BMCS2003 Artificial Intelligence

Dealing with Uncertainty: Fuzzy Expert System

Three ways to deal with uncertainty:

- √ Probability Theory
- √ Certainty Factors
- Fuzzy sets

Objectives:

- Fuzzy set theory / Fuzzy logic

  (Lofti Zadeh, 1965)
- Fuzzy expert system

Fuzzy Logic???

- Fuzzy logic is not logic that is fuzzy,
- but logic that is used to describe fuzziness.
- Fuzzy logic is the theory of fuzzy sets, sets that calibrate vagueness.
- Fuzzy logic is a set of mathematical principles for knowledge representation based on degrees of membership.

Fuzzy Logic

- The digital computing world is built on a structure of Boolean logic applied to binary values
  - one or zero, yes or no, in or out, 0 or 1.
- But this powerful structure is a gross oversimplification of the real world, where many shades of gray exist between black and white

0 Boolean logic **1**

0 Multi-valued logic **1**

Boolean Logic

- Boolean logic uses sharp distinctions. It forces us to draw lines between members of a class and non- members.
- For instance,

  Tom’s height is 181 cm. Is he tall?

- If we draw a line at 180 cm to differentiate between a tall man and a short man,

  David’s height is 179 cm. Is he tall?

Example

- E.g, the possibility that a man 179 cm tall is tall might be set to a value of 0.86.
  - Conclusion: It is likely that the man is tall

[in Boolean logic, 179cm is considered short]

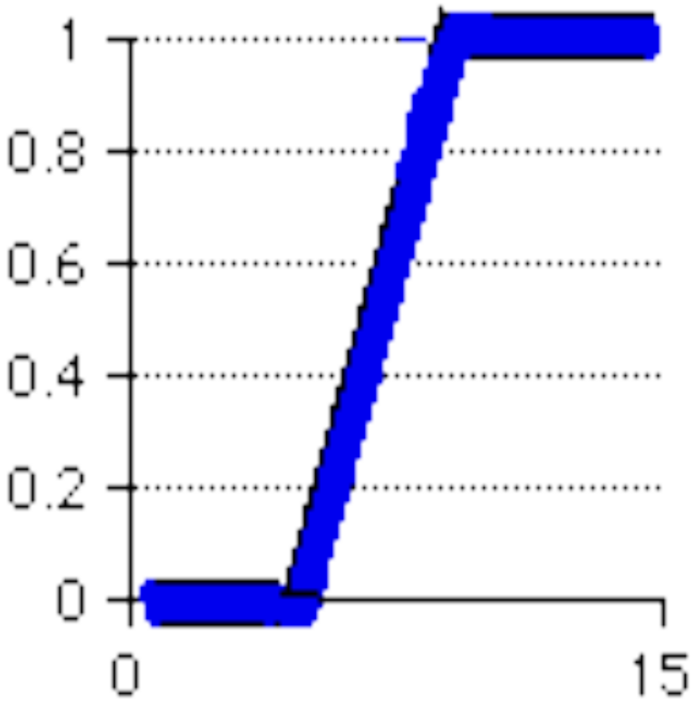
Degree of membership of 'tall men"

Name	Height, cm	Degree of Membership	
		Crisp	*Fuzzy*
Chris	208	1	1.00
Mark	205	1	1.00
John	198	1	0.98
Tom	181	1	0.82
David	179	0	0.78
Mike	172	0	0.24
Bob	167	0	0.15
Steven	158	0	0.06
Bill	155	0	0.01
Peter	152	0	0.00

Fuzzy Sets?

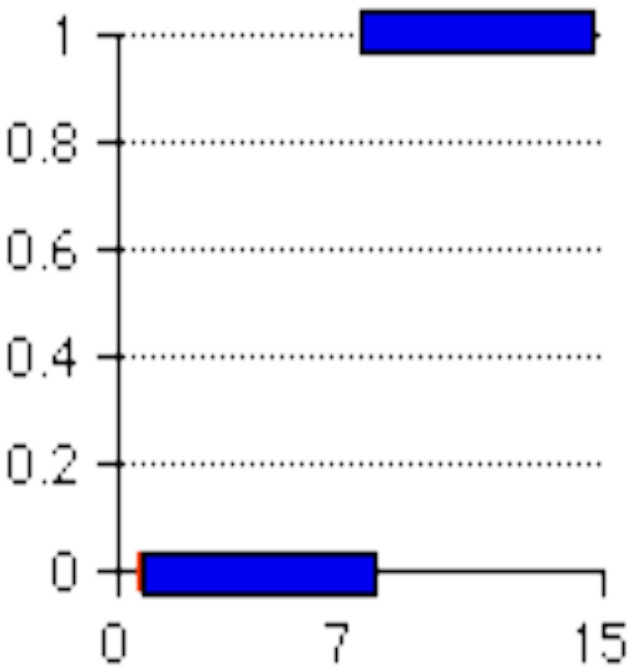
- The concept of **set**
- E.g.
  - *Group* indicates a set of students.
  - If an element is within a Boolean set → 1 (true), otherwise 0 (false)
- Fuzzy set?
  - A set with fuzzy boundaries
  - Elements belongs to a fuzzy set with a certain degree of membership
  - The degree is taken in the interval [0,1]
  - 0 represents absolute falseness and 1 represents absolute truth. [And others are in between 0 and 1]

Fuzzy set v.s. Boolean set



Fuzzy logic

X-axis represents the universe of discourse



Boolean

Context of Universe of Discourse

- Example:

  "Russell's height is 172 cm, is he short or tall"?

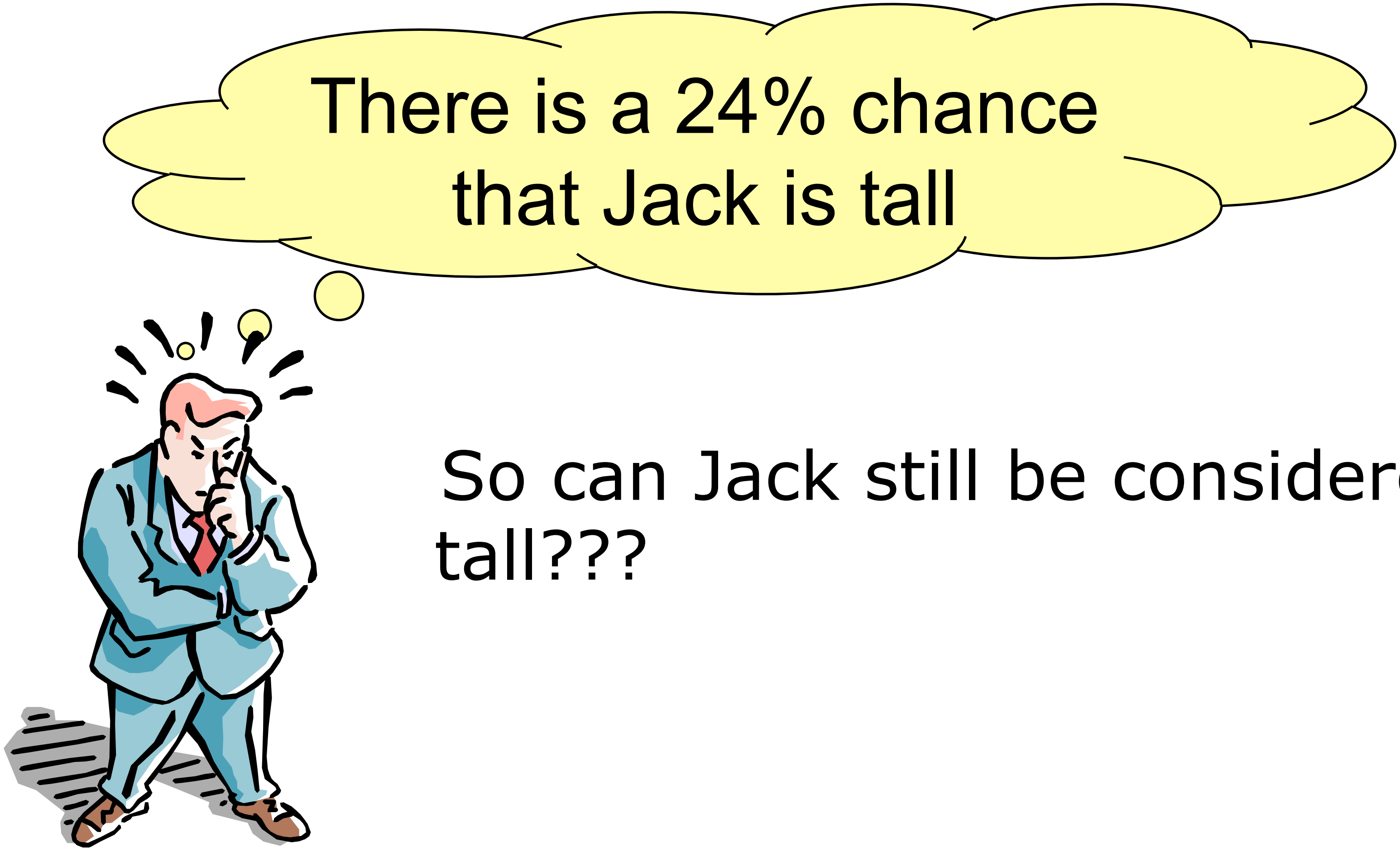
- The context of universe (the range of all possible values) may vary
  - e.g. men's heights vs. women's height

Probability v.s. Fuzzy

- Question:
  - since both ranged from 0 to 1, can
  - Probability = fuzzy logic?

- Consider the statement
  - "Jack is tall" {truth value 0.24}.

In Probability



So can Jack still be considered tall???

In Fuzzy Terminology

Jack's degree of membership in the set of tall people is 0.24

- if we take all the (fuzzy set of) **tall** people and line them up, Jack is positioned 24 of the way to the tallest.
  - Is Jack tall?
- In fuzzy logic,
  - we state this as $\mu_{TALL}(\text{Jack}) = 0.24$, where μ_{TALL} is the membership function.

Fuzzy Logic and Probability

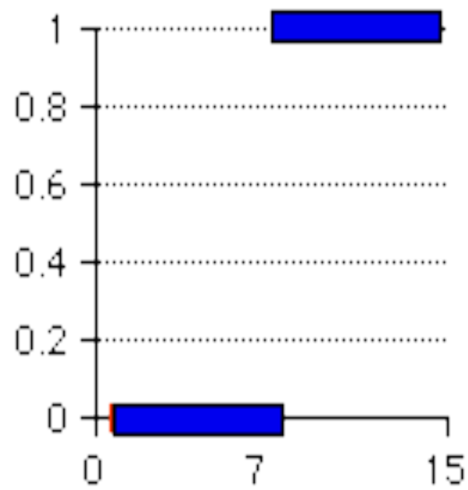
- clearly fuzzy logic ≠ probability
- In fuzzy logic:
  - We know Jack is somewhat tall, and
  - We may know Jack's height exactly (the assertion ‘Jack is tall (0.24)’ measures how well Jack’s height matches the sense of the word ‘tall’).
- In probability:
  - We may still not know whether Jack is tall, and
  - We don't actually know Jack's height.

Membership function in crisp set

- Let X be the universe of discourse and its elements be denoted as x.
- In the classical set theory, crisp set A of X is defined as function $f_A(x)$ called the characteristic function of A

$f_A(x)$: X ® {0, 1},
where

$$f_A(x) = \begin{cases} 1, \text{if } x \in A \\ 0, \text{if } x \notin A \end{cases}$$



Membership function in Fuzzy Set

- In the fuzzy theory, fuzzy set A of universe X is defined by function $\mu_A(x)$ called the membership function of set A

$$\mu_A(x): X \rightarrow [0, 1], \text{ where } \mu_A(x) = 1 \text{ if } x \text{ is totally in } A;$$
$$\mu_A(x) = 0 \text{ if } x \text{ is not in } A;$$
$$0 < \mu_A(x) < 1 \text{ if } x \text{ is partly in } A.$$

Example (People and Tallness)

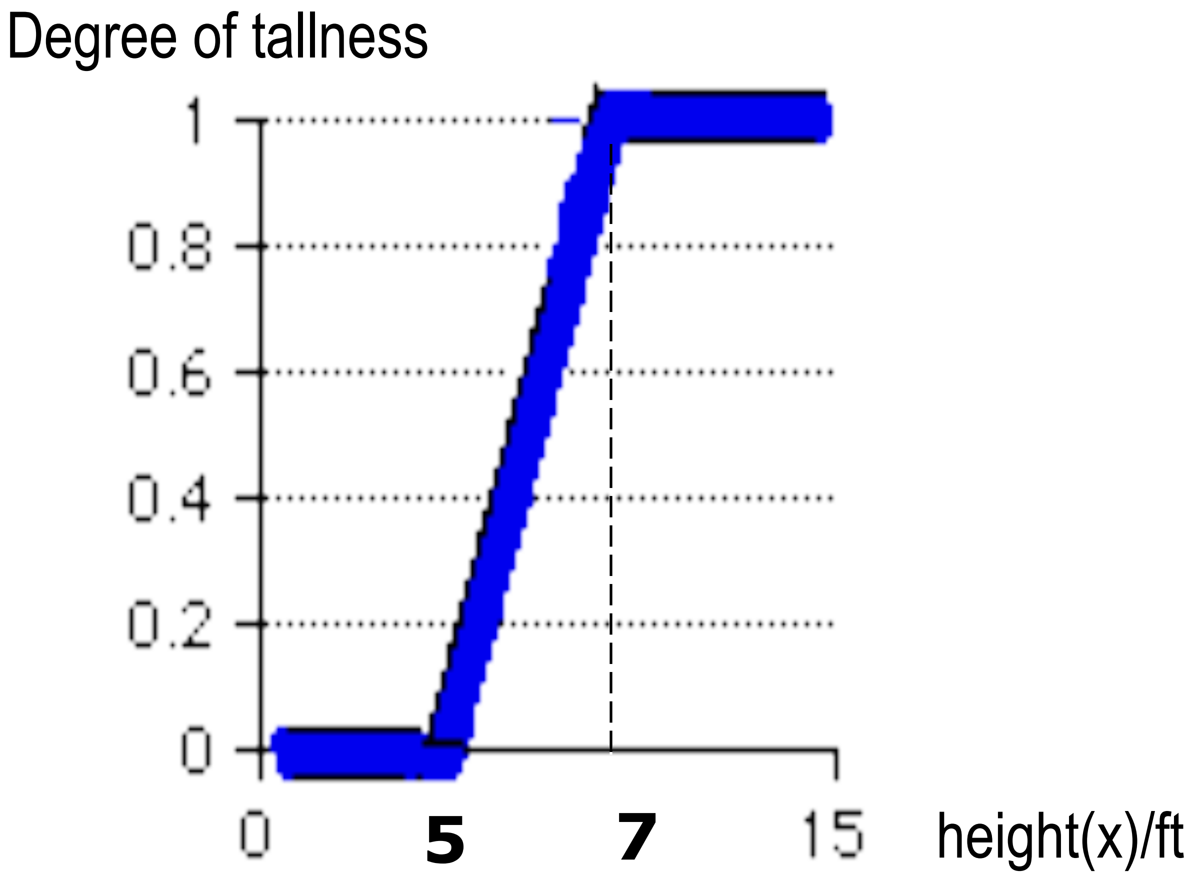
Question: To what degree is person x tall?

The easiest way to do this is with a membership function based on the person's height. And fix a strict changeover point

tall(x) =

{ 0,	if height(x) < 5 ft.,
(height(x)-5ft.)/2ft.,	if 5 ft. <= height (x) <= 7 ft.,
1,	if height(x) > 7 ft. }

Graph Based on Membership Function

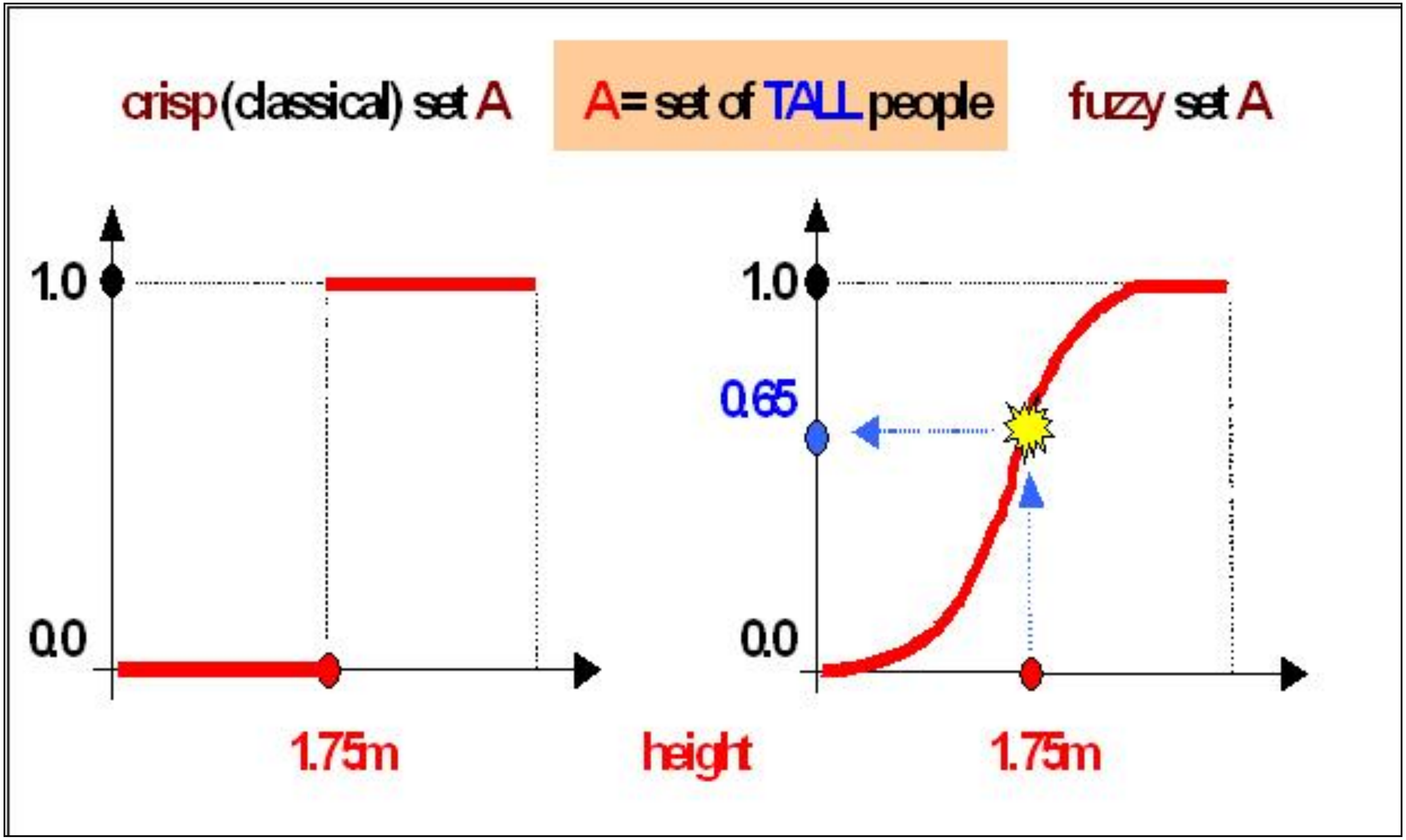


Fuzzy Sets

Person	Height	degree of tallness
Billy	3' 2”	0.00
Yoke	5' 5"	0.21
Drew	5' 9"	0.38
Erik	5' 10”	0.42
Mark	6' 1"	0.54
Kareem	7' 2"	1.00

- Expressions like "A is X" can be interpreted as degrees of truth, e.g., "Drew is TALL" = 0.38.

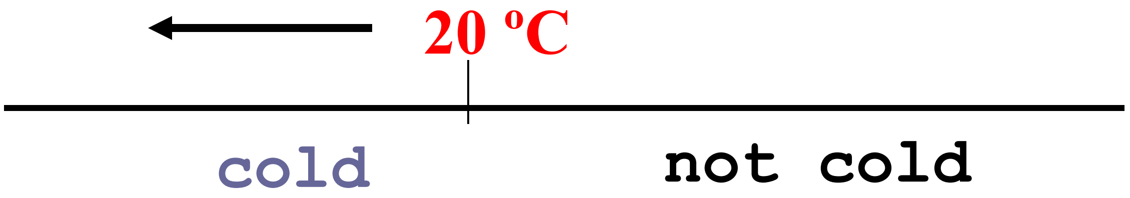
Another Graph Example



Fuzzy Logic – Creating fuzzy set

- In the context of a rule-based system, we might imagine asking 'Is the water cold?'

  Yes / No / very cold / a bit cold
- We might fixing a strict changeover point.



Fuzzy Sets

- For example

Temp		*Function Value*	
− 273	→	1	(cold)
− 40	→	1	(cold)
0	→	0.9	(not quite cold)
5	→	0.7	(on the cold side)
10	→	0.3	(a bit cold)
15	→	0.1	(barely cold)
100	→	0	(not cold)
1000	→	0	(not cold)

The degree of fuzzy sets is usually between 0 to 1.

Creating fuzzy set

- Alternatively, we may make such distinctions ‘fuzzy’, essentially by allowing a whole spectrum of ‘degrees of coldness ’.

temperature	**-10 ºC**		**20 ºC**
	Definitely cold	**cold**	**Definitely not cold**
Membership degree	1		0

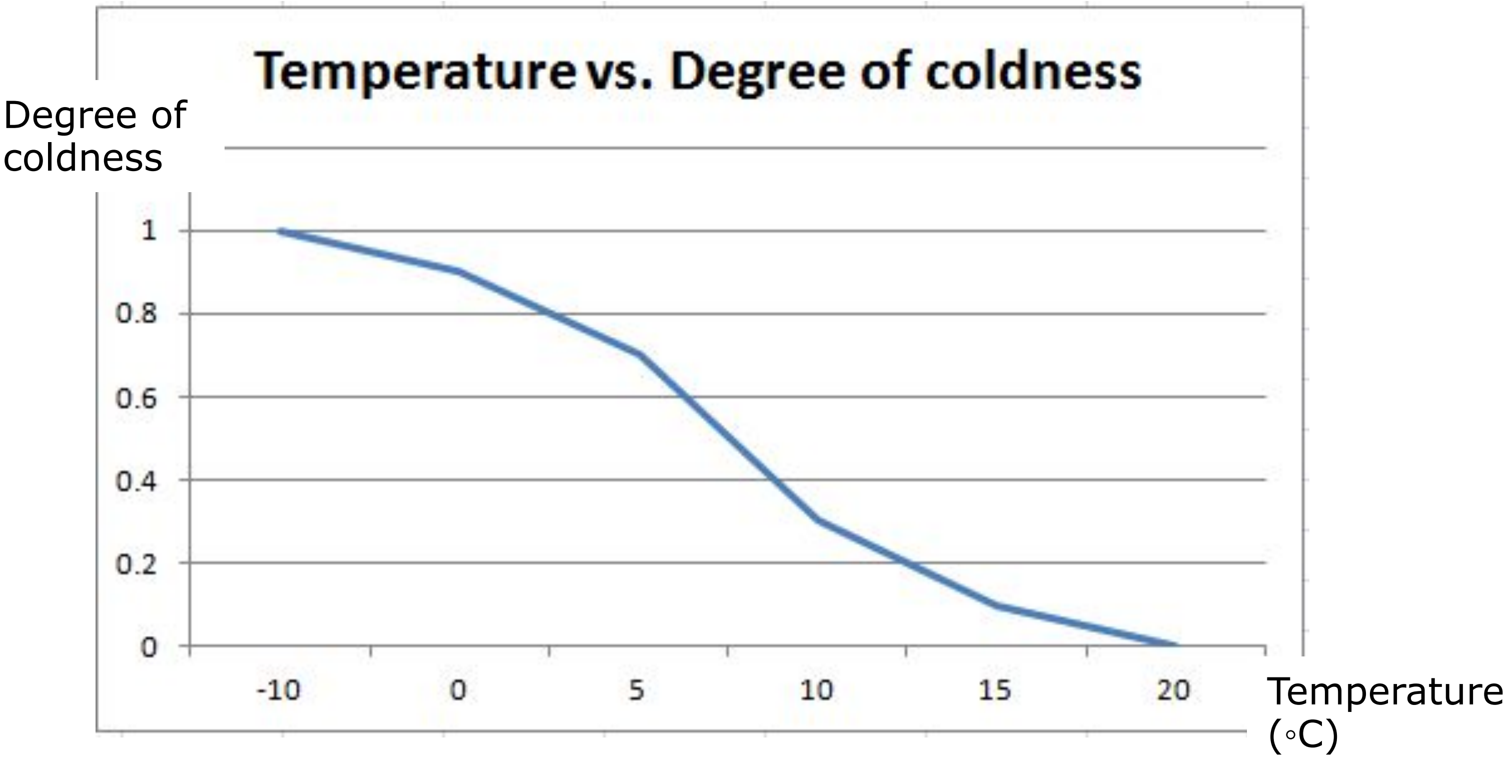
Function (Membership function)

- A property like coldness could be represented by a <u>function.</u> Given a temperature, the function will return a number representing the degree of coldness.

coldness(x) =

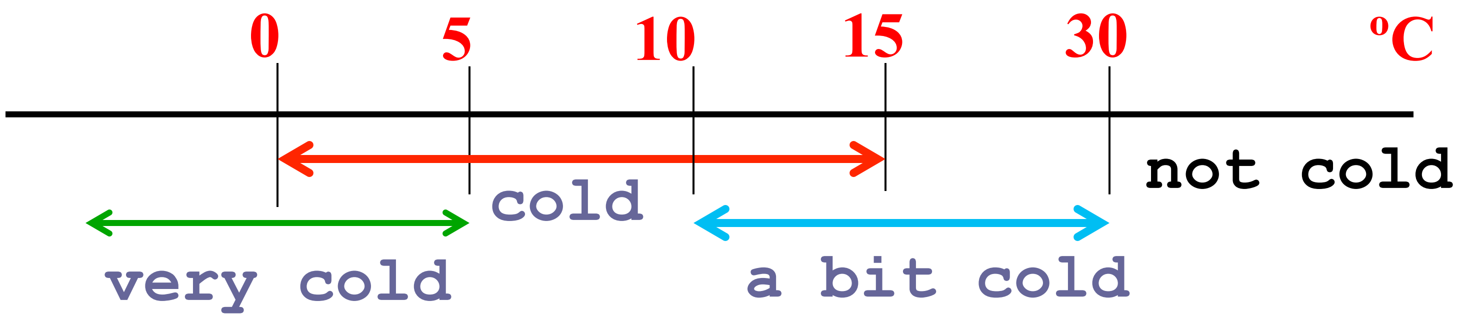
{ 0,	**if temp(x) > 20.,**
f(temp(x))	**if 0 <= temp(x) <= 20.,**
1,	**if temp(x) < -10 }**

The fuzzy membership function



Fuzzy sets

- We can also define more fuzzy sets to describe different levels of coldness.



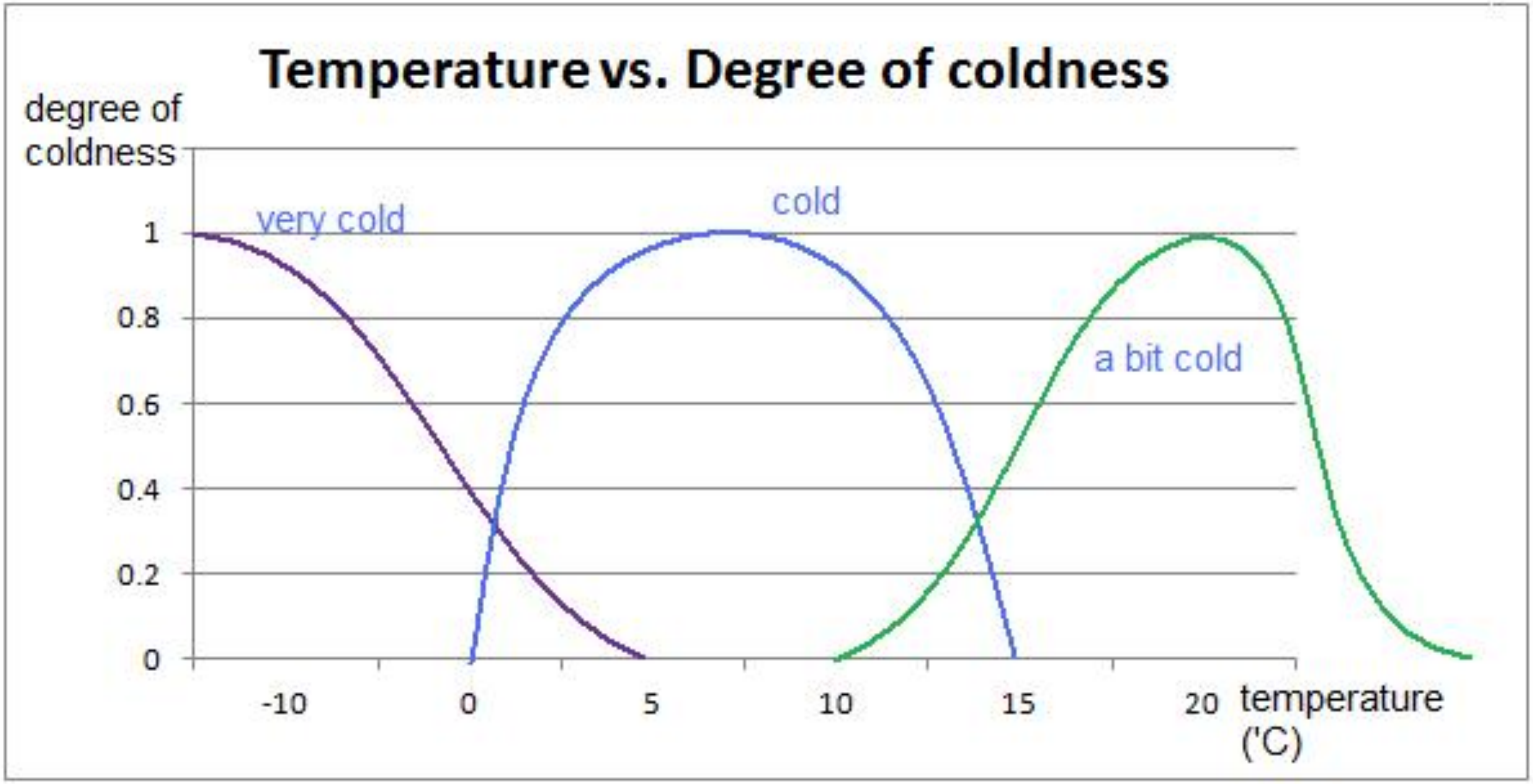
Three different fuzzy sets

very_cold(x) =	
{ 0,	if temp(x) > 5,
f(temp(x))	if -20 <= temp(x) <= 5,
1,	if temp(x) < -20 }

cold(x) =	
{ 0,	if temp(x) > 15 or temp(x) < 0,
f(temp(x))	if 0<= temp(x) <= 15
1,	if temp(x) = 7.5}

a_bit_cold(x) =	
{ 0,	if temp(x) < 10 or temp(x) >30,
f(temp(x))	if 10<= temp(x) <= 30.,
1,	if temp(x) = 20 }

Fuzzy membership functions



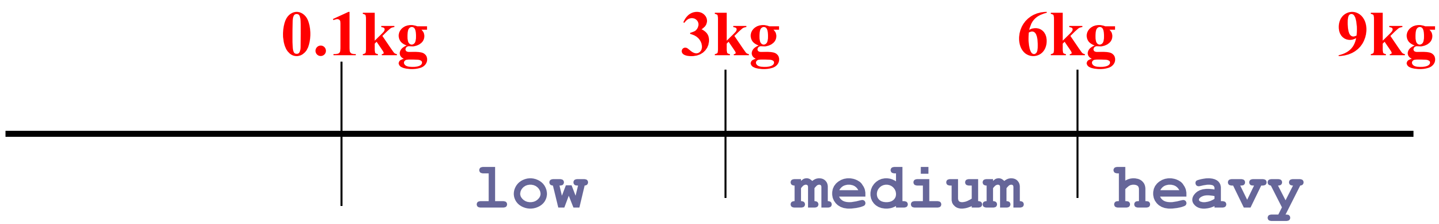
Exercise

- A washing machine can load up to 9kg clothes, and minimum 0.1kg clothes to perform its washing task. It has 3 options to weight the clothes: *low, medium, high*

  Create:
  - A fuzzy set to demonstrate the spectrum of ‘degrees of heaviness ’.
  - a membership function of the clothes heaviness,
  - a graph to represent the membership function.

Answer

Fuzzy set:



It is useful to ensure the range of each set is overlapping with another.

Answer

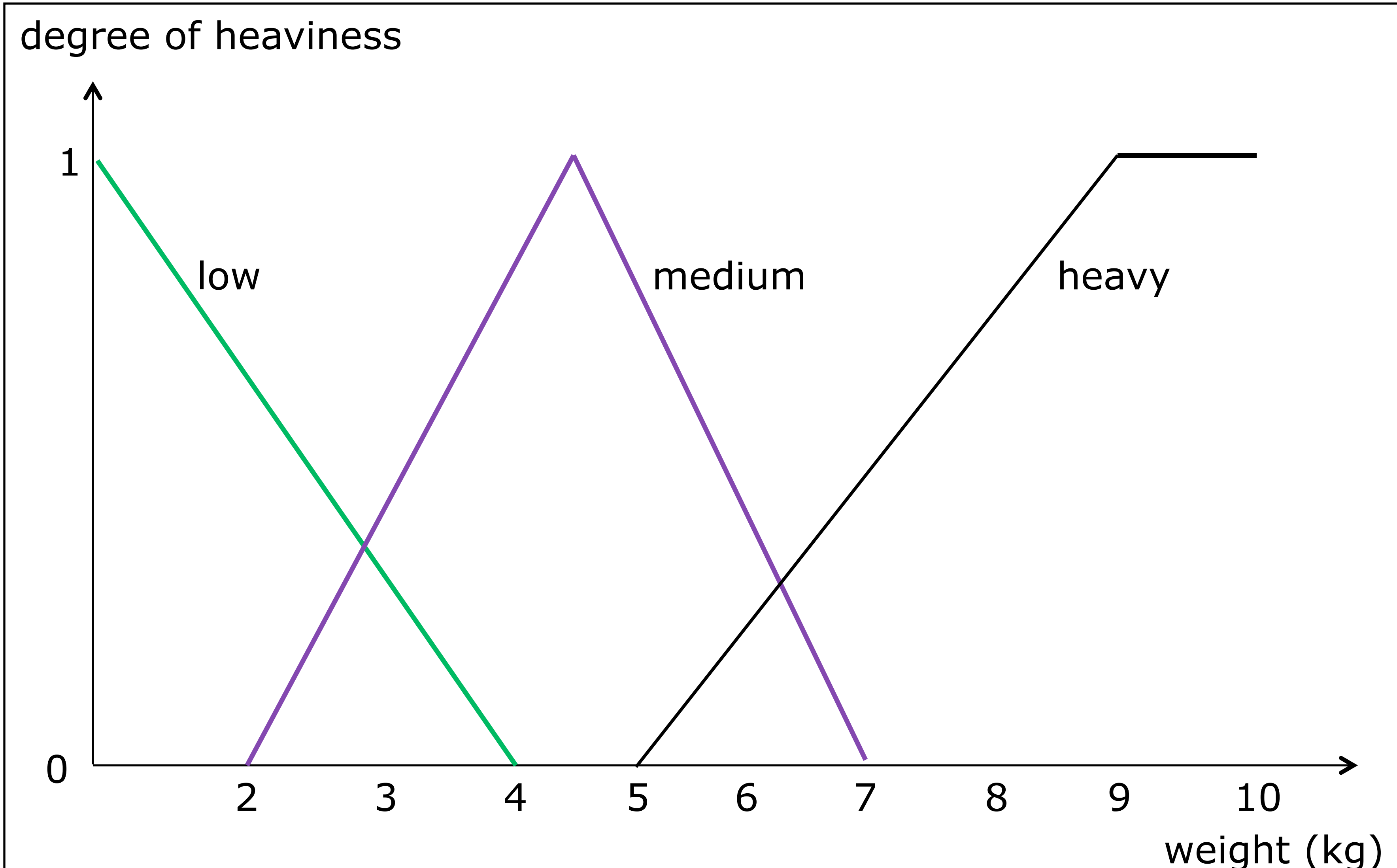
Membership functions:

$$
\text{low}(x) = \begin{cases} 0, & \text{if } \text{weight}(x) >= 4, \\ f(\text{weight}(x)) & \text{if } 0.1 < \text{weight}(x) < 4, \\ 1, & \text{if } \text{weight}(x) <= 0 \end{cases}
$$

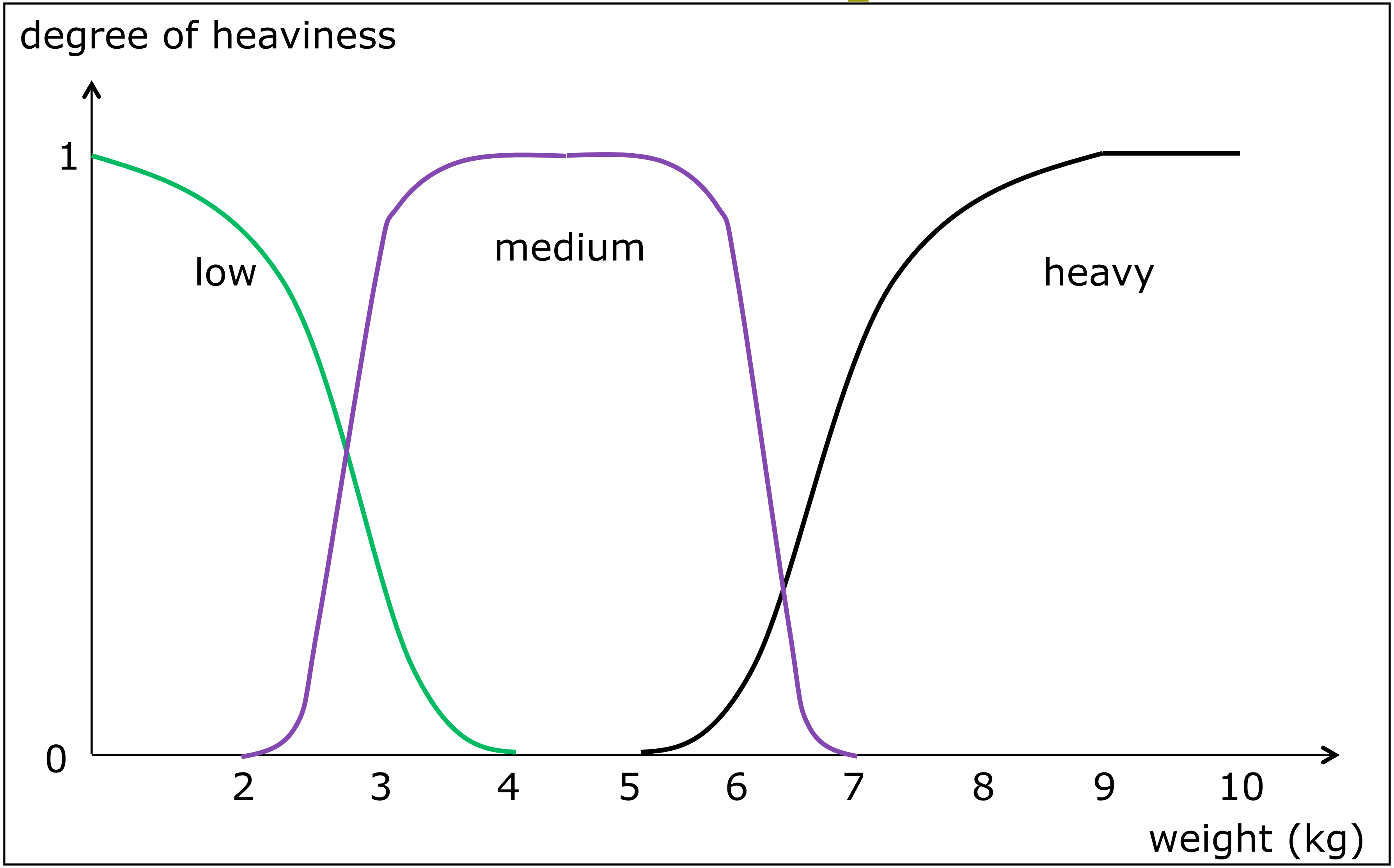
$$
\text{medium}(x) = \begin{cases} 0, & \text{if } \text{weight}(x) <= 2 \text{ or } \text{weight}(x) >= 7, \\ f(\text{weight}(x)) & \text{if } 2 < \text{weight}(x) < 7, \\ 1, & \text{if } \text{weight}(x) = 4.5 \end{cases}
$$

$$
\text{heavy}(x) = \begin{cases} 0, & \text{if } \text{weight}(x) <= 5, \\ f(\text{weight}(x)) & \text{if } 5 < \text{weight}(x) < 9, \\ 1, & \text{if } \text{weight}(x) >= 9 \end{cases}
$$

Membership Function



Alternative Membership Function



Creating Membership function

- Let's imagine that we have a function hot which, when given a temperature, returns the degree of hotness. Then the function very_hot will have the meaning suggested by the name.

$$very\_hot(T) = (hot(T))^{2}$$

- Similarly

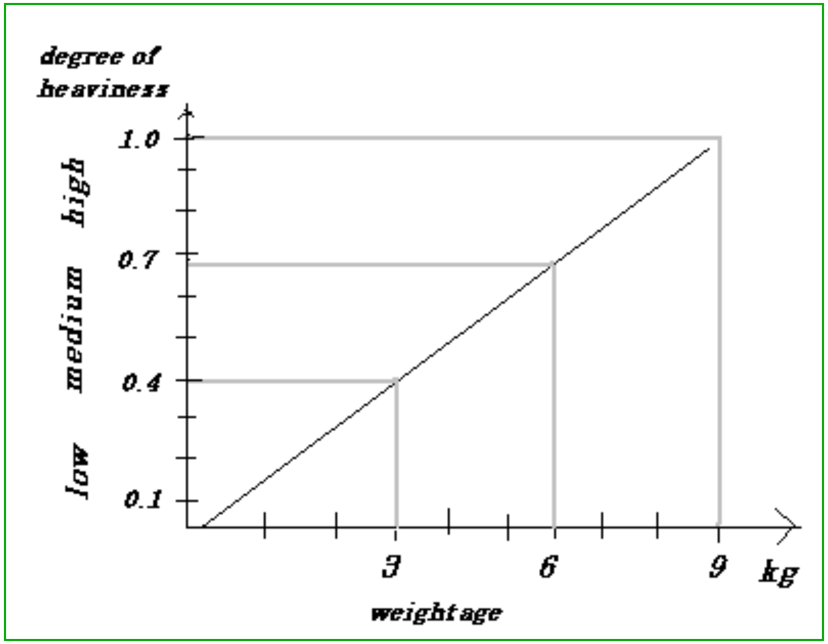
$$not\_hot(T) = 1 - hot(T)$$

and

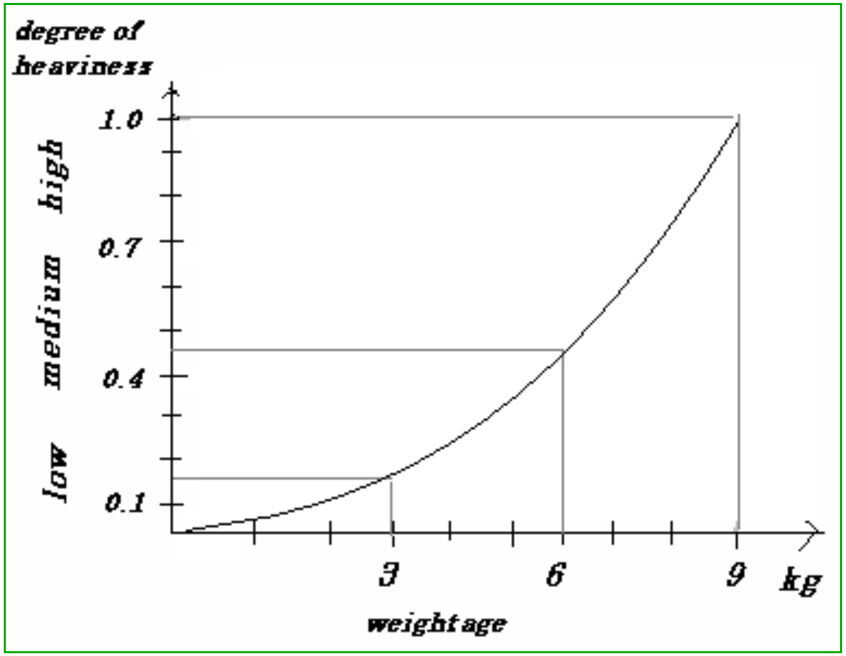
$$somewhat\_hot(T) = hot(T)^{0.5}$$

“Very” Membership Function?

- If the membership function for
  Very_Heavy(W) = Heavy(W)^2



Heavy(W)



Very_Heavy(W)

Drawing multiple fuzzy sets

"Medium" and "Heavy" Membership functions:

medium(x) =

{ 0, if weight(x) < 3,
f(weight(x)) if 3 <= weight(x)<6,
1, if 6 <=weight(x) <9 }

Heavy(x) =

{ 0, if weight(x) < 4.7
f(weight(x)) if 4.7<= weight(x)<=8,
1, if weight(x) >8 }

- If the membership functions for

  Very_Heavy(x) = Heavy(x)^2

- and

  Low(x) = 1- Heavy(x)

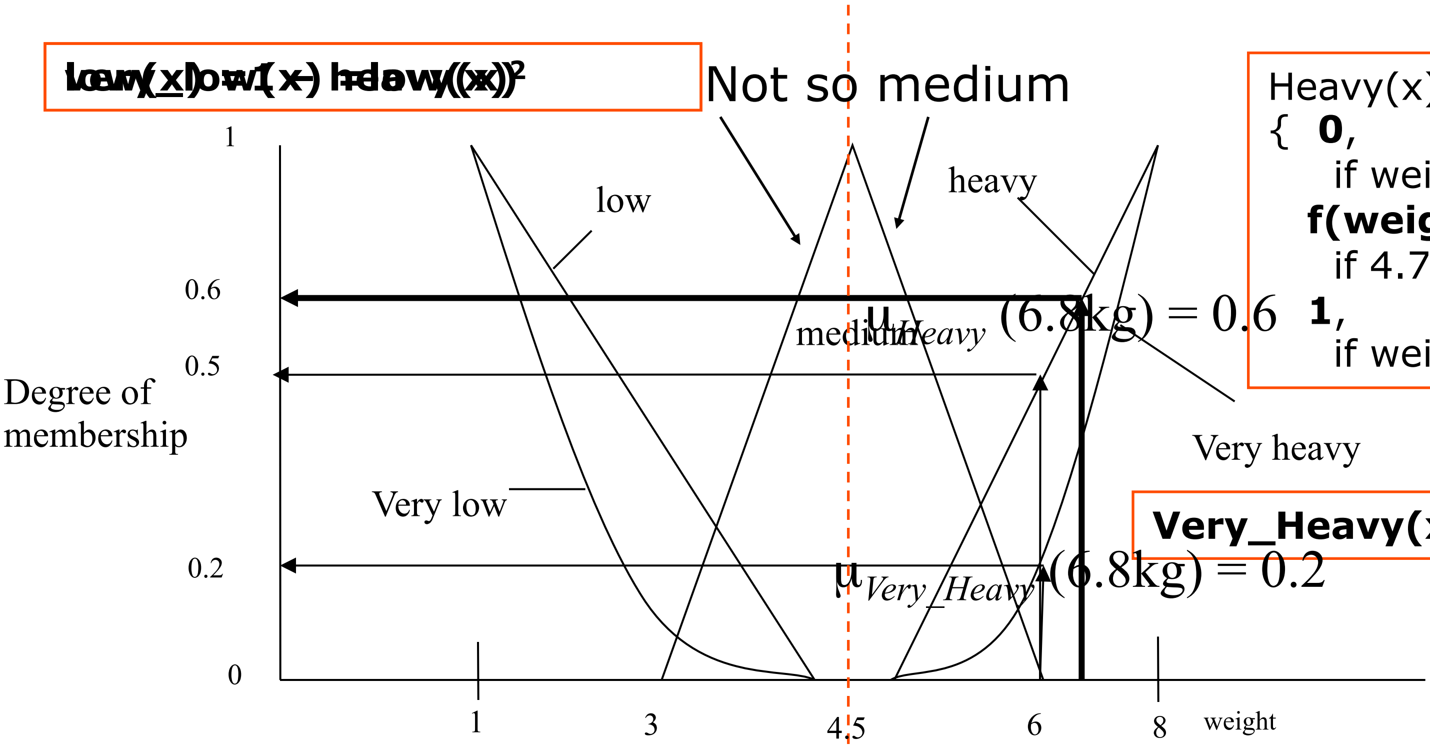
- and

  Very_Low(x) = Low(x)^2

Designing Graph

low(x) = 1 - heavy(x)

Heavy(x) =
{ **0**,
if weight(x) < 4.7
f(weight(x))
if 4.7<= weight(x)<=8,
1,
if weight(x) >8 }



Very_Heavy(x) =Heavy(X)²

Fuzzy Sets - Limitation

- And there are other mathematical formulas which can be given sensible interpretations, too.
- But: people do not actually think in fuzzy logic either. It has been usefully applied, however, in situations where input is provided by sensors rather than people (e.g. cameras and washing machines).

Fuzzy Sets - Advantage

- The benefit of all this is that the reasoning process of our expert system can use the idea of ‘coldness’ (knowledge), rather than the underlying idea of temperature (data). The rules in an expert system may refer to coldness; the data being input may be actual temperatures; fuzzy logic is a means to relate the two.
- Rule:

  IF weather is cold

  THEN reduce coldness

A fuzzy set will be set to determine “cold” weather

Fuzzy rules

- In 1973, Lotfi Zadeh published his second most influential paper. This paper outlined a new approach to analysis of complex systems, in which Zadeh suggested capturing human knowledge in fuzzy rules.

Fuzzy Rules???

- A fuzzy rule can be defined as a conditional statement in the form:

- IF x is A
- THEN y is B

- where x and y are linguistic variables; and A and B are linguistic values determined by fuzzy sets on the universe of discourses X and Y, respectively

Classic v.s. fuzzy rules?

- A classic IF-THEN rule uses binary logic, for example,

Rule: 1		**Rule: 2**	
IF	**speed is > 100**	**IF**	**speed is < 40**
THEN	**stopping_distance is <u>long</u>**	**THEN**	**stopping_distance is <u>short</u>**

The variable *speed* can have any numerical value between 0 and 220 km/h, but the linguistic variable *stopping_distance* can take either value ***long*** or ***short***. In other words, classical rules are expressed in the black-and-white language of Boolean logic.

We can also represent the stopping distance rules in a fuzzy form:

Rule: 1	Rule: 2
IF speed is fast	IF speed is slow
THEN stopping_distance is long	THEN stopping_distance is short

In fuzzy rules, the linguistic variable *speed* also has the range (the universe of discourse) between 0 and 220 km/h, but this range includes fuzzy sets, such as *slow*, *medium* and *fast*.

The universe of discourse of the linguistic variable *stopping_distance* can be between 0 and 300 m and may include such fuzzy sets as *short*, *medium* and *long*.

How to reason with fuzzy rules?

- Fuzzy rules relate fuzzy sets.
- In a fuzzy system, all rules fire to some extent, or in other words they fire partially. If the antecedent is true to some degree of membership, then the consequent is also true to that same degree.
- i.e.

Rule:

IF speed is fast ← Membership degree 0.4

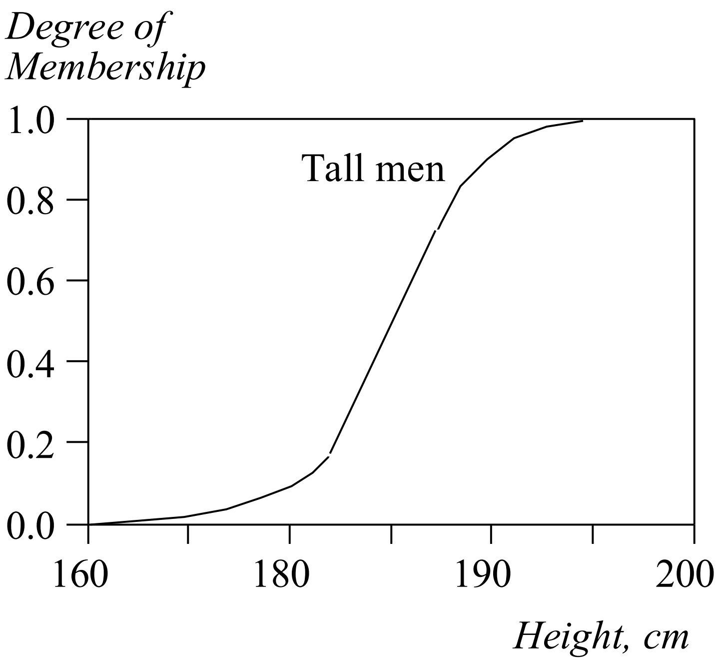
THEN stopping_distance is long ← Membership degree 0.4

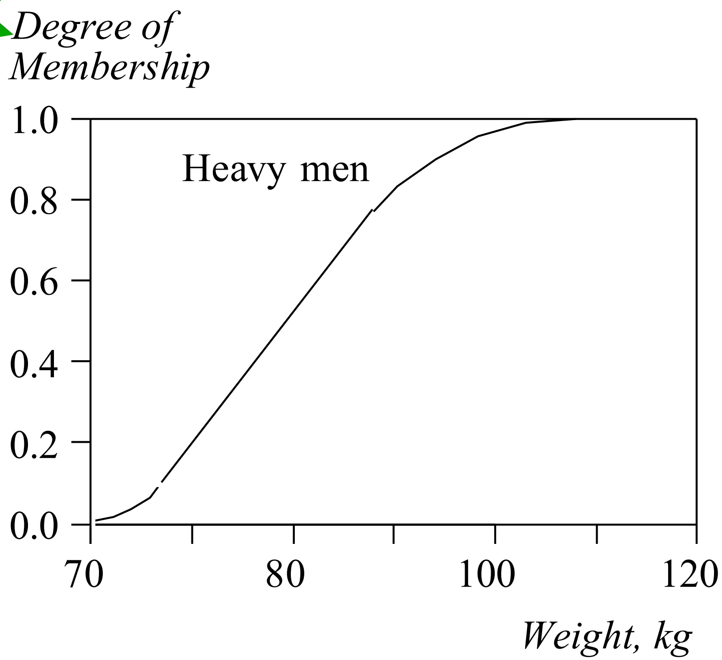
Example

- Fuzzy sets of tall and heavy men

IF height is tall
THEN weight is heavy

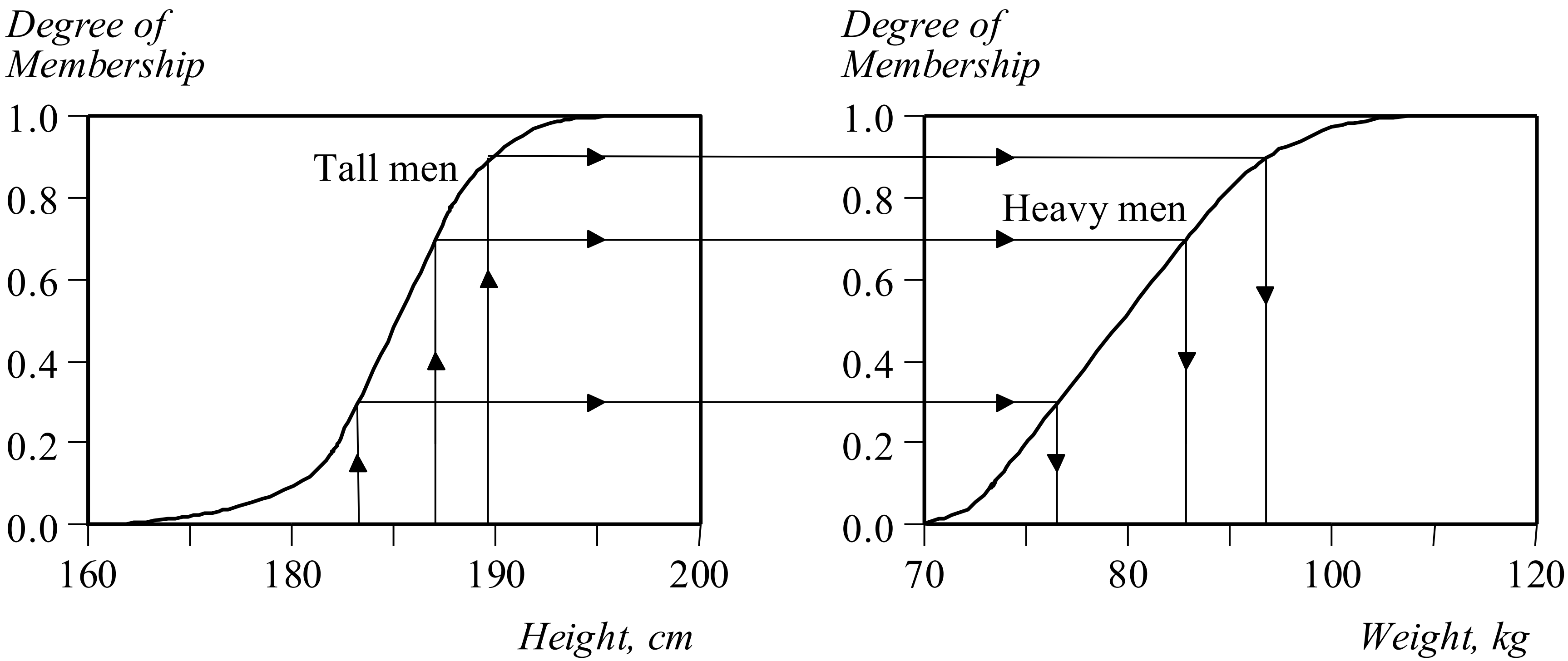
The model is based on a relationship between a man's height and his weight:





Fuzzy inference

- The fuzzy sets of height and weight provide the basis for a weight estimation model.



Fuzzy inference: Monotonic Selection

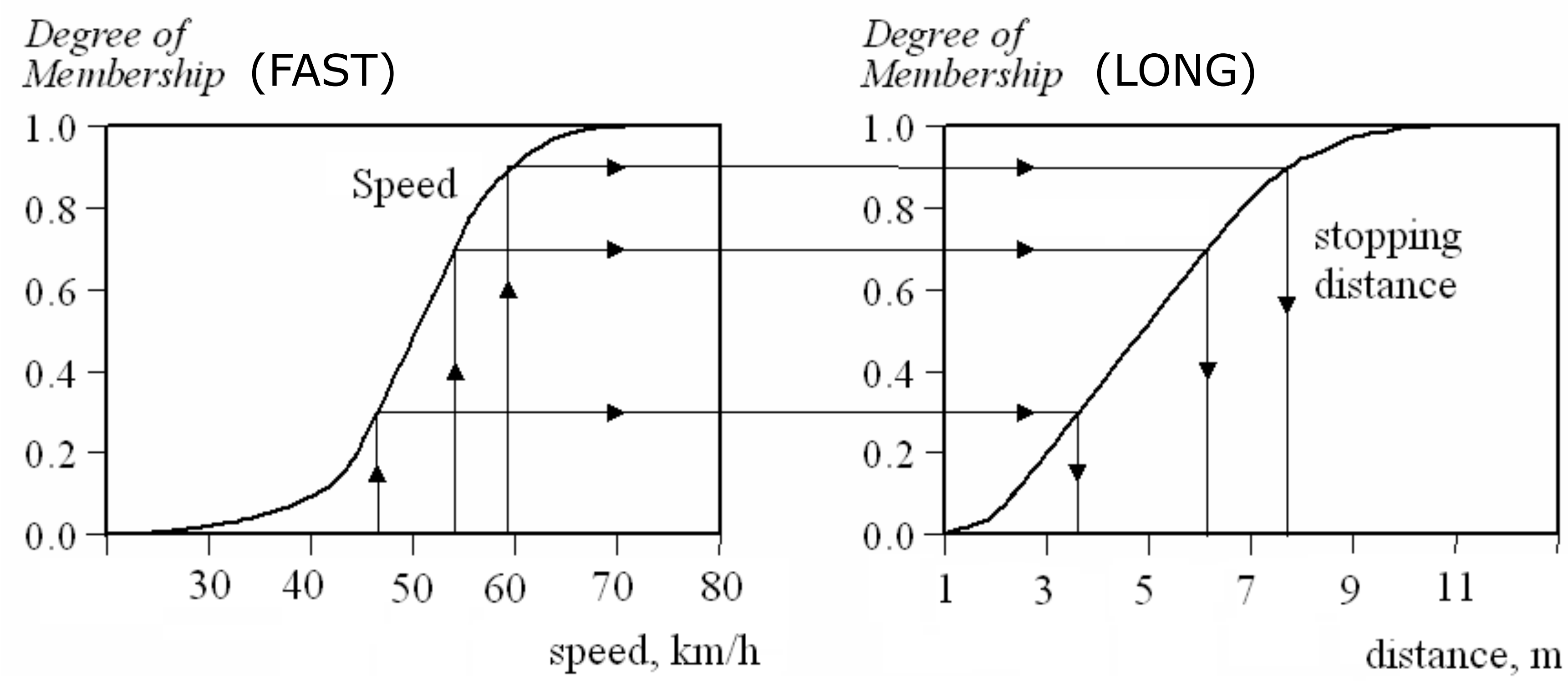
- The value of the output or a truth membership grade of the rule consequent can be estimated directly from a corresponding truth membership grade in the antecedent.
- This form of fuzzy inference uses a method called **monotonic selection**.

Exercise

Rule: 1	**Rule: 2**
IF speed is fast	**IF speed is slow**
THEN stopping_distance is <u>long</u>	**THEN stopping_distance is <u>short</u>**

- Based on the fuzzy rules above, use monotonic selection method to draw a stopping_distance estimation model.

Answer



Degree of Membership (FAST)
1.0
0.8
0.6
0.4
0.2
0.0
Speed
30
40
50
60
70
80
speed, km/h
Degree of Membership (LONG)
1.0
0.8
0.6
0.4
0.2
0.0
stopping distance
1
3
5
7
9
11
distance, m

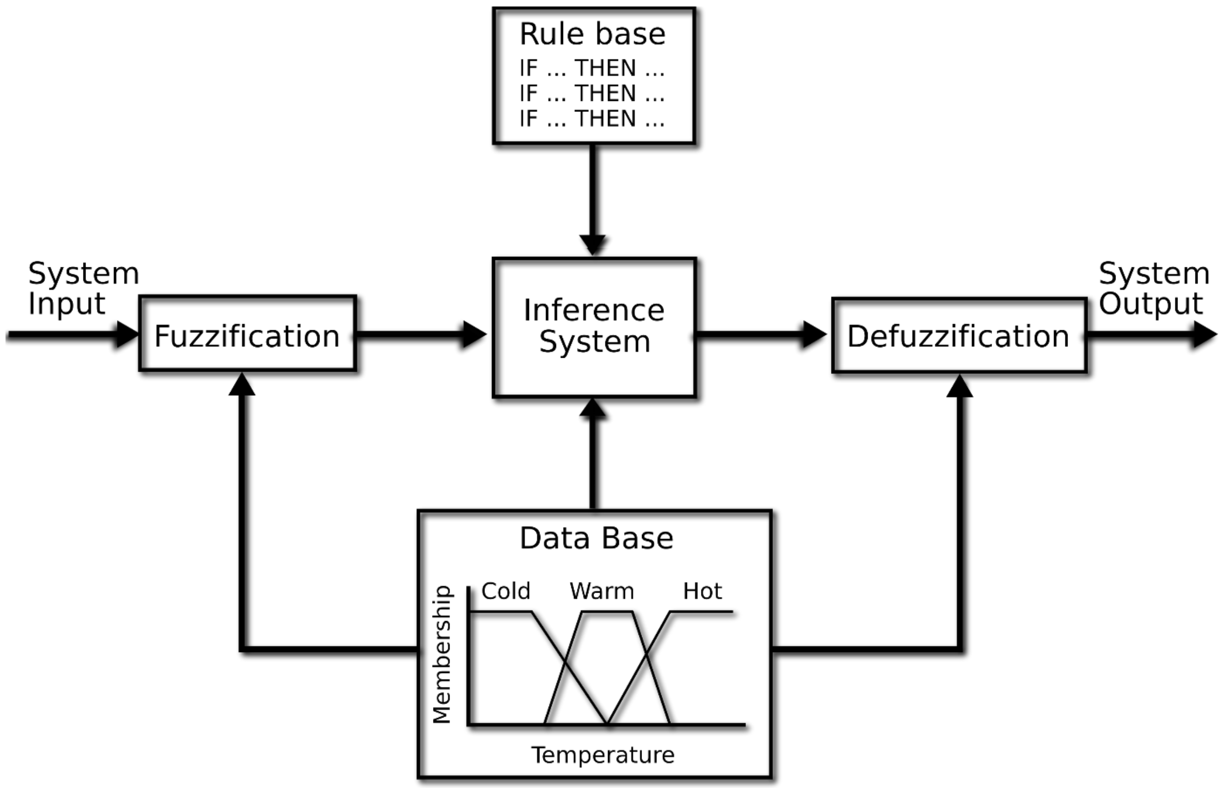
Question

Suggest the fuzzy inference system (monotonic selection) for Rule 2

- **IF speed is slow**
- **THEN stopping_distance is short**

Fuzzy Inference System

- Mamdani fuzzy inference
- Sugeno fuzzy inference



Fuzzy inference

The most commonly used fuzzy inference technique is the so-called Mamdani method. In 1975, Professor **Ebrahim Mamdani** of London University built one of the first fuzzy systems to control a steam engine and boiler combination. He applied a set of fuzzy rules supplied by experienced human operators.

Mamdani fuzzy inference

- The Mamdani-style fuzzy inference process is performed in four steps:
  - fuzzification of the input variables,
  - rule evaluation;
  - aggregation of the rule outputs, and finally
  - defuzzification.

We examine a simple two-input one-output problem that includes three rules:

Rule: 1
IF x **is** $A3$
OR y **is** $B1$
THEN z **is** $C1$

Rule: 2
IF x **is** $A2$
AND y **is** $B2$
THEN z **is** $C2$

Rule: 3
IF x **is** $A1$
THEN z **is** $C3$

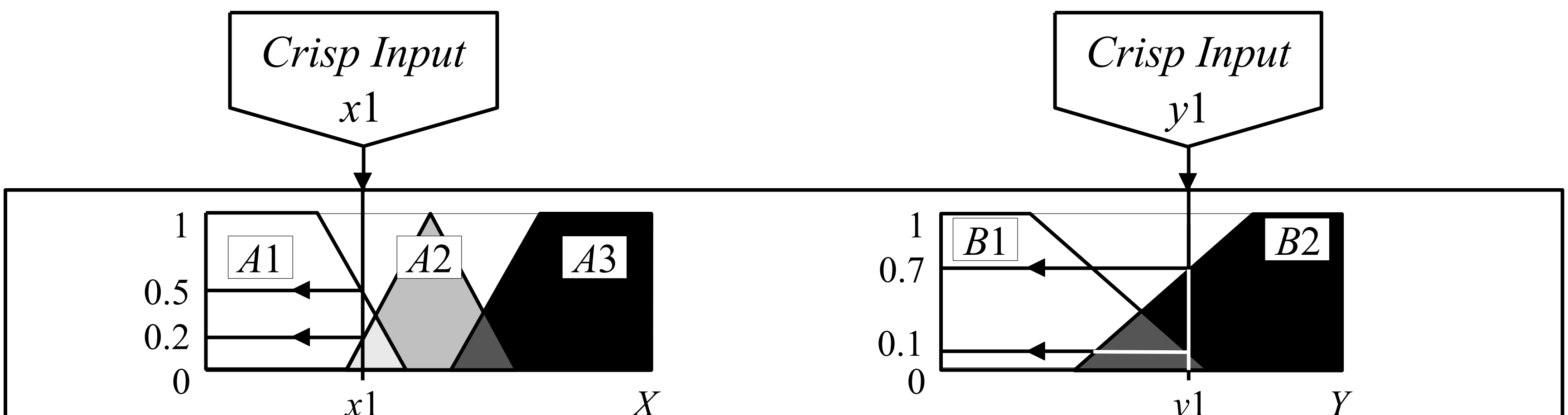
Rule: 1
IF ***project_funding*** **is** ***adequate***
OR ***project_staffing*** **is** ***small***
THEN ***risk*** **is** ***low***

Rule: 2
IF ***project_funding*** **is** ***marginal***
AND ***project_staffing*** **is** ***large***
THEN ***risk*** **is** ***normal***

Rule: 3
IF ***project_funding*** **is** ***inadequate***
THEN ***risk*** **is** ***high***

Step 1: Fuzzification

The first step is to take the crisp inputs, $x1$ and $y1$ (*project funding* and *project staffing*), and determine the degree to which these inputs belong to each of the appropriate fuzzy sets.



$\mu_{(x = A1)} = 0.5$
$\mu_{(x = A2)} = 0.2$

$\mu_{(y = B1)} = 0.1$
$\mu_{(y = B2)} = 0.7$

Step 2: Rule Evaluation

The second step is to take the fuzzified inputs, $\mu_{(x=A1)} = 0.5$, $\mu_{(x=A2)} = 0.2$, $\mu_{(y=B1)} = 0.1$ and $\mu_{(y=B2)} = 0.7$, and apply them to the antecedents of the fuzzy rules. If a given fuzzy rule has multiple antecedents, the fuzzy operator (AND or OR) is used to obtain a single number that represents the result of the antecedent evaluation. This number (the truth value) is then applied to the consequent membership function.

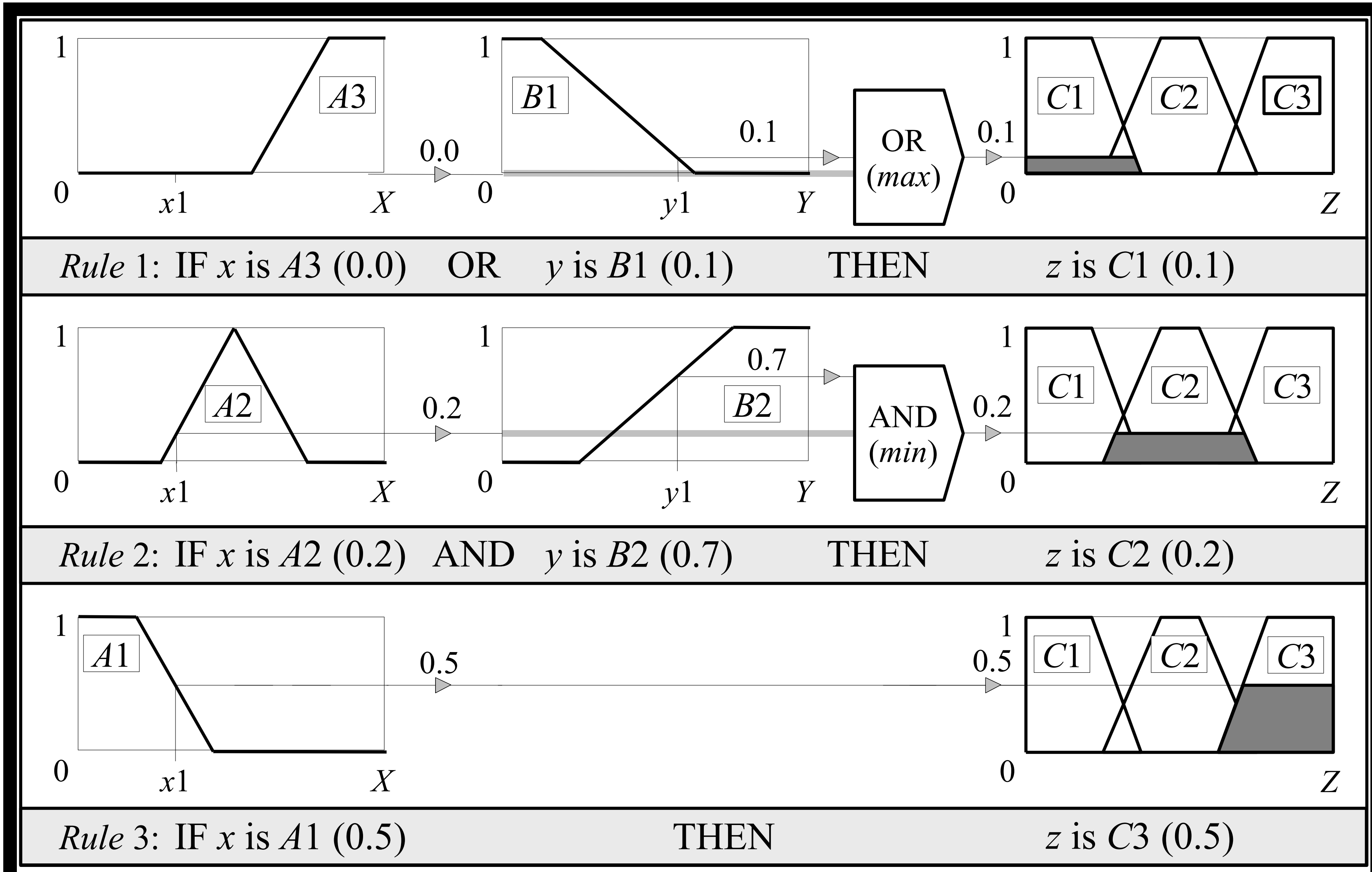
To evaluate the disjunction of the rule antecedents, we use the **OR fuzzy operation**. Typically, fuzzy expert systems make use of the classical fuzzy operation **union**:

$$\mu_A \cup_B(x) = \mathbf{\mathit{max}}\ [\mu_A(x),\ \mu_B(x)]$$

Similarly, in order to evaluate the conjunction of the rule antecedents, we apply the **AND fuzzy operation intersection**:

$$\mu_A \cap_B(x) = \mathbf{\mathit{min}}\ [\mu_A(x),\ \mu_B(x)]$$

Mamdani-style rule evaluation

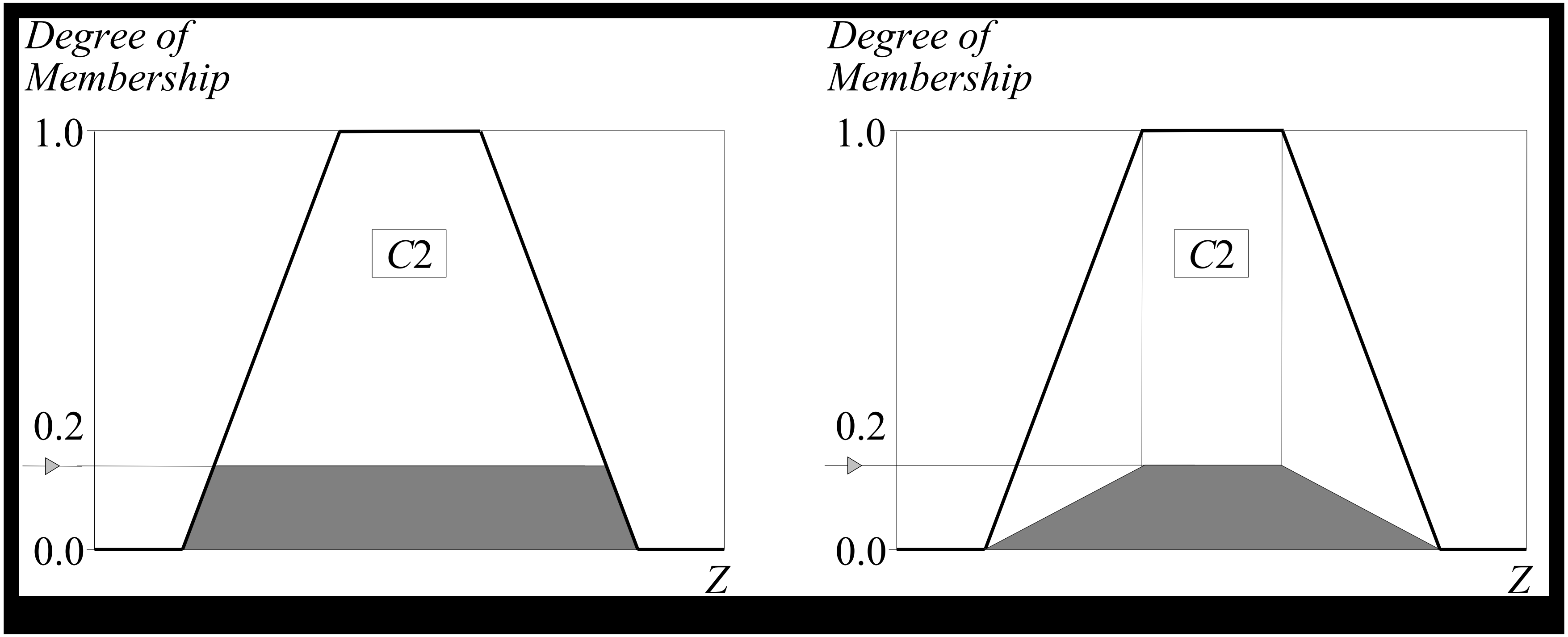


Now the result of the antecedent evaluation can be applied to the membership function of the consequent.

- The most common method of correlating the rule consequent with the truth value of the rule antecedent is to cut the consequent membership function at the level of the antecedent truth. This method is called **clipping**. Since the top of the membership function is sliced, the clipped fuzzy set loses some information. However, clipping is still often preferred because it involves less complex and faster mathematics, and generates an aggregated output surface that is easier to defuzzify.

- While clipping is a frequently used method, **scaling** offers a better approach for preserving the original shape of the fuzzy set. The original membership function of the rule consequent is adjusted by multiplying all its membership degrees by the truth value of the rule antecedent. This method, which generally loses less information, can be very useful in fuzzy expert systems.

Clipped and scaled membership functions

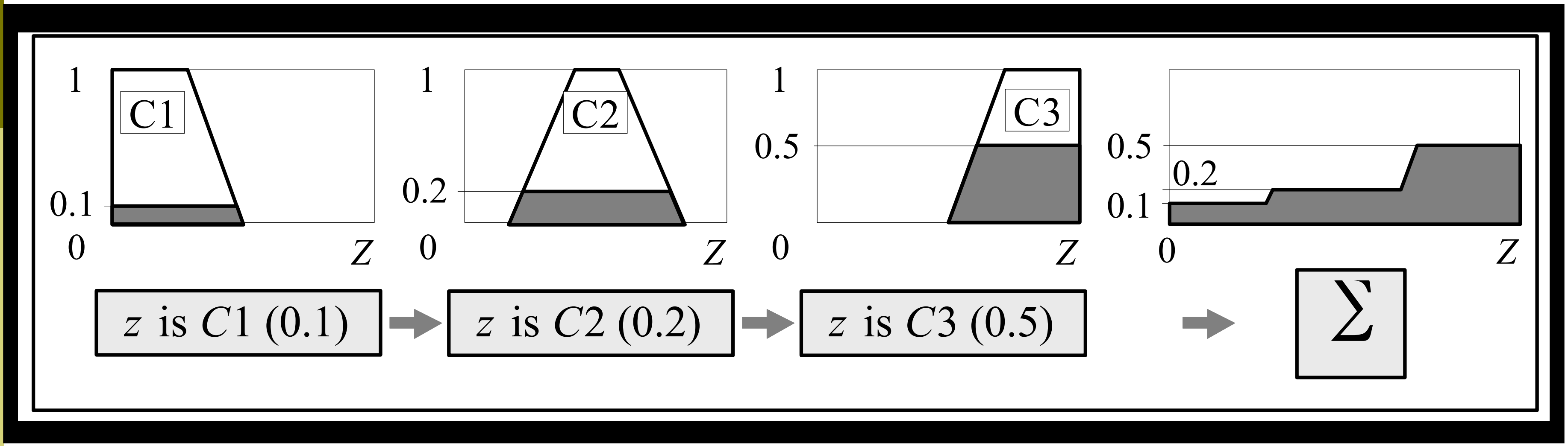


Step 3: Aggregation of the rule outputs

Aggregation is the process of unification of the outputs of all rules. We take the membership functions of all rule consequents previously clipped or scaled and combine them into a single fuzzy set.

The input of the aggregation process is the list of clipped or scaled consequent membership functions, and the output is one fuzzy set for each output variable.

Aggregation of the rule outputs



Step 4: Defuzzification

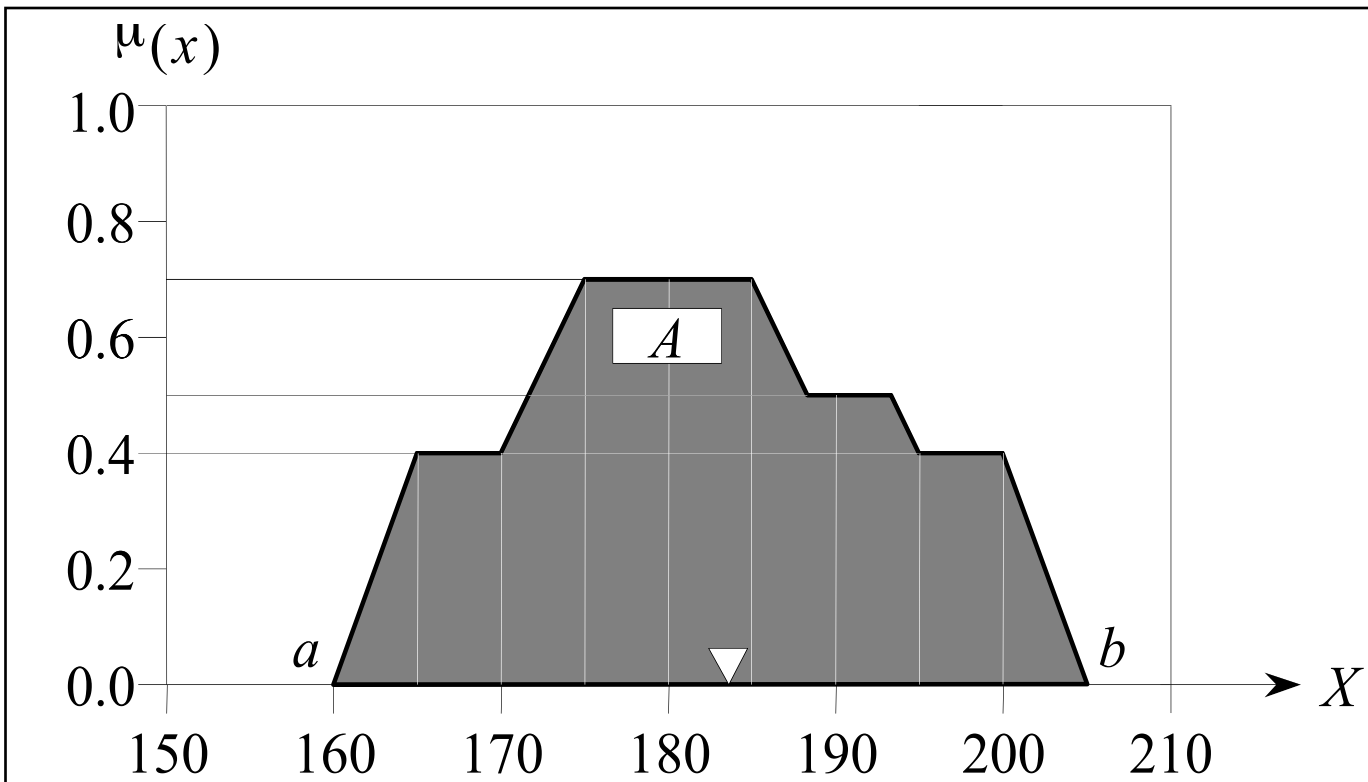
The last step in the fuzzy inference process is defuzzification. Fuzziness helps us to evaluate the rules, but the final output of a fuzzy system has to be a crisp number. The input for the defuzzification process is the aggregate output fuzzy set and the output is a single number.

- There are several defuzzification methods, but probably the most popular one is the **centroid technique**. It finds the point where a vertical line would slice the aggregate set into two equal masses. Mathematically this **centre of gravity (COG)** can be expressed as:

$$COG = \frac{\int_a^b \mu_A(x)\, x\, dx}{\int_a^b \mu_A(x)\, dx}$$

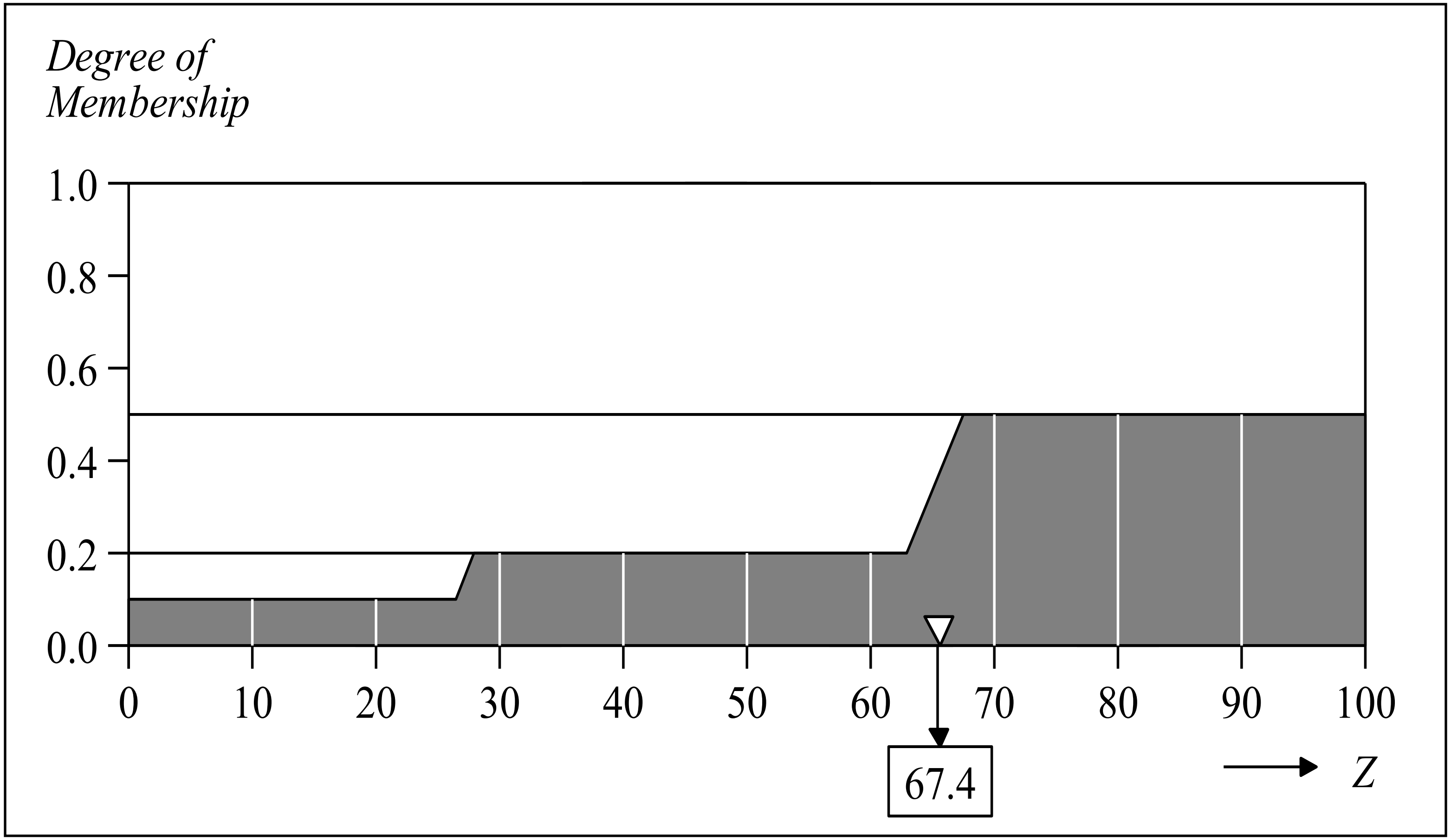
- Centroid defuzzification method finds a point representing the centre of gravity of the fuzzy set, *A*, on the interval, *ab*.

- A reasonable estimate can be obtained by calculating it over a sample of points.



Centre of gravity (COG):

$$COG = \frac{(0+10+20)\times 0.1 + (30+40+50+60)\times 0.2 + (70+80+90+100)\times 0.5}{0.1+0.1+0.1+0.2+0.2+0.2+0.2+0.5+0.5+0.5+0.5} = 67.4$$



Sugeno fuzzy inference

- Mamdani-style inference, as we have just seen, requires us to find the centroid of a two-dimensional shape by integrating across a continuously varying function. In general, this process is not computationally efficient.
- **Michio Sugeno** suggested to use a single spike, a *singleton*, as the membership function of the rule consequent. A **singleton**, or more precisely a **fuzzy singleton**, is a fuzzy set with a membership function that is unity at a single particular point on the universe of discourse and zero everywhere else.

Sugeno-style fuzzy inference is very similar to the Mamdani method. Sugeno changed only a rule consequent. Instead of a fuzzy set, he used a mathematical function of the input variable. The format of the **Sugeno-style fuzzy rule** is

IF x is A
AND y is B
THEN z is $f(x, y)$

where x, y and z are linguistic variables; A and B are fuzzy sets on universe of discourses X and Y, respectively; and $f(x, y)$ is a mathematical function.

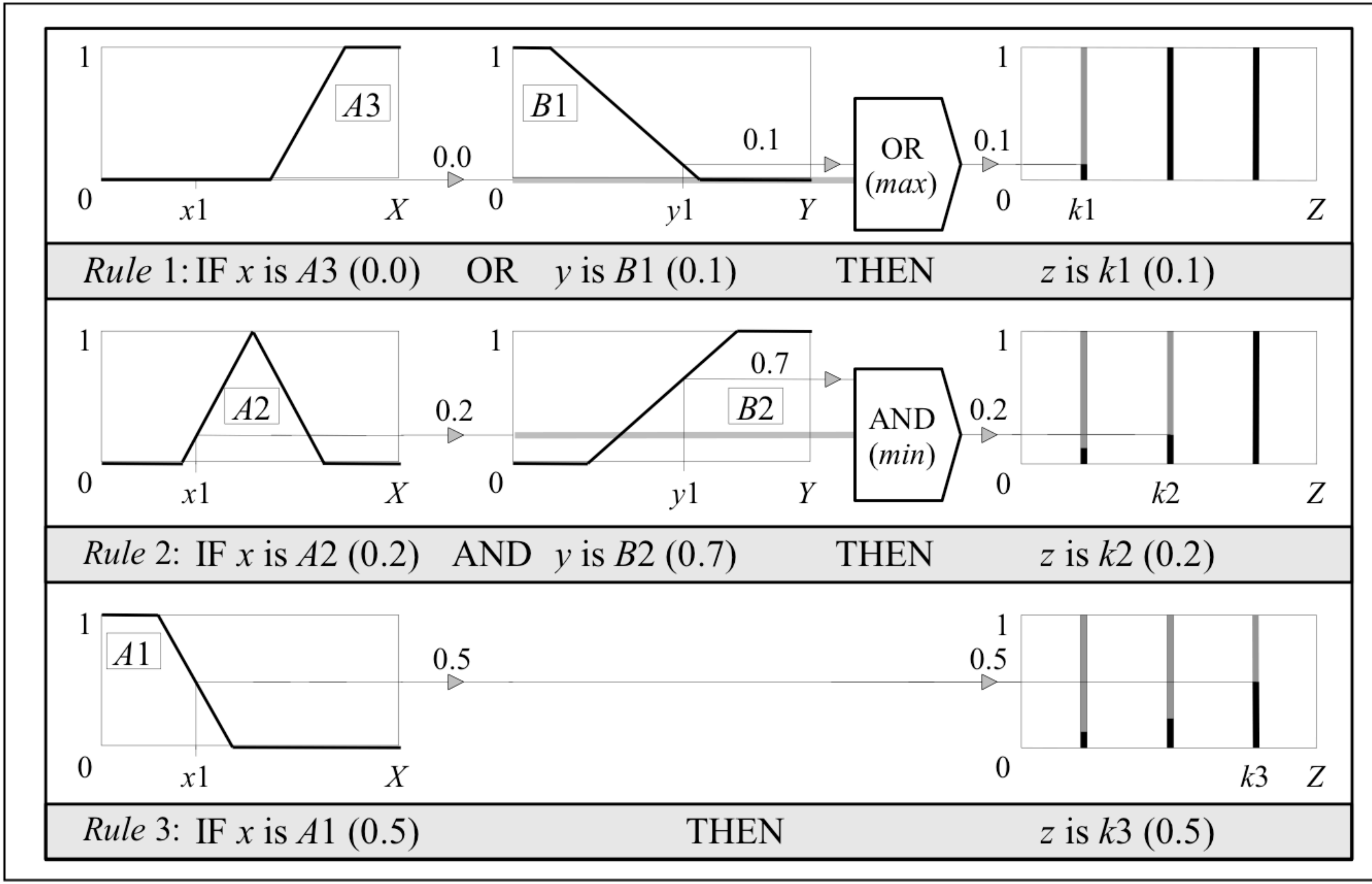
The most commonly used **zero-order Sugeno fuzzy model** applies fuzzy rules in the following form:

IF x is A
AND y is B
THEN z is k

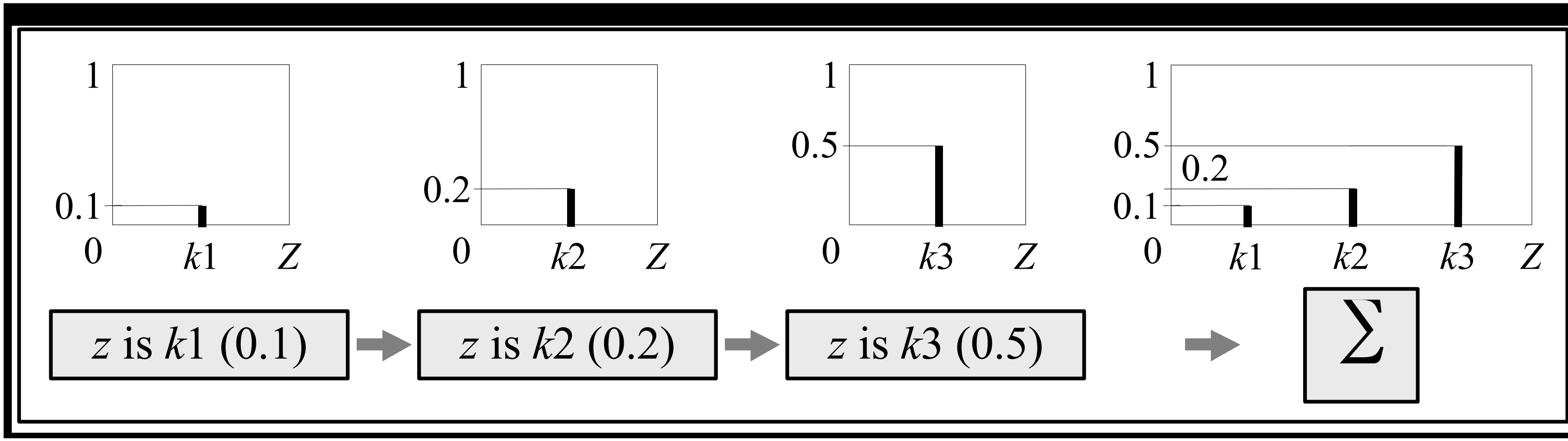
where k is a constant.

In this case, the output of each fuzzy rule is constant. All consequent membership functions are represented by singleton spikes.

Sugeno-style rule evaluation



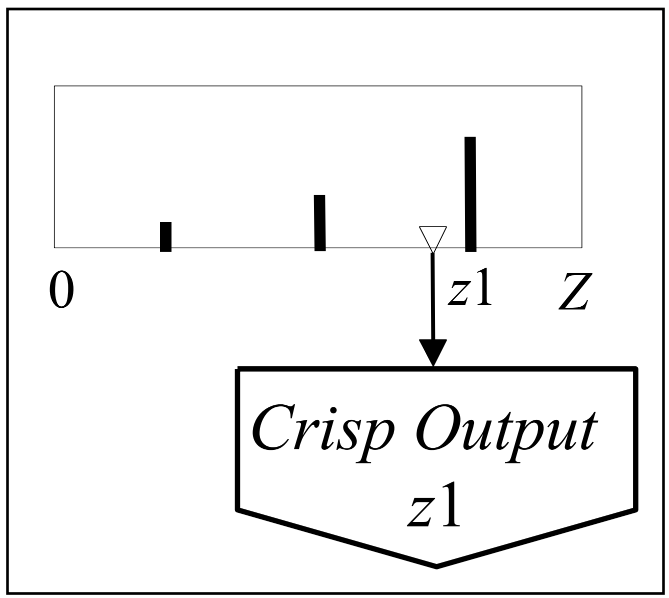
Sugeno-style aggregation of the rule outputs



Weighted average (WA):

$$WA = \frac{\mu(k1)\times k1 + \mu(k2)\times k2 + \mu(k3)\times k3}{\mu(k1)+\mu(k2)+\mu(k3)} = \frac{0.1\times 20 + 0.2\times 50 + 0.5\times 80}{0.1+0.2+0.5} = 65$$

Sugeno-style defuzzification



How to make a decision on which method to apply - Mamdani or Sugeno?

- Mamdani method is widely accepted for capturing expert knowledge. It allows us to describe the expertise in more intuitive, more human-like manner. However, Mamdani-type fuzzy inference entails a substantial computational burden.
- On the other hand, Sugeno method is computationally effective and works well with optimisation and adaptive techniques, which makes it very attractive in control problems, particularly for dynamic nonlinear systems.