y}) = \mu_{A_i}(\mathbf{x}) \wedge \mu_{B_i}(\mathbf{y})$$

Example: Conjunction, Aggregation

1. Each rule

$$\textbf{If } \tilde{x} \textbf{ is } A_i \textbf{ then } \tilde{y} \textbf{ is } B_i$$

is represented as a fuzzy relation on $X \times Y$:

$$R_i = A_i \times B_i \qquad \mu_{R_i}(\boldsymbol{x}, \boldsymbol{y}) = \mu_{A_i}(\boldsymbol{x}) \wedge \mu_{B_i}(\boldsymbol{y})$$

2. The entire rule base's relation is the union:

$$R = \bigcup_{i=1}^{K} R_i \qquad \mu_R(\boldsymbol{x}, \boldsymbol{y}) = \max_{1 \le i \le K} [\mu_{R_i}(\boldsymbol{x}, \boldsymbol{y})]$$

Example: Conjunction, Aggregation, and Composition

1. Each rule

$$\textbf{If } \tilde{x} \textbf{ is } A_i \textbf{ then } \tilde{y} \textbf{ is } B_i$$

is represented as a fuzzy relation on $X \times Y$:

$$R_i = A_i \times B_i \qquad \mu_{R_i}(\boldsymbol{x}, \boldsymbol{y}) = \mu_{A_i}(\boldsymbol{x}) \wedge \mu_{B_i}(\boldsymbol{y})$$

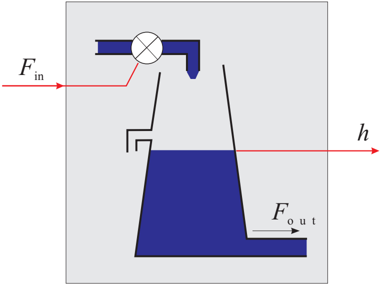
2. The entire rule base's relation is the union:

$$R = \bigcup_{i=1}^{K} R_i \qquad \mu_R(\boldsymbol{x}, \boldsymbol{y}) = \max_{1 \le i \le K} [\mu_{R_i}(\boldsymbol{x}, \boldsymbol{y})]$$

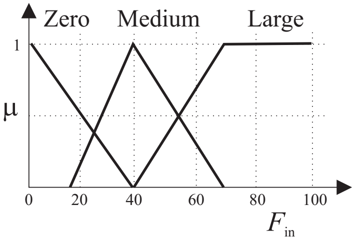
3. Given an input value A' the output value B' is:

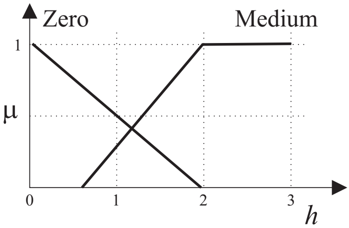
$$B' = A' \circ R \qquad \mu_{B'}(\boldsymbol{y}) = \max_X [\mu_{A'}(\boldsymbol{x}) \wedge \mu_R(\boldsymbol{x}, \boldsymbol{y})]$$

Example: Modeling of Liquid Level

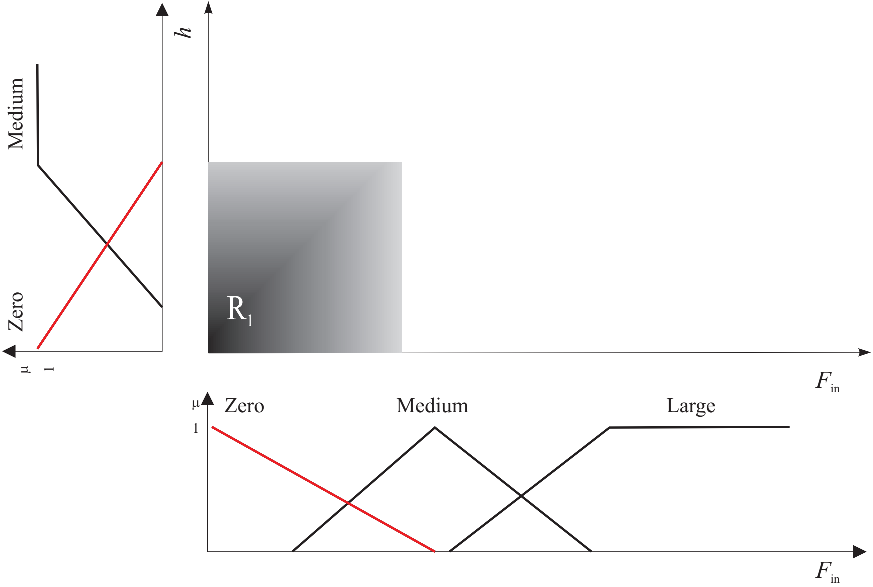


- **If** F_{in} **is** Zero **then** h **is** Zero
- **If** F_{in} **is** Med **then** h **is** Med
- **If** F_{in} **is** Large **then** h **is** Med

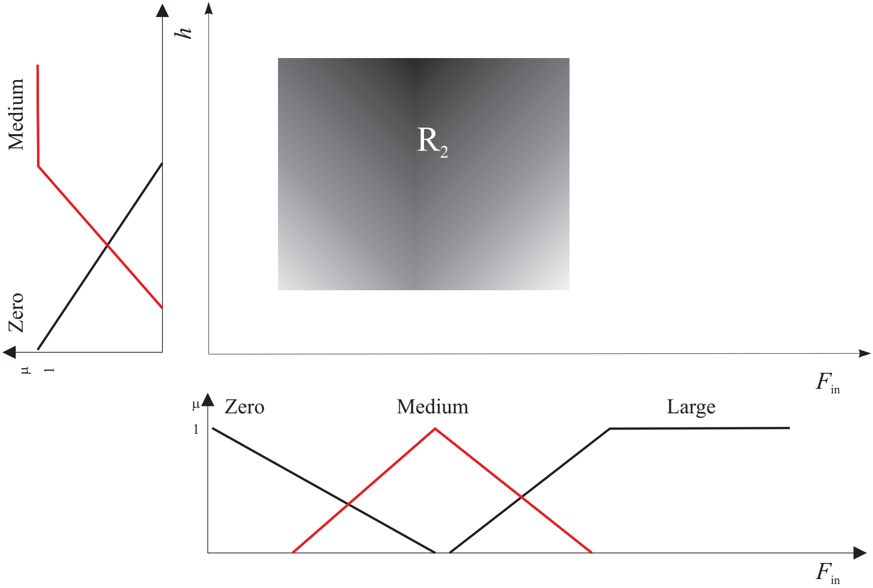




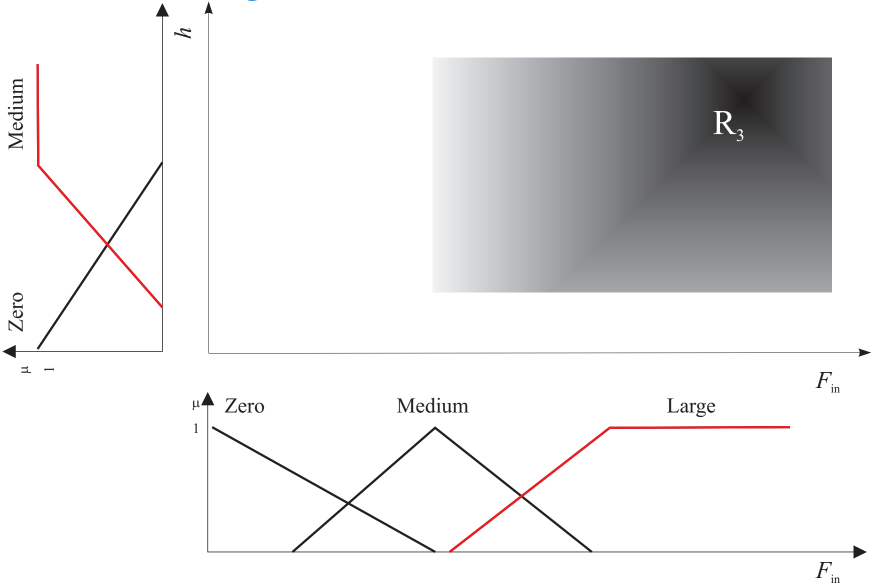
$\mathcal{R}_1$ If Flow is Zero then Level is Zero



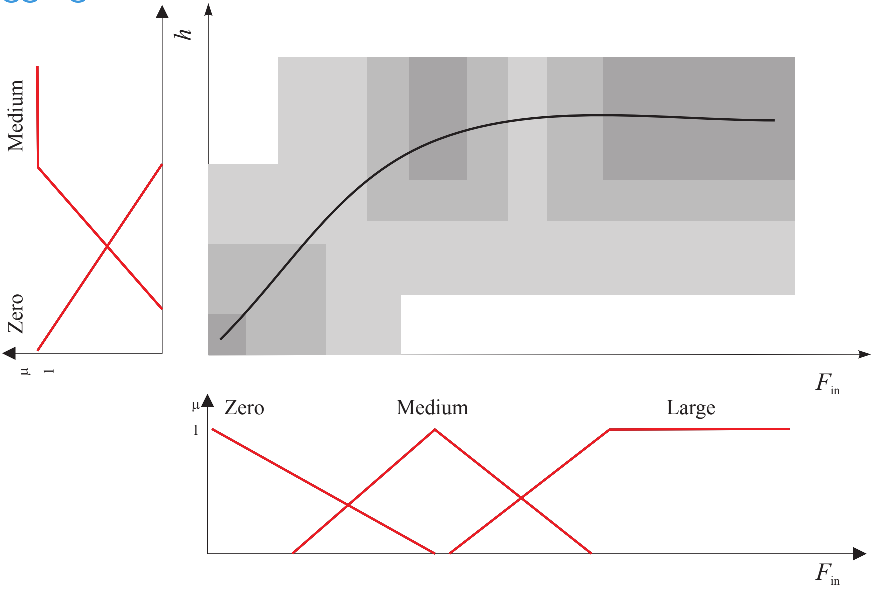
$\mathcal{R}_2$ If Flow is Medium then Level is Medium



$\mathcal{R}_3$ If Flow is Large then Level is Medium



Aggregated Relation

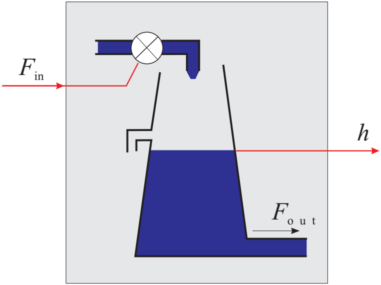


Simplified Approach

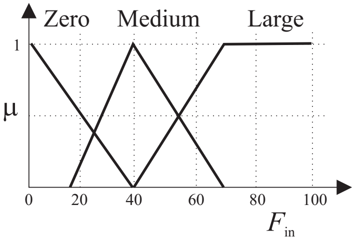
1. Compute the match between the input and the antecedent membership functions (*degree of fulfillment*).
2. Clip the corresponding output fuzzy set for each rule by using the degree of fulfillment.
3. Aggregate output fuzzy sets of all the rules into one fuzzy set.

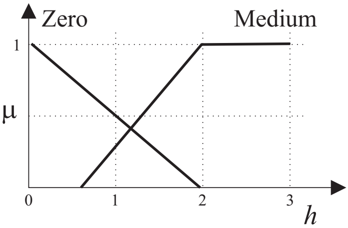
This is called the *Mamdani* or *max-min* inference method.

Water Tank Example

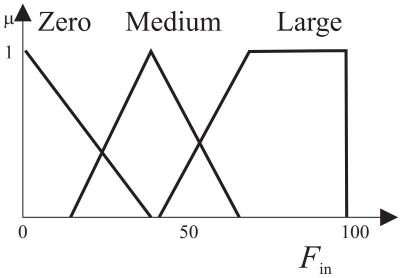


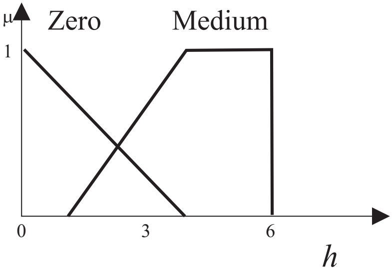
- **If** F_{in} **is** Zero **then** h **is** Zero
- **If** F_{in} **is** Med **then** h **is** Med
- **If** F_{in} **is** Large **then** h **is** Med



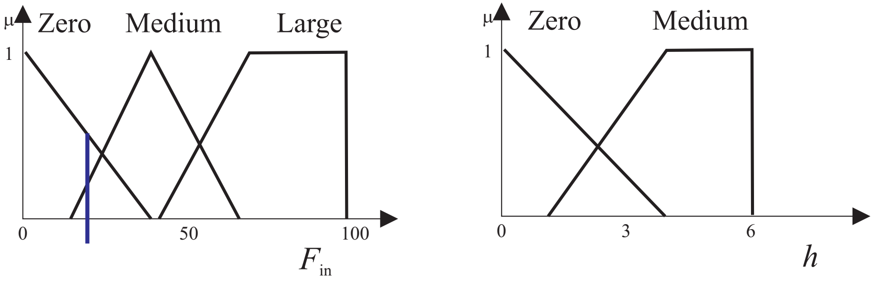


Mamdani Inference: Example



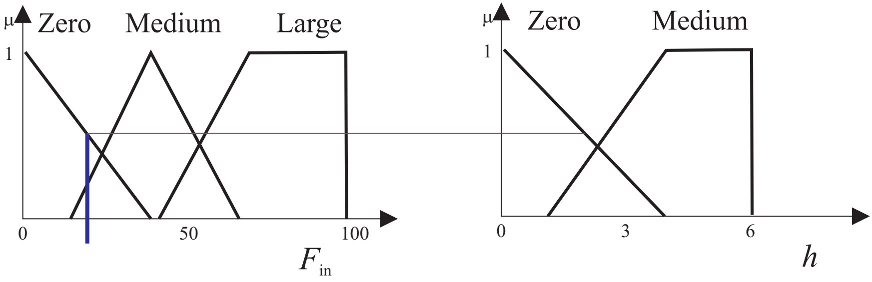


Mamdani Inference: Example



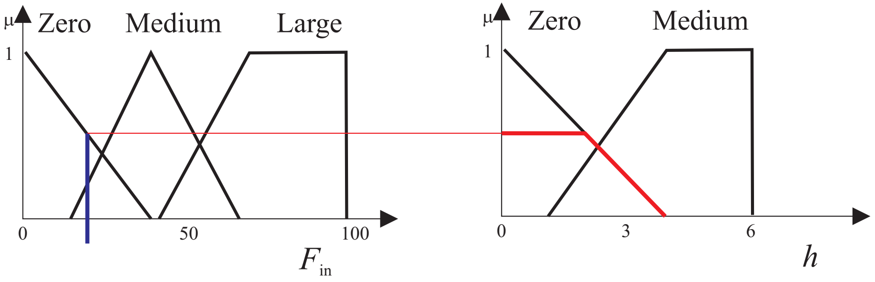
Given a crisp (numerical) input (F_{in}).

If F_{in} is Zero then ...



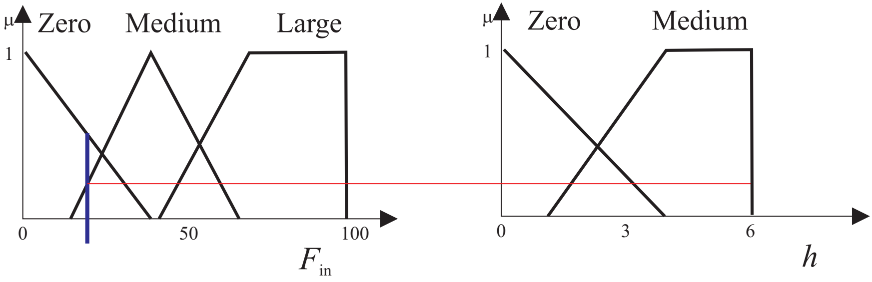
Determine the degree of fulfillment (truth) of the first rule.

If F_{in} is Zero then h is Zero



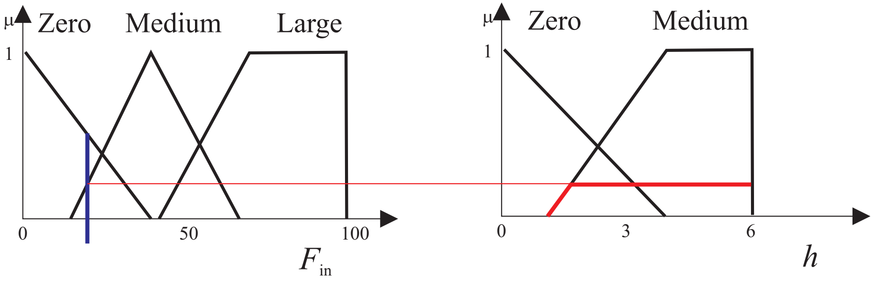
Clip consequent membership function of the first rule.

If F_{in} is Medium then . . .



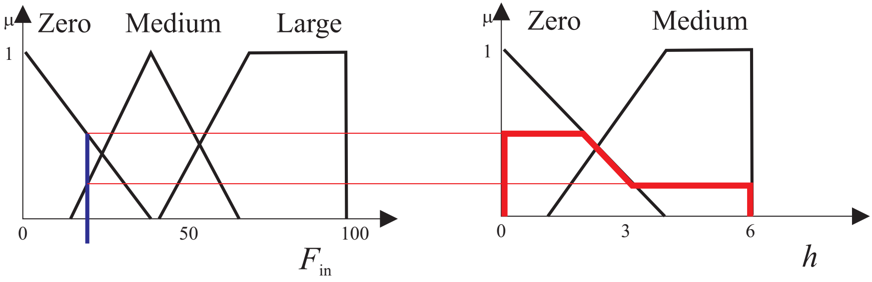
Determine the degree of fulfillment (truth) of the second rule.

If F_{in} is Medium then h is Medium



Clip consequent membership function of the second rule.

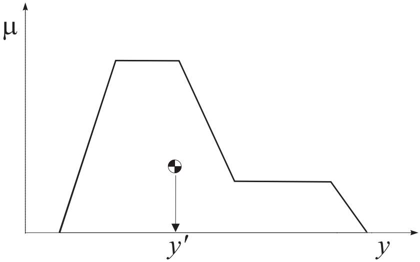
Aggregation



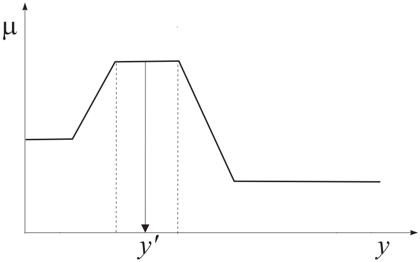
Combine the result of the two rules (union).

Defuzzification

conversion of a fuzzy set to a crisp value



(a) center of gravity

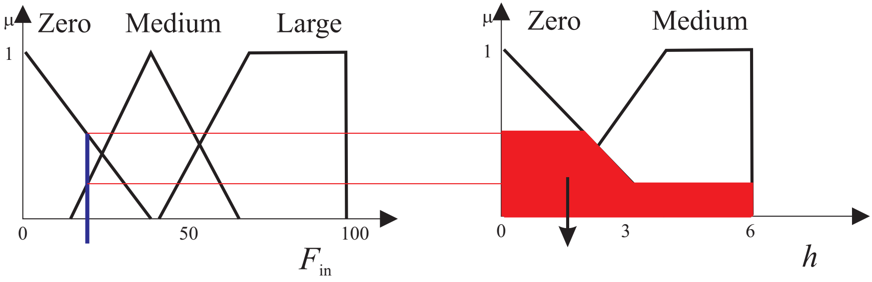


(b) mean of maxima

Center-of-Gravity Method

$$y_0 = \frac{\sum_{j=1}^{F} \mu_{B'}(y_j) y_j}{\sum_{j=1}^{F} \mu_{B'}(y_j)}$$

Defuzzification



Compute a crisp (numerical) output of the model (center-of-gravity method).

Outline

1. Singleton and Takagi–Sugeno fuzzy system.
2. Knowledge based fuzzy modeling.
3. Data-driven construction.

Singleton Fuzzy Model

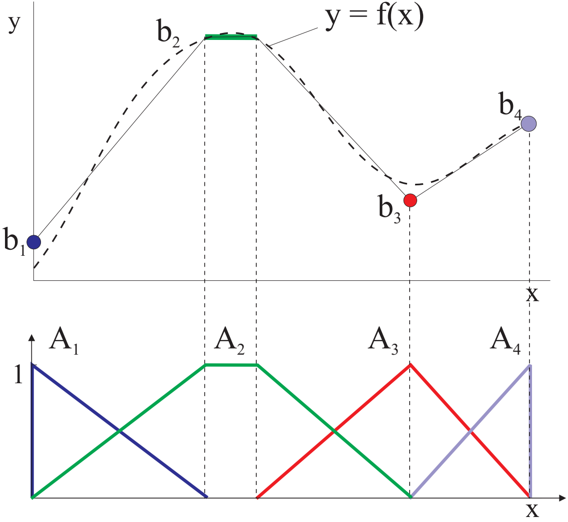
$$\textbf{If } x \textbf{ is } A_i \textbf{ then } y = b_i$$

Inference/defuzzification:

$$y = \frac{\sum_{i=1}^{K} \mu_{A_i}(x) b_i}{\sum_{i=1}^{K} \mu_{A_i}(x)}$$

- well-understood approximation properties
- straightforward parameter estimation

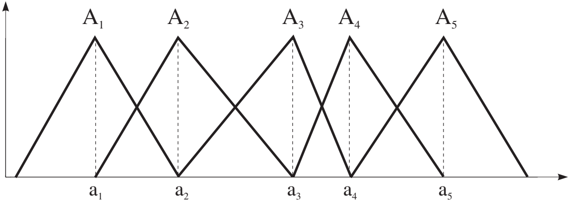
Piece-wise Linear Approximation



Linear Mapping with a Singleton Model

$$y = \boldsymbol{k}^T \boldsymbol{x} + q = \sum_{j=1}^{p} k_j x_j + q$$

- Triangular partition:



- Consequent singletons are equal to:

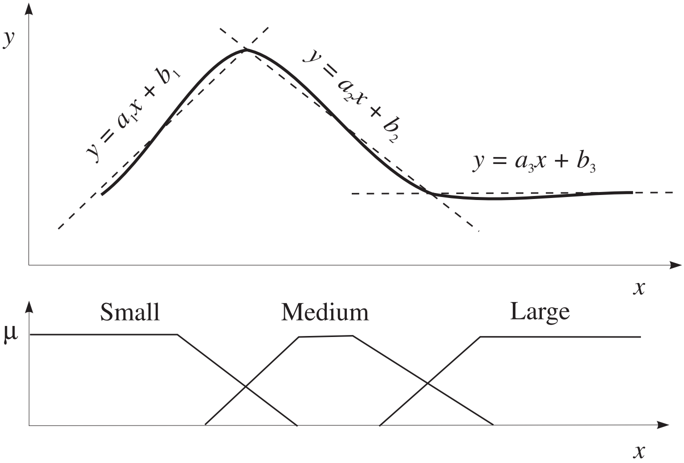
$$b_i = \sum_{j=1}^{p} k_j a_{i,j} + q$$

Takagi–Sugeno (TS) Fuzzy Model

$$\textbf{If } x \textbf{ is } A_i \textbf{ then } y_i = a_i x + b_i$$

$$y = \frac{\sum_{i=1}^{K} \mu_{A_i}(x) y_i}{\sum_{i=1}^{K} \mu_{A_i}(x)} = \frac{\sum_{i=1}^{K} \mu_{A_i}(x)(a_i x + b_i)}{\sum_{i=1}^{K} \mu_{A_i}(x)}$$

Input-Output Mapping of the TS Model



Consequents are approximate local linear models of the system.

TS Model is a Quasi-Linear System

$$y = \frac{\sum_{i=1}^{K} \mu_{A_i}(\boldsymbol{x}) y_i}{\sum_{j=1}^{K} \mu_{A_j}(\boldsymbol{x})} = \frac{\sum_{i=1}^{K} \mu_{A_i}(\boldsymbol{x})(\boldsymbol{a}_i^T \boldsymbol{x} + b_i)}{\sum_{j=1}^{K} \mu_{A_j}(\boldsymbol{x})}$$

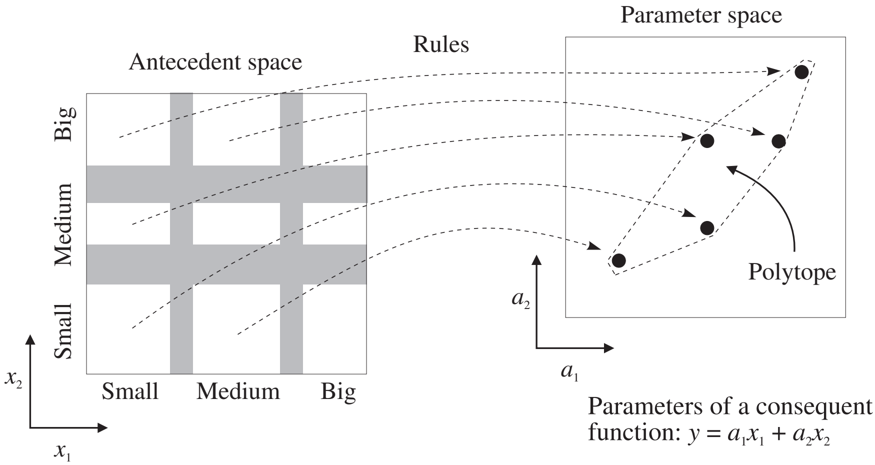
TS Model is a Quasi-Linear System

$$y = \frac{\sum_{i=1}^{K} \mu_{A_i}(\boldsymbol{x}) y_i}{\sum_{j=1}^{K} \mu_{A_j}(\boldsymbol{x})} = \frac{\sum_{i=1}^{K} \mu_{A_i}(\boldsymbol{x})(\boldsymbol{a}_i^T \boldsymbol{x} + b_i)}{\sum_{j=1}^{K} \mu_{A_j}(\boldsymbol{x})}$$

$$y = \underbrace{\left(\sum_{i=1}^{K} \gamma_i(\boldsymbol{x}) \boldsymbol{a}_i^T \right)}_{\boldsymbol{a}(\boldsymbol{x})^T} \boldsymbol{x} + \underbrace{\sum_{i=1}^{K} \gamma_i(\boldsymbol{x}) b_i}_{b(\boldsymbol{x})}$$

linear in parameters a_i and b_i, pseudo-linear in $\boldsymbol{x}$ (LPV)

TS Model is a Polytopic System



Construction of Fuzzy Models

Modeling Paradigms

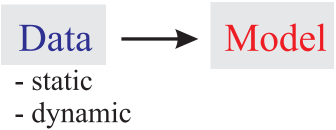
- Mechanistic (white-box, physical)
- Qualitative (naive physics, knowledge-based)
- Data-driven (black-box, inductive)

Often combination of different approaches semi-mechanistic, gray-box modeling.

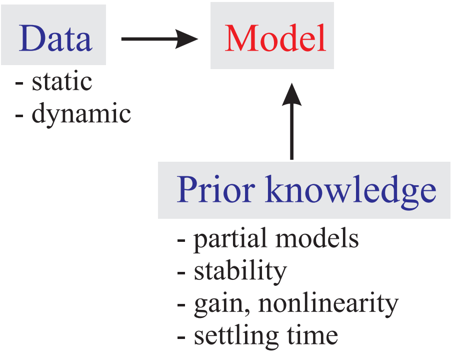
Parameterization of nonlinear models

- polynomials, splines
- look-up tables
- fuzzy systems
- neural networks
- (neuro-)fuzzy systems
- radial basis function networks
- wavelet networks
- ...

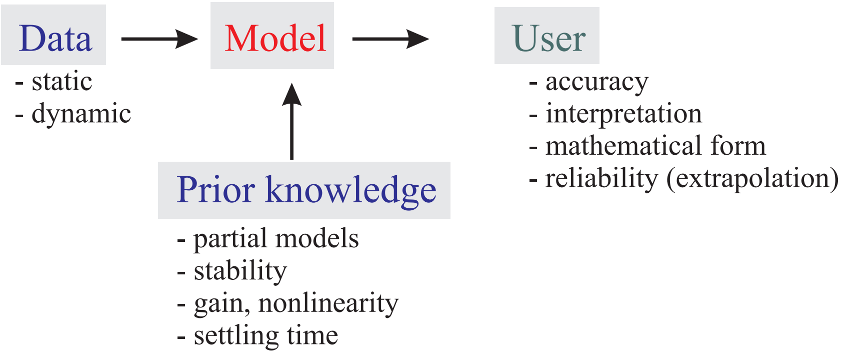
Modeling of Complex Systems



Modeling of Complex Systems



Modeling of Complex Systems



Building Fuzzy Models

Knowledge-based approach:

- expert knowledge $\rightarrow$ rules & membership functions
- fuzzy model of human operator
- linguistic interpretation

Building Fuzzy Models

Knowledge-based approach:

- expert knowledge $\rightarrow$ rules & membership functions
- fuzzy model of human operator
- linguistic interpretation

Data-driven approach:

- nonlinear mapping, universal approximation
- extract rules & membership functions from data

Knowledge-Based Modeling

- Problems where little or no data are available.
- Similar to expert systems.
- Presence of both quantitative and qualitative variables or parameters.

Typical applications: fuzzy control and decision support, but also modeling of poorly understood processes

Wear Prediction for a Trencher

Trencher T-850 (Vermeer)

Chain Detail

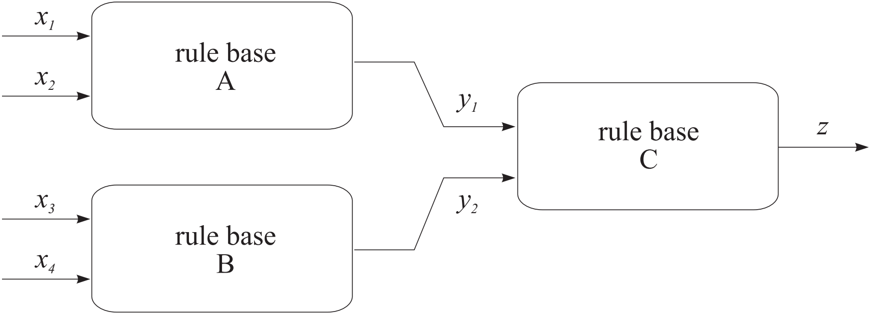
Goal: Given the terrain properties, predict bit wear and production rate of trencher.

Why Knowledge-Based Modeling?

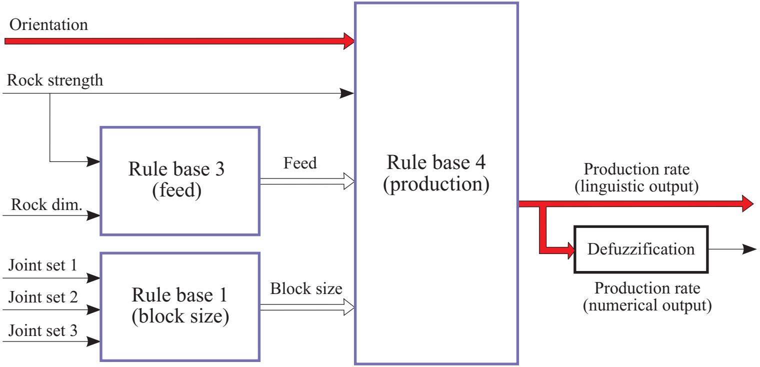
- Interaction between tool and environment is complex, dynamic and highly nonlinear, rigorous mathematical models are not available.
- Little data (15 data points) to develop statistical regression models.
- Input data are a mixture of numerical measurements (rock strength, joint spacing, trench dimensions) and qualitative information (joint orientation).
- Precise numerical output not needed, qualitative assessment is sufficient.

Dimensionality Problem: Hierarchical Structure

Assume 5 membership functions for each input
625 rules in a flat rule base vs. 75 rules in a hierarchical one

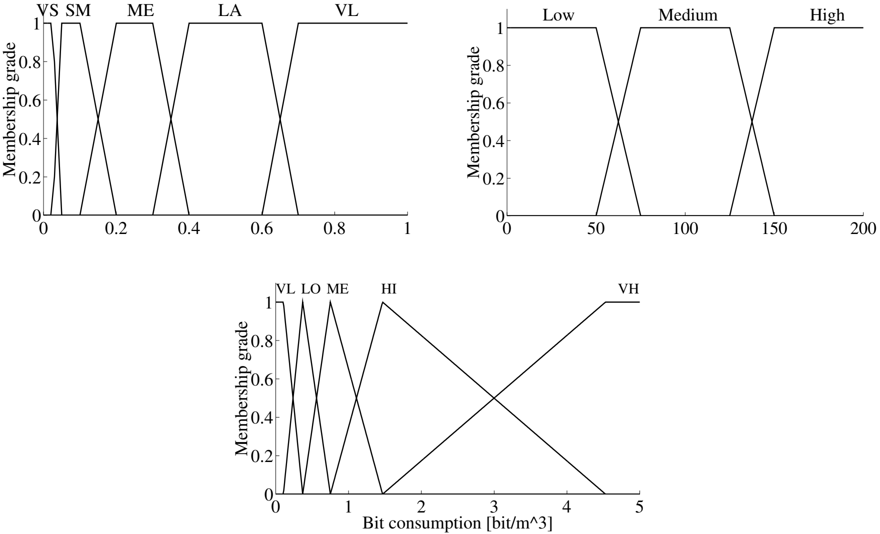


Trencher: Fuzzy Rule Bases




```
If TRENCH-DIM is SMALL  and STRENGTH is LOW    Then FEED is VERY-HIGH;
If TRENCH-DIM is SMALL  and STRENGTH is MEDIUM Then FEED is HIGH;
....
If JOINT-SP is EXTRA-LARGE and FEED is VERY-HIGH Then PROD is VERY-HIGH
```

Example of Membership Functions



Output: Prediction of Production Rate

```
data no.  measured value   predicted linguistic value(s)
---------------------------------------------------------

1               2.07       VERY-LOW       1.00
2               5.56       HIGH           1.00
3              23.60       VERY-HIGH      0.50
4              11.90       HIGH           0.40    VERY-HIGH 0.60
5               7.71       MEDIUM         1.00
6               7.17       LOW            0.72
7               8.05       MEDIUM         0.80
8               7.39       LOW            1.00
9               4.58       LOW            0.50
10              8.74       MEDIUM         1.00
11            134.84       EXTREMELY-HIGH 1.00
```

Data-Driven Construction

Structure and Parameters

Structure:

- Input and output variables. For dynamic systems also the representation of the dynamics.
- Number of membership functions per variable, type of membership functions, number of rules.

Parameters:

- Consequent parameters (least squares).
- Antecedent membership functions (various methods).

Least-Squares Estimation of Singletons

$$R_i: \quad \textbf{If } \boldsymbol{x} \textbf{ is } A_i \textbf{ then } y = b_i$$

- Given A_i and a set of input–output data:

$$\{(\boldsymbol{x}_k, y_k) \mid k = 1, 2, \ldots, N\}$$

- Estimate optimal consequent parameters b_i.

Least-Squares Estimation of Singletons

1. Compute the membership degrees $\mu_{A_i}(\mathbf{x}_k)$
2. Normalize

$$\gamma_{ki} = \mu_{A_i}(\mathbf{x}_k) / \sum_{j=1}^{K} \mu_{A_j}(\mathbf{x}_k)$$

(Output: $y_k = \sum_{i=1}^{K} \gamma_{ki} b_i$, in a matrix form: $\mathbf{y} = \mathbf{\Gamma} \mathbf{b}$)

3. Least-squares estimate: $\mathbf{b} = \left[\mathbf{\Gamma}^T \mathbf{\Gamma} \right]^{-1} \mathbf{\Gamma}^T \mathbf{y}$

Least-Square Estimation of TS Consequents

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \\ \vdots \\ \boldsymbol{x}_N^T \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad \boldsymbol{\Gamma}_i = \begin{bmatrix} \gamma_{i1} & 0 & \cdots & 0 \\ 0 & \gamma_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_{iN} \end{bmatrix}$$

$$\boldsymbol{\theta}_i = \left[\boldsymbol{a}_i^T \quad b_i\right]^T, \quad \boldsymbol{X}_e = [\boldsymbol{X} \quad \mathbf{1}]$$

Least-Square Estimation of TS Consequents

- Global LS: $\boldsymbol{\theta}' = \left[(\boldsymbol{X}')^T \boldsymbol{X}'\right]^{-1} (\boldsymbol{X}')^T \boldsymbol{y}$

  with $\boldsymbol{X}' = [\boldsymbol{\Gamma}_1 \boldsymbol{X}_e \quad \boldsymbol{\Gamma}_2 \boldsymbol{X}_e \quad \ldots \quad \boldsymbol{\Gamma}_c \boldsymbol{X}_e]$

  and $\boldsymbol{\theta}' = \left[\boldsymbol{\theta}_1^T \quad \boldsymbol{\theta}_2^T \quad \ldots \quad \boldsymbol{\theta}_c^T\right]^T$

Least-Square Estimation of TS Consequents

- Global LS: $\boldsymbol{\theta}' = \left[(\boldsymbol{X}')^T \boldsymbol{X}'\right]^{-1} (\boldsymbol{X}')^T \boldsymbol{y}$

  with $\boldsymbol{X}' = [\boldsymbol{\Gamma}_1 \boldsymbol{X}_e \quad \boldsymbol{\Gamma}_2 \boldsymbol{X}_e \quad \ldots \quad \boldsymbol{\Gamma}_c \boldsymbol{X}_e]$

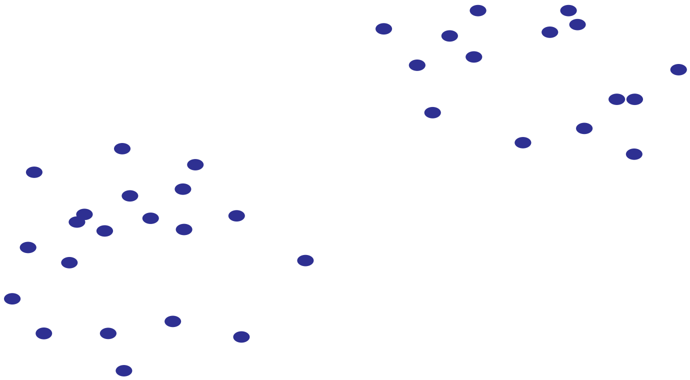
  and $\boldsymbol{\theta}' = \left[\boldsymbol{\theta}_1^T \quad \boldsymbol{\theta}_2^T \quad \ldots \quad \boldsymbol{\theta}_c^T\right]^T$

- Local LS: $\boldsymbol{\theta}_i = \left[\boldsymbol{X}_e^T \boldsymbol{\Gamma}_i \boldsymbol{X}_e\right]^{-1} \boldsymbol{X}_e^T \boldsymbol{\Gamma}_i \boldsymbol{y}$

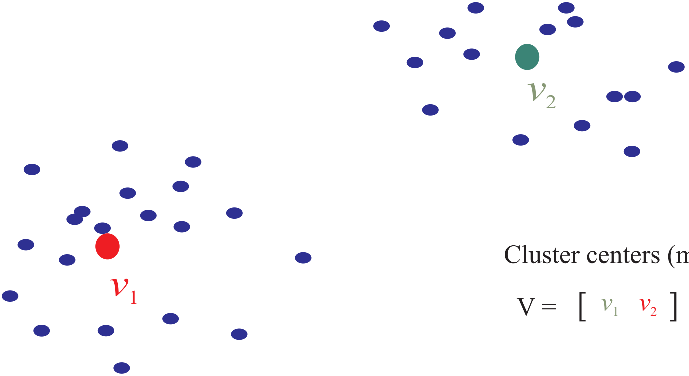
Antecedent Membership Functions

- templates (grid partitioning),
- nonlinear optimization (neuro-fuzzy methods),
- tree-construction
- product space fuzzy clustering

Fuzzy Clustering: Data



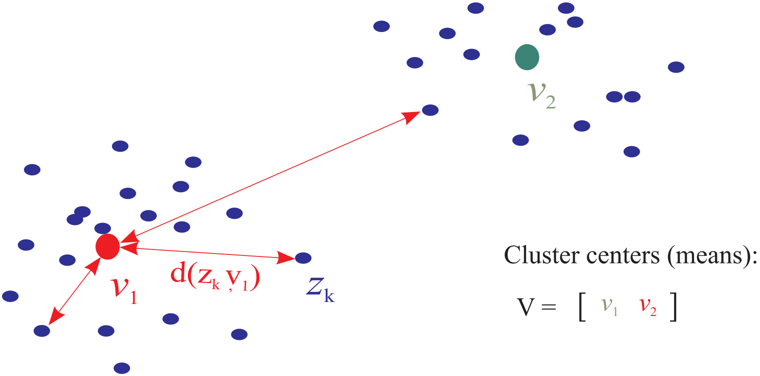
Fuzzy Clustering: Prototypes



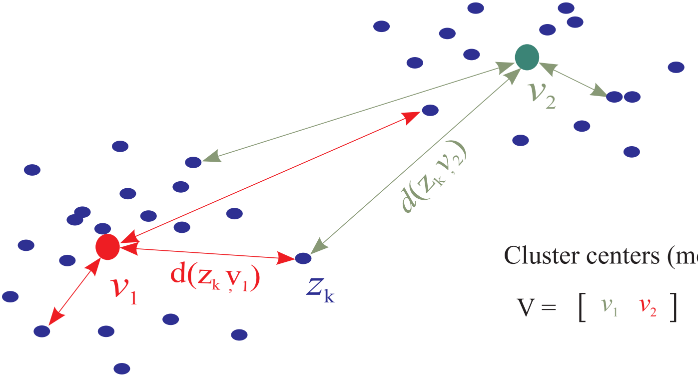
Cluster centers (means):

$$V = [\ v_1 \ \ v_2\ ]$$

Fuzzy Clustering: Distance



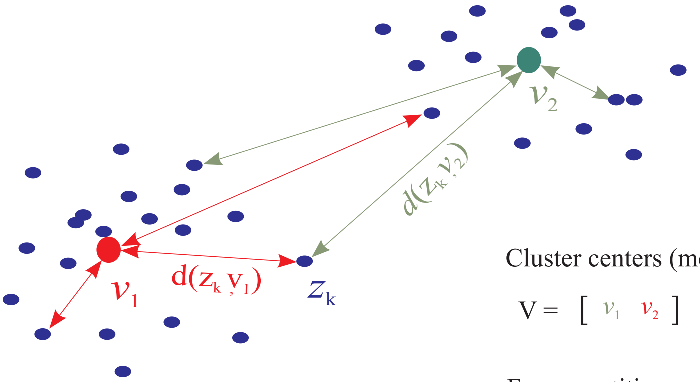
Fuzzy Clustering: Distance



Cluster centers (means):

$V = [\ v_1 \quad v_2\ ]$

Fuzzy Clustering: Partition Matrix



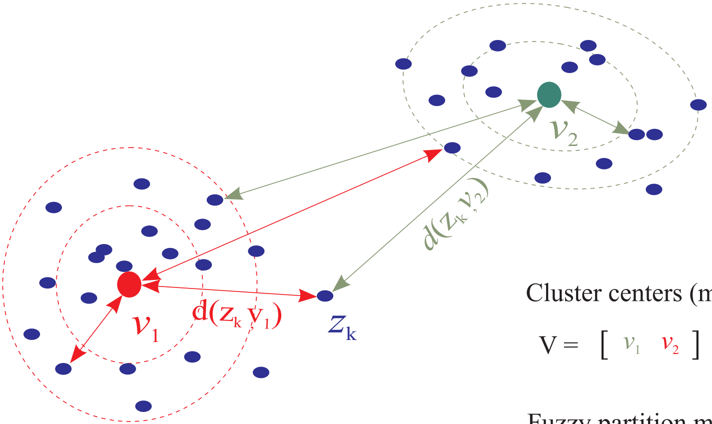
Cluster centers (means):

$$V = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$$

Fuzzy partition matrix:

$$U = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1N} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2N} \end{bmatrix}$$

Fuzzy Clustering: Shapes



Cluster centers (means):

$$V = [\ v_1 \quad v_2\ ]$$

Fuzzy partition matrix:

$$U = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1N} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2N} \end{bmatrix}$$

Fuzzy Clustering Problem

Given the data:

$$\boldsymbol{z}_k = [z_{1k}, z_{2k}, \ldots, z_{nk}]^T \in \mathbb{R}^n, \quad k = 1, \ldots, N$$

Find:

the fuzzy partition matrix:

$$\boldsymbol{U} = \begin{bmatrix} \mu_{11} & \ldots & \mu_{1k} & \ldots & \mu_{1N} \\ \vdots & \ldots & \vdots & \ldots & \vdots \\ \mu_{c1} & \ldots & \mu_{ck} & \ldots & \mu_{cN} \end{bmatrix}$$

and the cluster centers:

$$\boldsymbol{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_c\}, \quad \boldsymbol{v}_i \in \mathbb{R}^n$$

Fuzzy Clustering: an Optimization Approach

Objective function (least-squares criterion):

$$J(Z; V, U, A) = \sum_{i=1}^{c} \sum_{j=1}^{N} \mu_{i,j}^{m} d_{A_i}^{2}(\mathbf{z}_j, \mathbf{v}_i)$$

subject to constraints:

$0 \le \mu_{i,j} \le 1,$	$i = 1, \ldots, c, \; j = 1, \ldots, N$	membership degree
$0 < \sum_{j=1}^{N} \mu_{i,j} < N,$	$i = 1, \ldots, c$	no cluster empty
$\sum_{i=1}^{c} \mu_{i,j} = 1,$	$j = 1, \ldots, N$	total membership

Fuzzy c-Means Algorithm

Repeat:

1. Compute cluster prototypes (means): $v_i = \frac{\sum_{k=1}^{N} \mu_{i,k}^m z_k}{\sum_{k=1}^{N} \mu_{i,k}^m}$

Fuzzy c-Means Algorithm

Repeat:

1. Compute cluster prototypes (means): $v_i = \frac{\sum_{k=1}^{N} \mu_{i,k}^m \mathbf{z}_k}{\sum_{k=1}^{N} \mu_{i,k}^m}$
2. Calculate distances: $d_{ik} = (\mathbf{z}_k - \mathbf{v}_i)^T (\mathbf{z}_k - \mathbf{v}_i)$

Fuzzy c-Means Algorithm

Repeat:

1. Compute cluster prototypes (means): $v_i = \frac{\sum_{k=1}^{N} \mu_{i,k}^m \mathbf{z}_k}{\sum_{k=1}^{N} \mu_{i,k}^m}$
2. Calculate distances: $d_{ik} = (\mathbf{z}_k - \mathbf{v}_i)^T (\mathbf{z}_k - \mathbf{v}_i)$
3. Update partition matrix: $\mu_{ik} = \frac{1}{\sum_{j=1}^{c} (d_{ik}/d_{jk})^{1/(m-1)}}$

until $\|\Delta \mathbf{U}\| < \epsilon$

Distance Measures

- Euclidean norm:

$$d^2(\boldsymbol{z}_j, \boldsymbol{v}_i) = (\boldsymbol{z}_j - \boldsymbol{v}_i)^T(\boldsymbol{z}_j - \boldsymbol{v}_i)$$

- Inner-product norm:

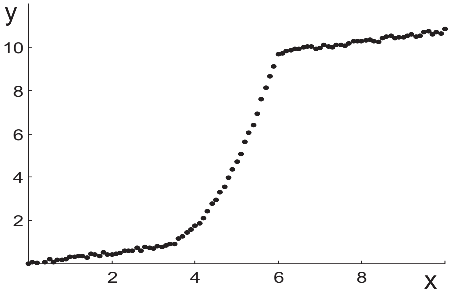
$$d^2_{A_i}(\boldsymbol{z}_j, \boldsymbol{v}_i) = (\boldsymbol{z}_j - \boldsymbol{v}_i)^T A_i(\boldsymbol{z}_j - \boldsymbol{v}_i)$$

- Many other possibilities ...

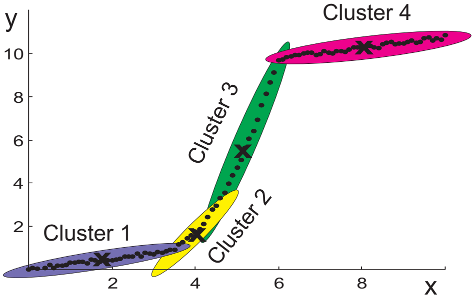
Fuzzy Clustering – Demo

1. Fuzzy c-means

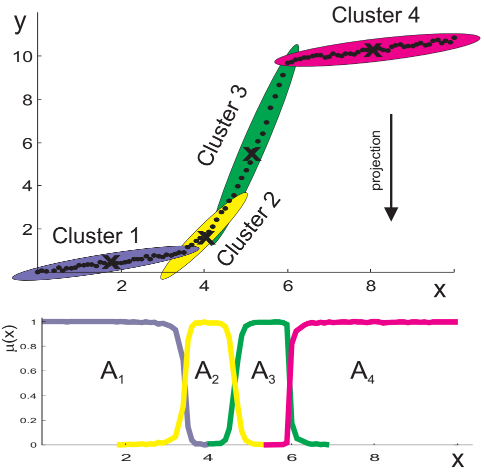
Extraction of Rules by Fuzzy Clustering



Extraction of Rules by Fuzzy Clustering



Extraction of Rules by Fuzzy Clustering



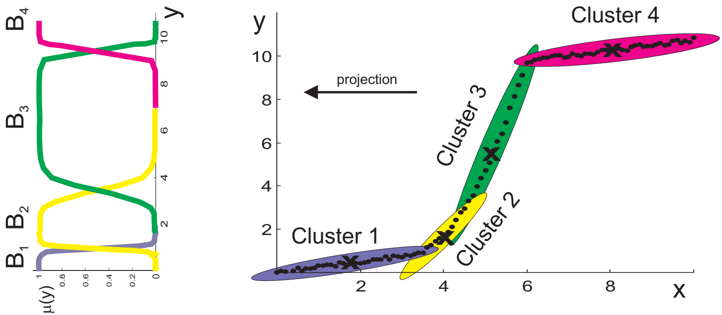
Takagi-Sugeno model

Rule-based description:

If x is A_1 then y = a_1x + b_1
If x is A_2 then y = a_2x + b_2

etc...

Extraction of Rules by Fuzzy Clustering



Rule-based description:

If y is B_1 then $x = a_1 y + b_1$
If y is B_2 then $x = a_2 y + b_2$

etc...

Inverse Takagi-Sugeno model

Rule Extraction – Demo

- Extraction of Takagi–Sugeno rules