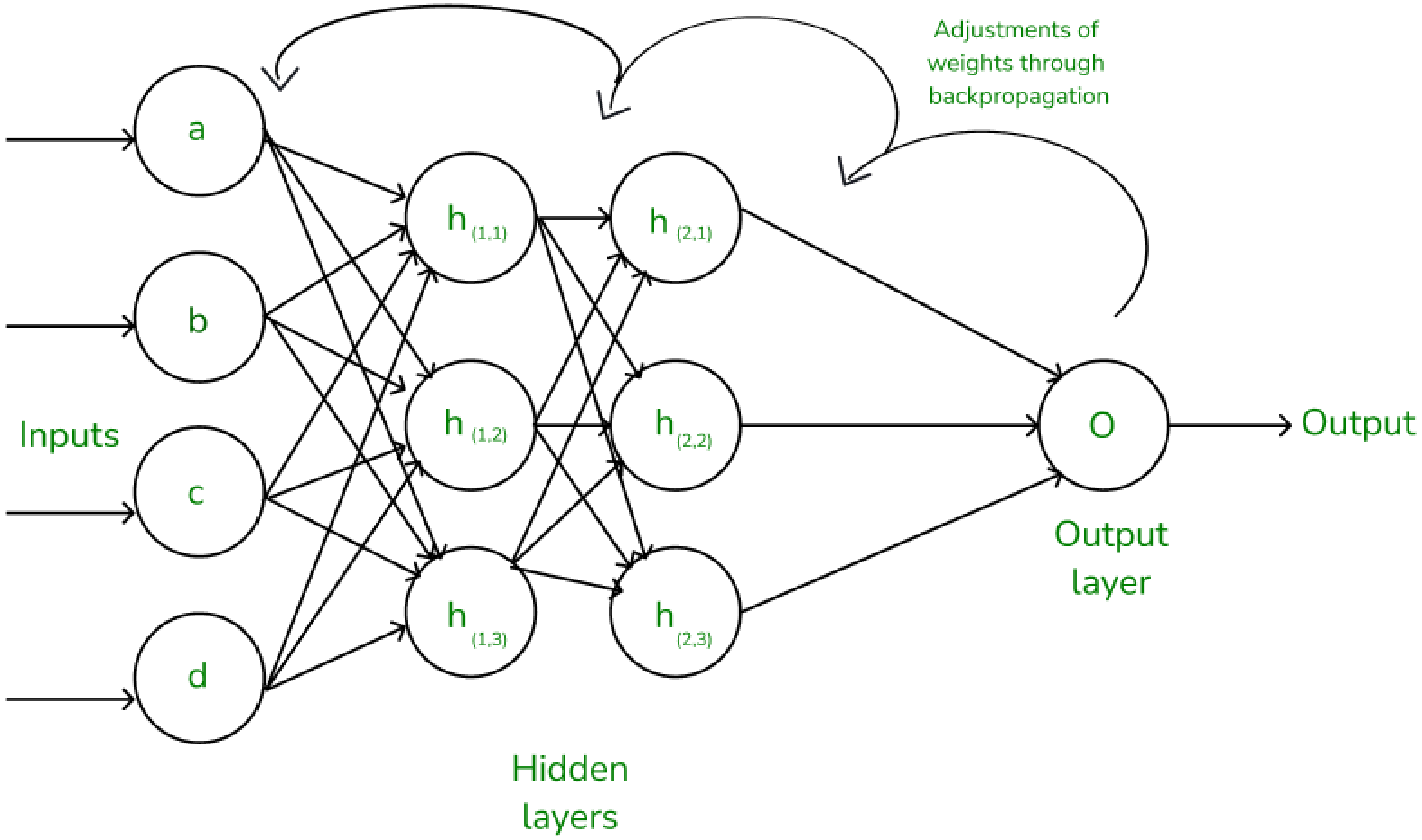
INTRODUCTION TO ARTIFICIAL INTELLIGENCE

PGS. TS. Nguyễn Thanh Bình

ERROR BACK-PROPAGATION ALGORITHM



ERROR BACK-PROPAGATION ALGORITHM

- We already trained linear networks by gradient descent method.
- Could we do similarly for multi-layer networks?
- Difficulties:
  - Do not have the target values for the hidden units!!!
- That is an unsolved problems in 1950s.
- 30 years later, the error back-propagation algorithm was proposed to train hidden units, leading to a new wave of neural network research and applications.

ERROR BACK-PROPAGATION ALGORITHM

- The algorithm provides a way to train networks with any number of hidden units arranged in any number of layers.
- The network does not have to be organized in layers.
- There must be a way to order the units such that all connections go from “earlier” (closer to the input) to “later” ones (closer to the output).
- Their connection pattern must not contain any cycles.
- The network that respect this constraint are called feed-forward networks and their connection pattern forms a directed acyclic graph

ERROR BACK-PROPAGATION ALGORITHM

- We want to train a multi-layer feed forward network by gradient descent to approximate an unknown function, based on some training data consisting of pairs (x,t).
- The vector x represents a pattern input to the network and the vector t the corresponding target (desired output).
- The overall gradient with respect to the entire training set is just the sum of the gradients for each pattern.

ERROR BACK-PROPAGATION ALGORITHM

- We will describe how to compute the gradient for just a single training pattern.
- We denote the weight from unit j to unit i by w_{ij}
- Some definitions:
  - The error signal for unit j:

$$\partial_j = -\frac{\partial E}{\partial \mathrm{net}_j}$$

  - The gradient for weight w_{ij}

$$\Delta w_{ij} = -\frac{\partial E}{\partial w_{ij}}$$

ERROR BACK-PROPAGATION ALGORITHM

- Some definitions:
  - The set of nodes anterior to unit I

$$A_i = \{j : \exists w_{ij}\}$$

  - The set of nodes posterior to unit j:

$$P_j = \{i : \exists w_{ij}\}$$

ERROR BACK-PROPAGATION ALGORITHM

- **The gradient**: we expand the gradient into two factors by using the chain rule:

$$\Delta w_{ij} = -\frac{\partial E}{\partial \mathrm{net}_i} \frac{\partial \mathrm{net}_i}{\partial w_{ij}}$$

- Here:

$$\frac{\partial \mathrm{net}_i}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}} \sum_{k \in A_i} w_{ik} y_k = y_j$$

- To compute this gradient, we need to know the activity and the error for all relevant nodes in the network.

ERROR BACK-PROPAGATION ALGORITHM

- Forward activation:
    - The activity of the input units is determined by the network's external input x.
    - For all other units, the activity are propagated forward:

$$y_i = f_i \sum_{j \in A_i} w_{ij} y_j$$

ERROR BACK-PROPAGATION ALGORITHM

- Calculating output error: assuming that we are using the sum-squared loss function:

$$E = \frac{1}{2} \sum_{o} (t_o - y_o)^2$$

- The error for the output unit is as following:

$$\delta_o = t_o - y_o$$

ERROR BACK-PROPAGATION ALGORITHM

- Error back-propagation: for hidden units, we must propagate the error back from the output nodes. Again, using the chain rule, we can expand the error of a hidden unit in terms of its posterior nodes:

$$\delta_j = \sum_{i \in P_j} \frac{\partial E}{\partial net_i} \frac{\partial net_i}{\partial y_j} \frac{\partial y_j}{\partial net_j}$$

ERROR BACK-PROPAGATION ALGORITHM

- The first is just the error of node i.
- The second is:

$$\frac{\partial net_i}{\partial y_j} = \frac{\partial}{\partial y_j} \sum_{k \in A_i} w_{ik} y_k = w_{ij}$$

- While the third term is the derivative of node j's activation function:

$$\frac{\partial y_j}{\partial net_j} = f_j'(net_j)$$

- So

$$\delta_j = f_j'(net_j) \sum_{i \in P_j} \delta_i w_{ij}$$

FUZZY EXPERT SYSTEMS: FUZZY LOGIC

- Introduction, or what is fuzzy thinking?
- Fuzzy sets
- Linguistic variables and hedges
- Operations of fuzzy sets
- Fuzzy rules
- Summary

INTRODUCTION, OR WHAT IS FUZZY THINKING?

- Experts rely on **common sense** when they solve problems.
- **How can we represent expert knowledge that uses vague and ambiguous terms in a computer?**
- 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 based on the idea that all things admit of degrees. Temperature, height, speed, distance, beauty – all come on a sliding scale. The motor is running ***really hot.*** Tom is a ***very tall*** guy

INTRODUCTION, OR WHAT IS FUZZY THINKING?

- Boolean logic uses sharp distinctions. It forces us to draw lines between members of a class and non- members. For instance, we may say, Tom is tall because his height is 181 cm. If we drew a line at 180 cm, we would find that David, who is 179 cm, is small. Is David really a small man or we have just drawn an arbitrary line in the sand?
- Fuzzy logic reflects how people think. It attempts to model our sense of words, our decision making and our common sense. As a result, it is leading to new, more human, intelligent systems.

INTRODUCTION, OR WHAT IS FUZZY THINKING?

- Fuzzy, or multi-valued logic was introduced **in the 1930s by Jan Lukasiewicz** , a Polish philosopher. While classical logic operates with only two values 1 (true) and 0 (false), Lukasiewicz introduced logic that extended the range of truth values to all real numbers in the interval between 0 and 1. He used a number in this interval to represent the possibility that a given statement was true or false. For example, the possibility that a man 181 cm tall is really tall might be set to a value of 0.86. It is likely that the man is tall. This work led to an inexact reasoning technique often called **possibility theory**.

INTRODUCTION, OR WHAT IS FUZZY THINKING?

- Later, in 1937, **Max Black** published a paper called “Vagueness: an exercise in logical analysis”. In this paper, he argued that a continuum implies degrees. Imagine, he said, a line of countless “chairs”. At one end is a Chippendale. Next to it is a near-Chippendale, in fact indistinguishable from the first item. Succeeding “chairs” are less and less chair-like, until the line ends with a log. When does a chair become a log? Max Black stated that if a continuum is discrete, a number can be allocated to each element. He accepted **vagueness as a matter of probability**.

INTRODUCTION, OR WHAT IS FUZZY THINKING?

- In 1965 **Lotfi Zadeh**, published his famous paper “Fuzzy sets”. Zadeh extended the work on possibility theory into a formal system of mathematical logic, and introduced a new concept for applying natural language terms. This new logic for representing and manipulating fuzzy terms was called **fuzzy logic**, and Zadeh became the Master of fuzzy logic.

INTRODUCTION, OR WHAT IS FUZZY THINKING?

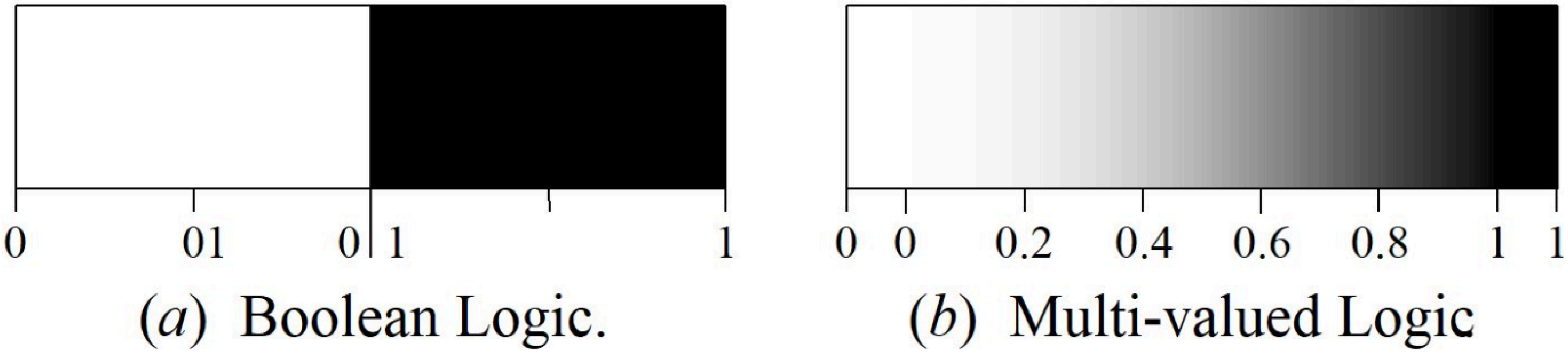
- Why fuzzy?
  - As Zadeh said, the term is concrete, immediate and descriptive; we all know what it means. However, many people in the West were repelled by the word fuzzy , because it is usually used in a negative sense.
- Why logic?
  - Fuzziness rests on fuzzy set theory, and fuzzy logic is just a small part of that theory

INTRODUCTION, OR WHAT IS FUZZY THINKING?

- **Fuzzy logic** is a set of mathematical principles for knowledge representation based on degrees of membership.
- Unlike two-valued Boolean logic, fuzzy logic is multi-valued. It deals with degrees of membership and degrees of truth. Fuzzy logic uses the continuum of logical values between 0 (completely false) and 1 (completely true). Instead of just black and white, it employs the spectrum of colours, accepting that things can be partly true and partly false at the same time

INTRODUCTION, OR WHAT IS FUZZY THINKING?

- Range of logical values in Boolean and fuzzy logic



FUZZY SETS

- The concept of a **set** is fundamental to mathematics.
- However, our own language is also the supreme expression of sets. For example, **car** indicates the **set of cars**. When we say a car , we mean one out of the set of cars.

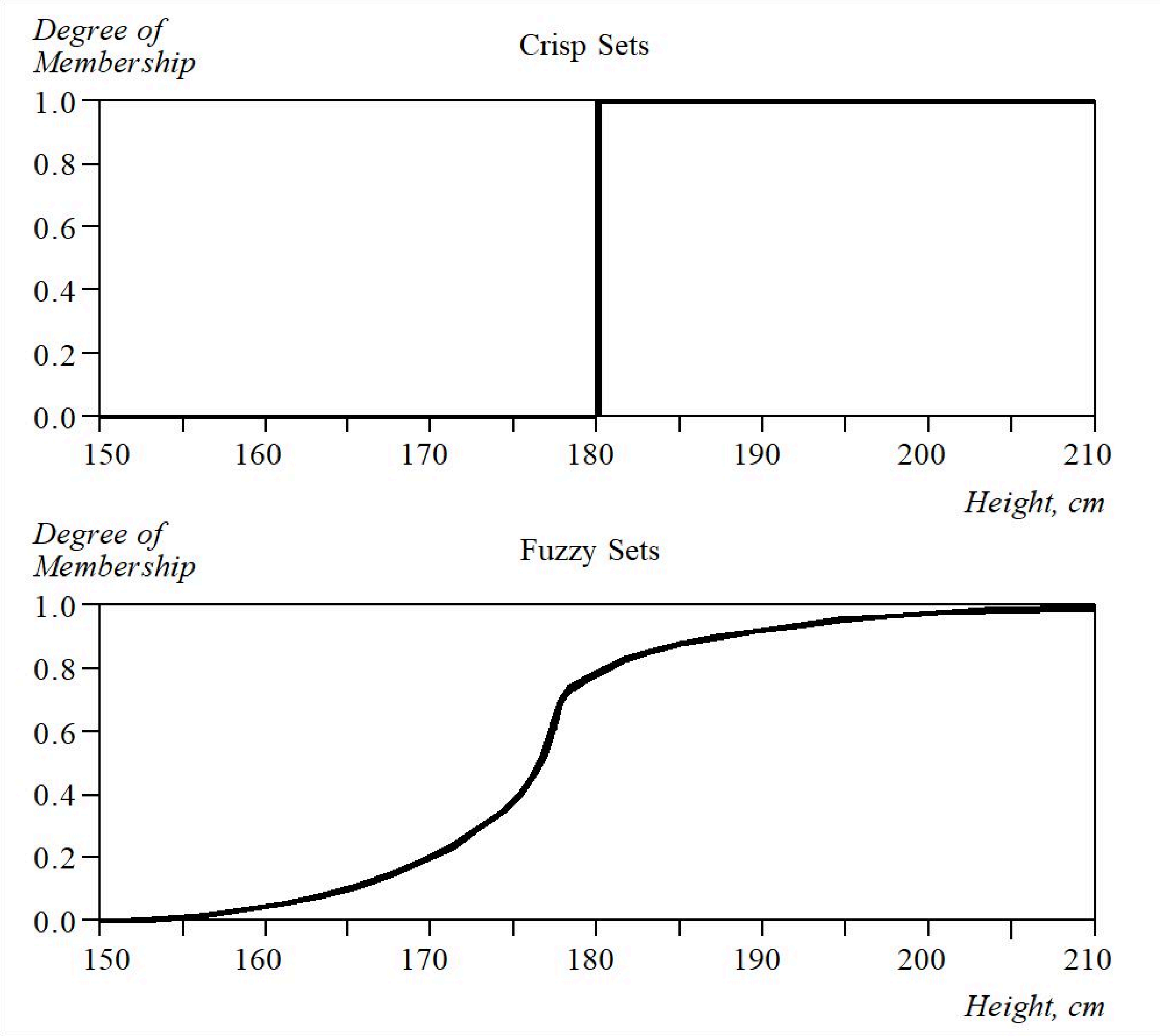
FUZZY SETS

- The classical example in fuzzy sets is tall men. The elements of the fuzzy set “tall men” are all men, but their degrees of membership depend on their height.

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

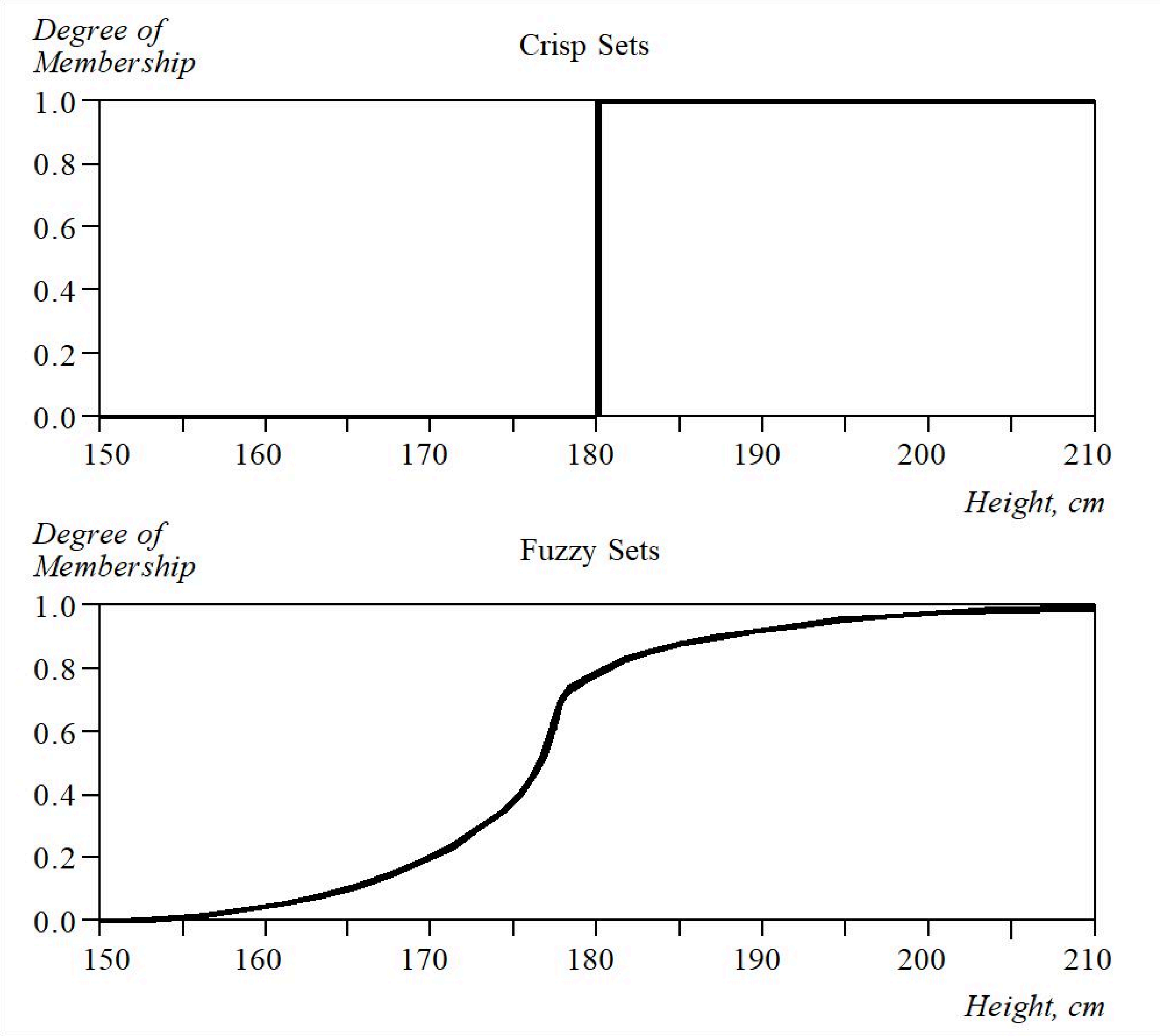
FUZZY SETS

- Crisp and fuzzy sets of “tall men”
    - The x-axis represents the **universe of discourse** – the range of all possible values applicable to a chosen variable. In our case, the variable is the man height. According to this representation, the universe of men’sheights consists of all tall men.



FUZZY SETS

- Crisp and fuzzy sets of “tall men”
  - The y-axis represents the **membership value of the fuzzy set**. In our case, the fuzzy set of “tall men” maps height values into corresponding membership values



FUZZY SETS

- A **fuzzy set** is a set with fuzzy boundaries.
- 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**
- $f_A(x) : X \to \{0, 1\}$, where

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

This set maps universe X to a set of two elements. For any element x of universe X, characteristic function $f_A(x)$ is equal to 1 if x is an element of set A, and is equal to 0 if x is not an element of A.

FUZZY SETS

- In the fuzzy theory, fuzzy set A of universe X is defined by function $\mu_A(x)$ called the **membership function** of set A
- $\mu_A(x) : X \rightarrow [0,1]$, where

$$\begin{cases} \mu_A(x) = 1 & \text{if } x \text{ is totally in } A; \\ \mu_A(x) = 0 & \text{if } x \text{ is not in } A; \\ 0 < \mu_A(x) < 1 & \text{if } x \text{ is partly in } A. \end{cases}$$

This set allows a continuum of possible choices. For any element x of universe X, membership function $\mu_A(x)$ equals the degree to which x is an element of set A. This degree, a value between 0 and 1, represents the **degree of membership**, also called **membership value**, of element x in set A.

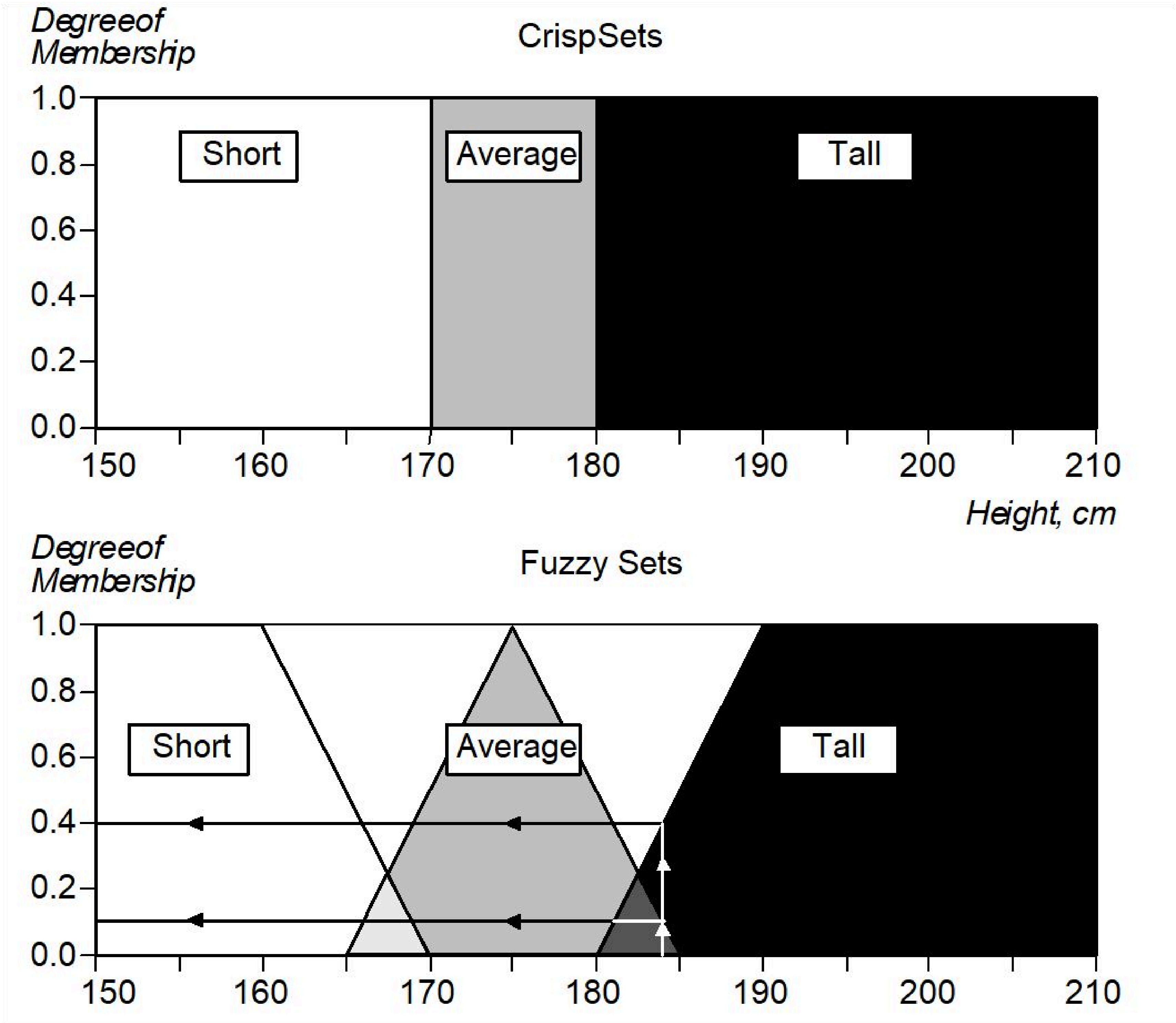
FUZZY SETS

How to represent a fuzzy set in a computer?

- First, we determine the membership functions. In our “tall men” example, we can obtain fuzzy sets of tall, short and average men.
- The universe of discourse – the men’s heights – consists of three sets: short, average and tall men. As you will see, a man who is 184 cm tall is a member of the average men set with a degree of membership of 0.1, and at the same time, he is also a member of the tall men set with a degree of 0.4.

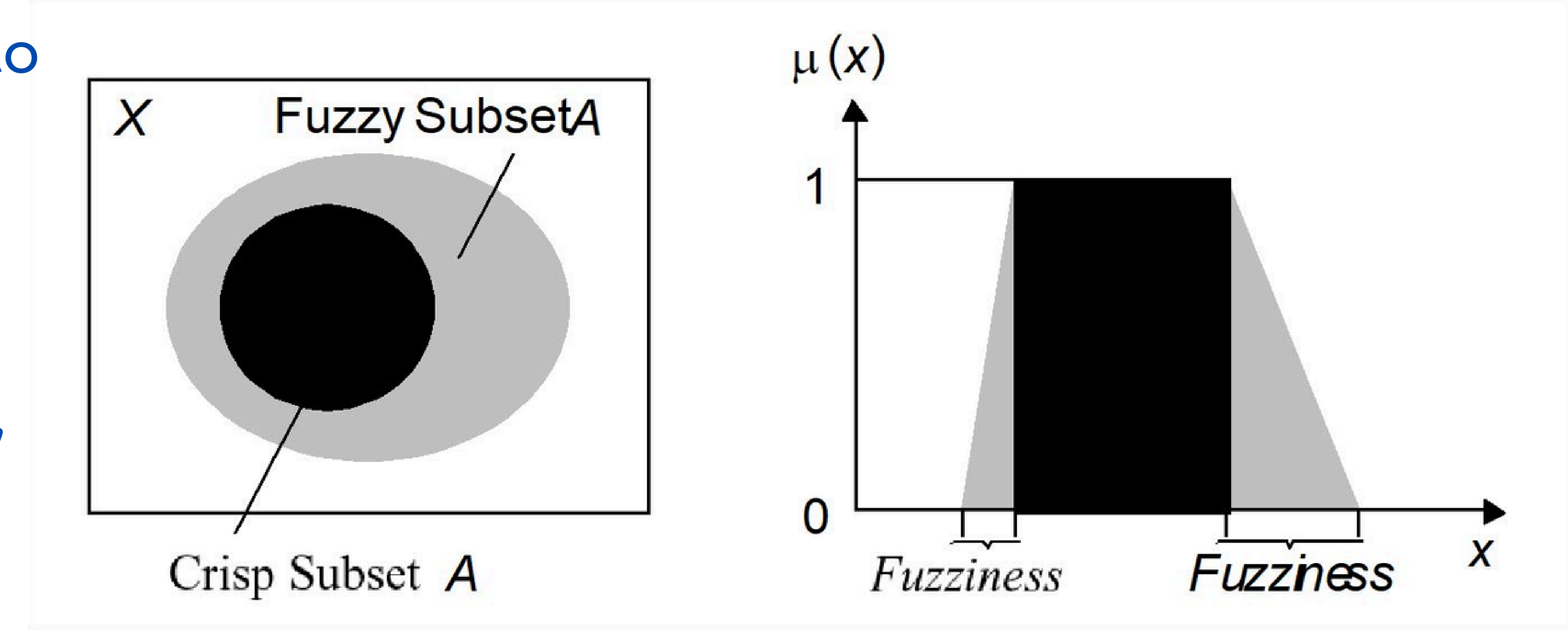
FUZZY SETS

Crisp and fuzzy sets of short, average and tall men



REPRESENTATION OF CRISP AND FUZZY SUBSETS

Typical functions that can be used to represent a fuzzy set are sigmoid, gaussian and pi. However, these functions increase the time of computation. Therefore, in practice, most applications use **linear fit functions.**



LINGUISTIC VARIABLES AND HEDGES

- In fuzzy expert systems, linguistic variables are used in fuzzy rules.
- For example:

IF wind is strong
THEN sailing is good

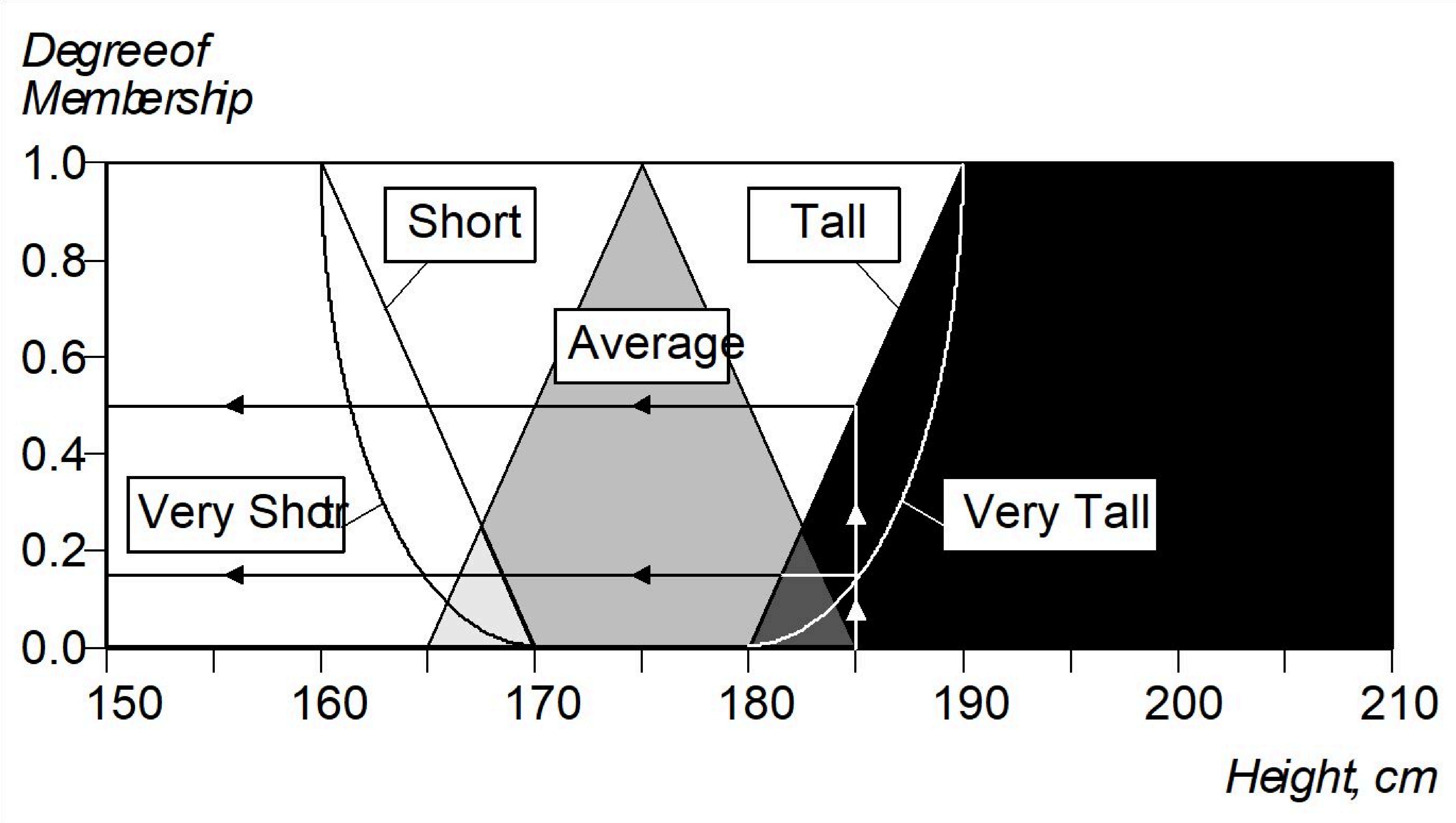
IF project_duration is long
THEN completion_risk is high

IF speed is slow
THEN stopping_distance is short

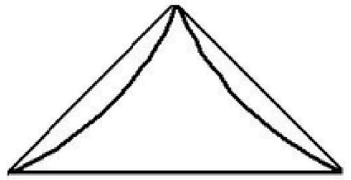
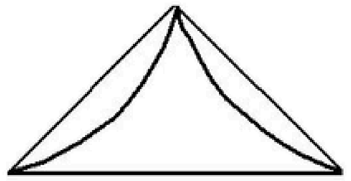
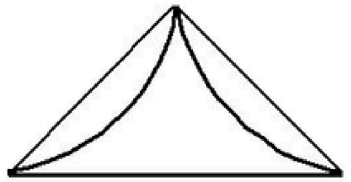
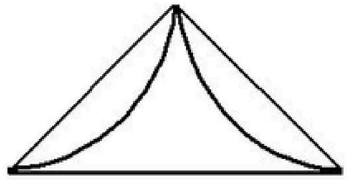
LINGUISTIC VARIABLES AND HEDGES

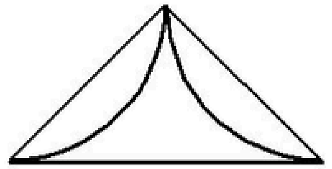
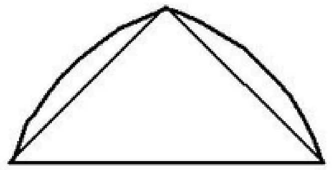
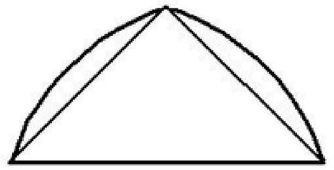
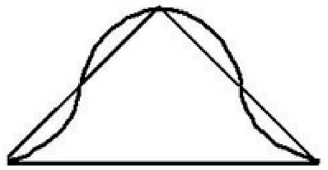
- The range of possible values of a linguistic variable represents the universe of discourse of that variable. For example, the universe of discourse of the linguistic variable speed might have the range between 0 and 220 km/h and may include such fuzzy subsets as very slow, slow, medium, fast, and very fast.
- A linguistic variable carries with it the concept of fuzzy set qualifiers, called **hedges**.
- **Hedges are terms that modify the shape of fuzzy sets. They include adverbs such as very, somewhat, quite, more or less and slightly.**

FUZZY SETS WITH THE HEDGE VERY



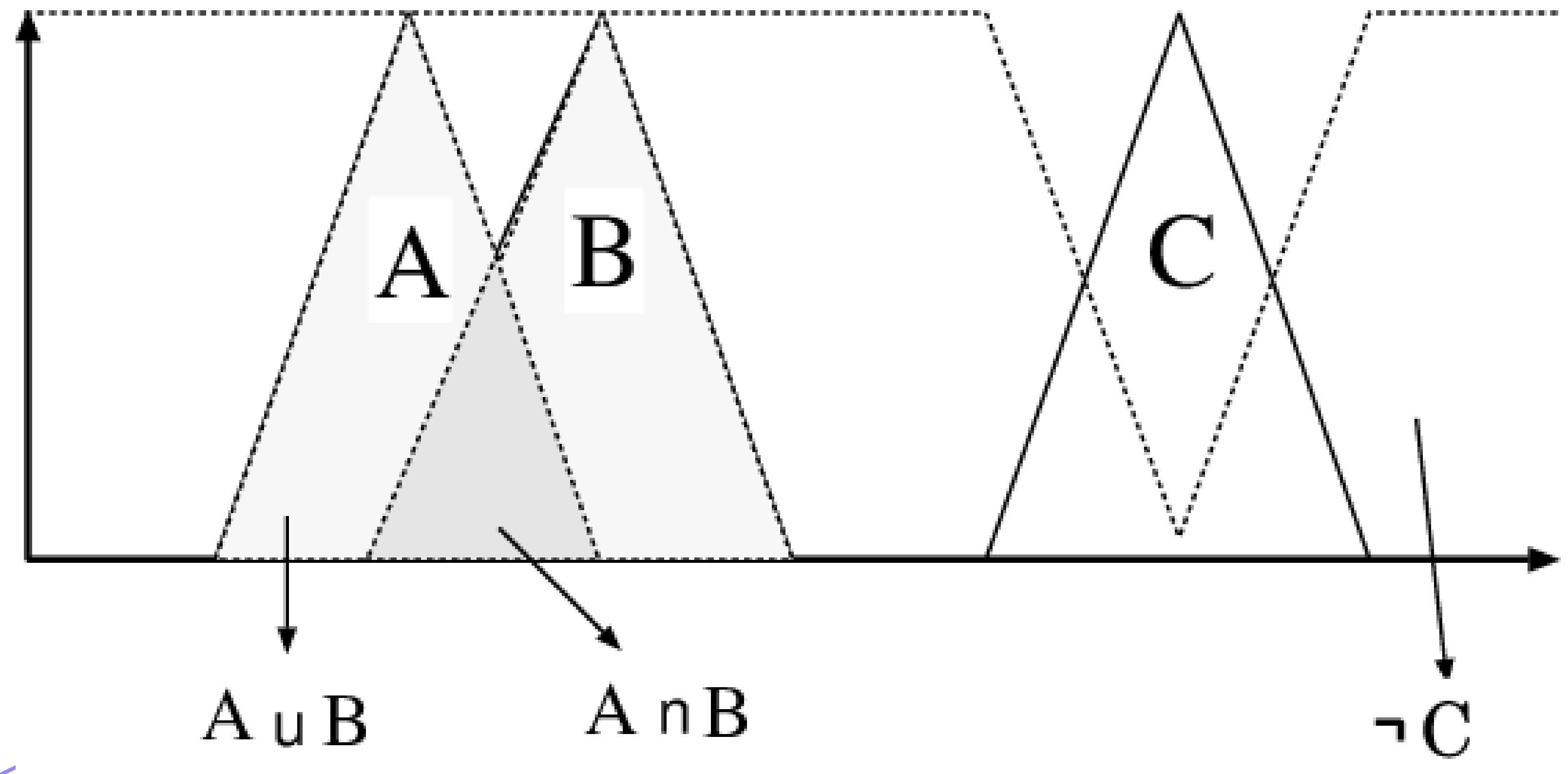
REPRESENTATION OF HEDGES IN FUZZY LOGIC

Hedge	*Mathematical Expression*	*Graphical Representation*
A little	$[\mu_A(x)]^{1.3}$	
Slightly	$[\mu_A(x)]^{1.7}$	
Very	$[\mu_A(x)]^{2}$	
Extremely	$[\mu_A(x)]^{3}$	

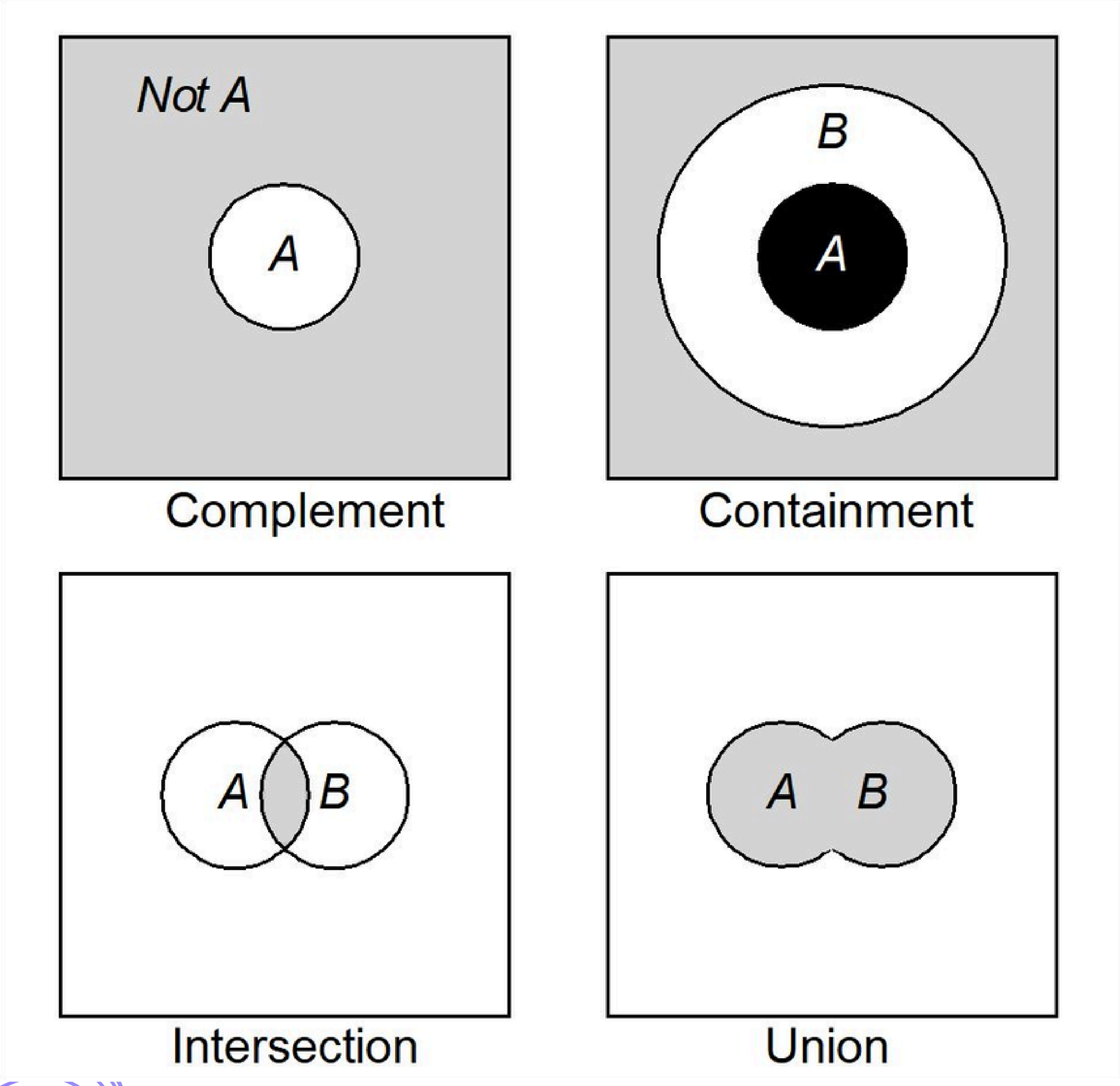
Hedge	*Mathematical Expression*	*Graphical Representation*
Very very	$[\mu_A(x)]^4$	
More or less	$\sqrt{\mu_A(x)}$	
Somewhat	$\sqrt{\mu_A(x)}$	
Indeed	$2\,[\mu_A(x)]^2$ if $0 \le \mu_A \le 0.5$ $1 - 2\,[1 - \mu_A(x)]^2$ if $0.5 < \mu_A \le 1$	

OPERATIONS OF FUZZY SETS

- The classical set theory developed in the late 19th century by Georg Cantor describes how crisp sets can interact. These interactions are called **operations**



CANTOR'S SETS



COMPLEMENT

Crisp Sets: Who does not belong to the set?

Fuzzy Sets: How much do elements not belong to the set?

The complement of a set is an opposite of this set. For example, if we have the set of tall men, its complement is the set of NOT tall men. When we remove the tall men set from the universe of discourse, we obtain the complement. If A is the fuzzy set, its complement ØA can be found as follows:

$$\overline{\mu_A(x)} = 1 - \mu_A(x)$$

CONTAINMENT

Crisp Sets: Which sets belong to which other sets?

Fuzzy Sets: Which sets belong to other sets?

- Similar to a Chinese box, a set can contain other sets. The smaller set is called the **subset**.
- For example, the set of tall men contains all tall men; very tall men is a subset of tall men. However, the tall men set is just a subset of the set of men.
- In crisp sets, all elements of a subset entirely belong to a larger set. In fuzzy sets, however, each element can belong less to the subset than to the larger set. Elements of the fuzzy subset have smaller memberships in it than in the larger set.

INTERSECTION

Crisp Sets: Which element belongs to both sets?

Fuzzy Sets: How much of the element is in both sets?

- In classical set theory, an intersection between two sets contains the elements shared by these sets.
- For example, the intersection of the set of tall men and the set of fat men is the area where these sets overlap.
- In fuzzy sets, an element may partly belong to both sets with different memberships. A fuzzy intersection is the lower membership in both sets of each element. The fuzzy intersection of two fuzzy sets A and B on universe of discourse X: $\mu_{A\cap B}(x) = \min[\mu_A(x), \mu_B(x)] = \mu_A(x) \cap \mu_B(x), \text{where } x \in X$

UNION

Crisp Sets: Which element belongs to either set?

Fuzzy Sets: How much of the element is in either set?

- The union of two crisp sets consists of every element that falls into either set.
- For example, the union of tall men and fat men contains all men who are tall OR fat.
- In fuzzy sets, the union is the reverse of the intersection. That is, the union is the largest membership value of the element in either set. The fuzzy operation for forming the union of two fuzzy sets A and B on universe X can be given as:

$$\mu_{A\cup B}(x) = \max[\mu_A(x), \mu_B(x)] = \mu_A(x) \cup \mu_B(x), \text{where } x \in X$$

UNION

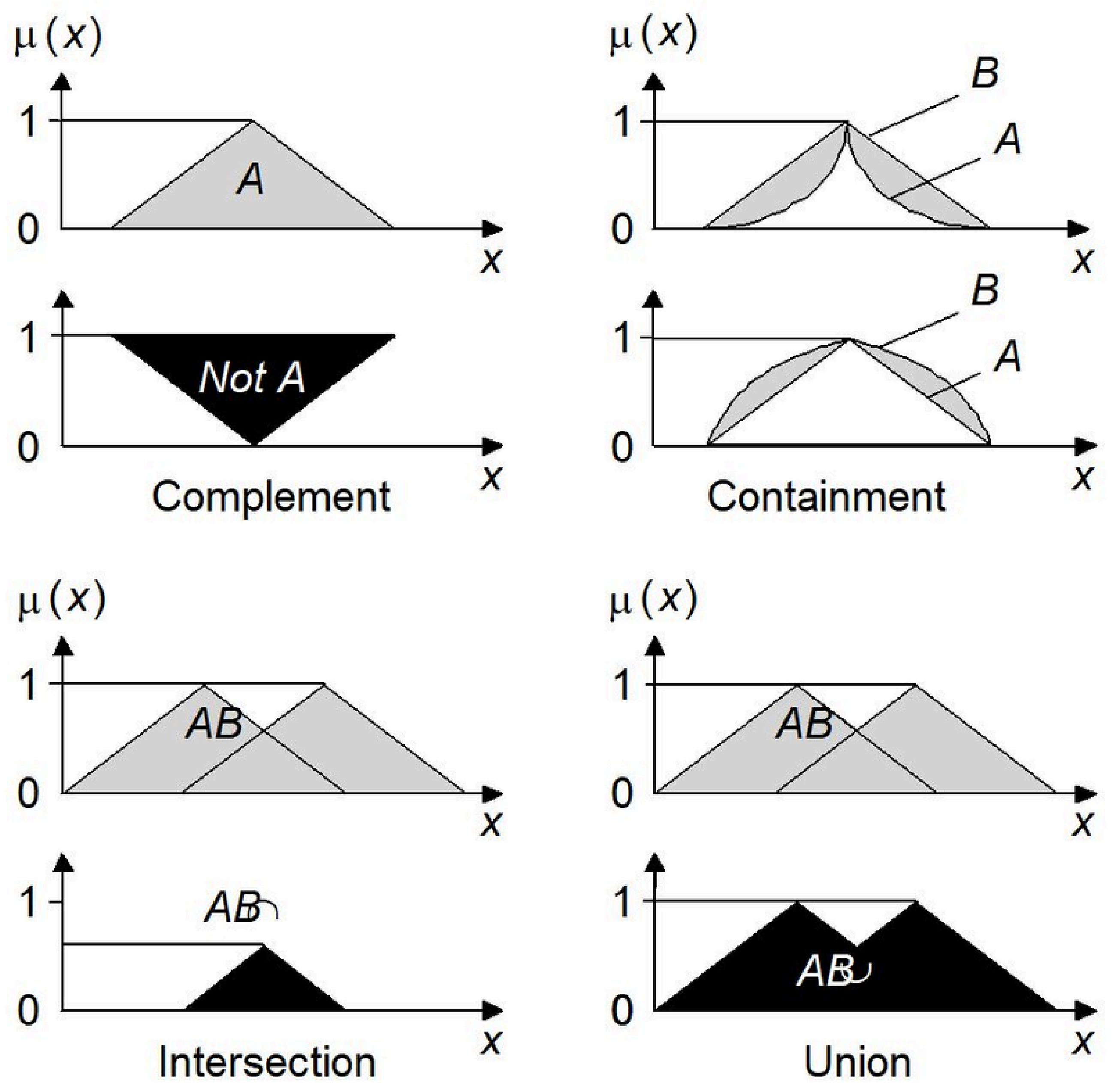
Crisp Sets: Which element belongs to either set?

Fuzzy Sets: How much of the element is in either set?

- The union of two crisp sets consists of every element that falls into either set.
- For example, the union of tall men and fat men contains all men who are tall OR fat.
- In fuzzy sets, the union is the reverse of the intersection. That is, the union is the largest membership value of the element in either set. The fuzzy operation for forming the union of two fuzzy sets A and B on universe X can be given as:

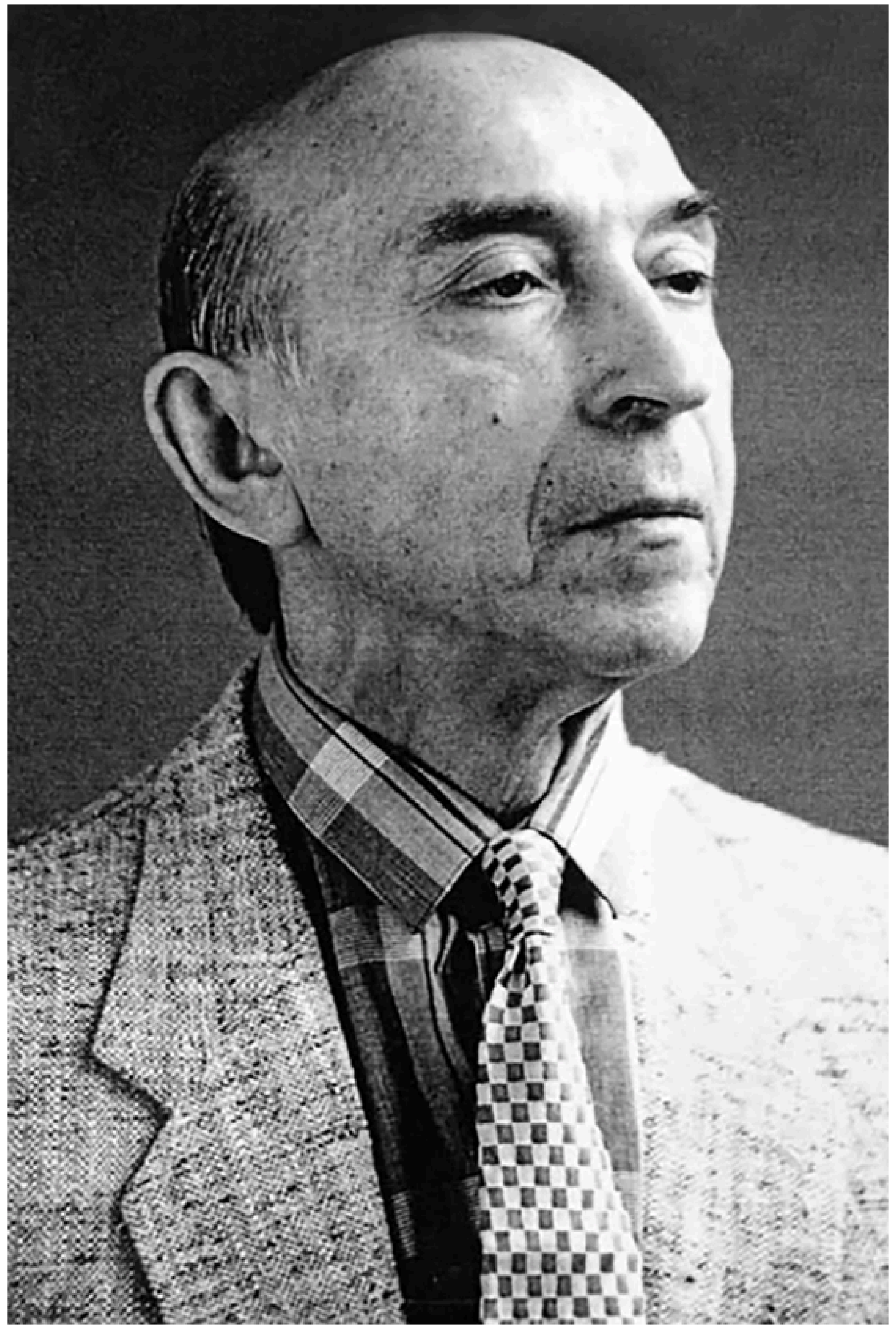
$$\mu_{A \cup B}(x) = \max[\mu_A(x), \mu_B(x)] = \mu_A(x) \cup \mu_B(x), \text{where } x \in X$$

OPERATIONS OF FUZZY SETS



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

What is a fuzzy rule?

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

$$\begin{array}{rl} \textbf{IF} & \text{x is A} \\ \textbf{THEN} & \text{y is B} \end{array}$$

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

FUZZY RULES

What is the difference between classical and fuzzy rules?

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

Rule 1:
IF speed is fast
THEN stopping_distance is long

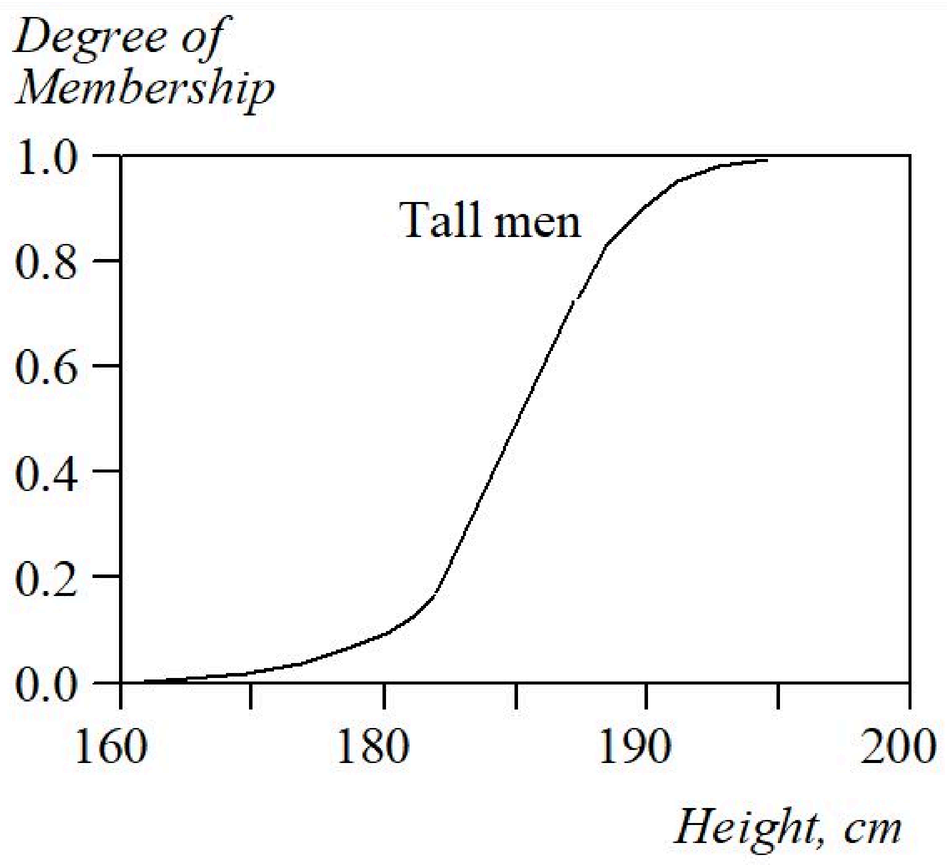
Rule 2:
IF speed is slow
THEN stopping_distance is short

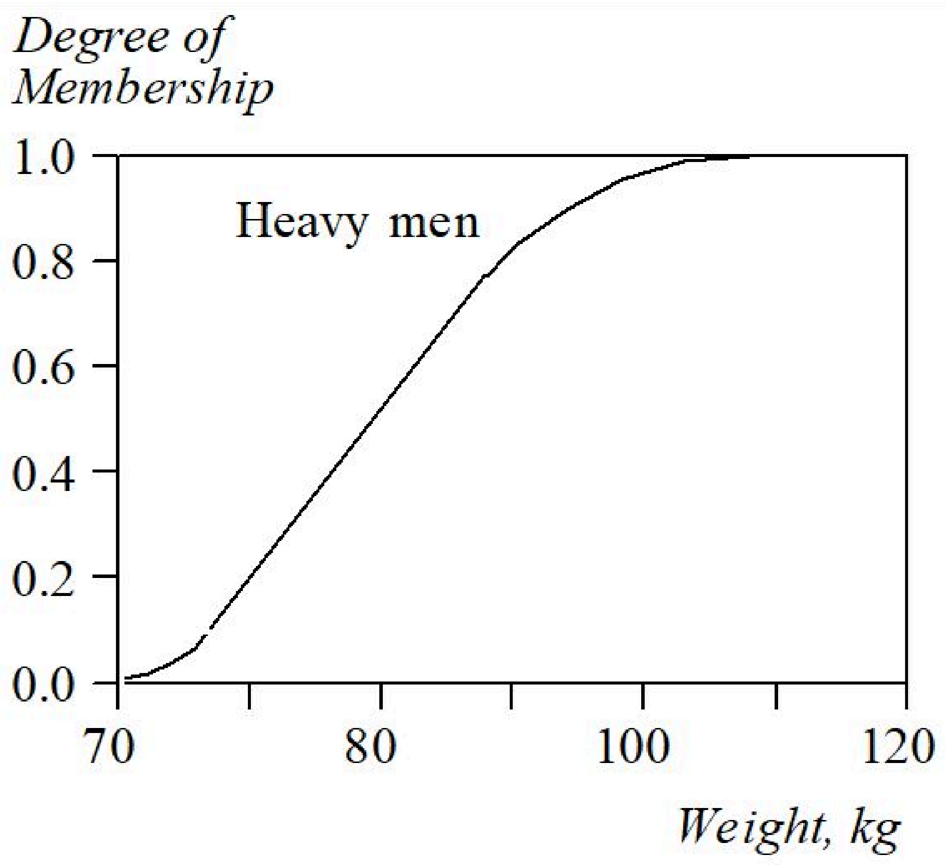
- In fuzzy rules, the linguistic variable speed also has the range (the universe of discourse) between 0 and 220 km/h, but this range includes fuzzy sets, such as slow, medium and fast. The universe of discourse of the linguistic variable stopping_distance can be between 0 and 300 m and may include such fuzzy sets as short, medium and long

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.

FUZZY SETS OF TALL AND HEAVY MEN

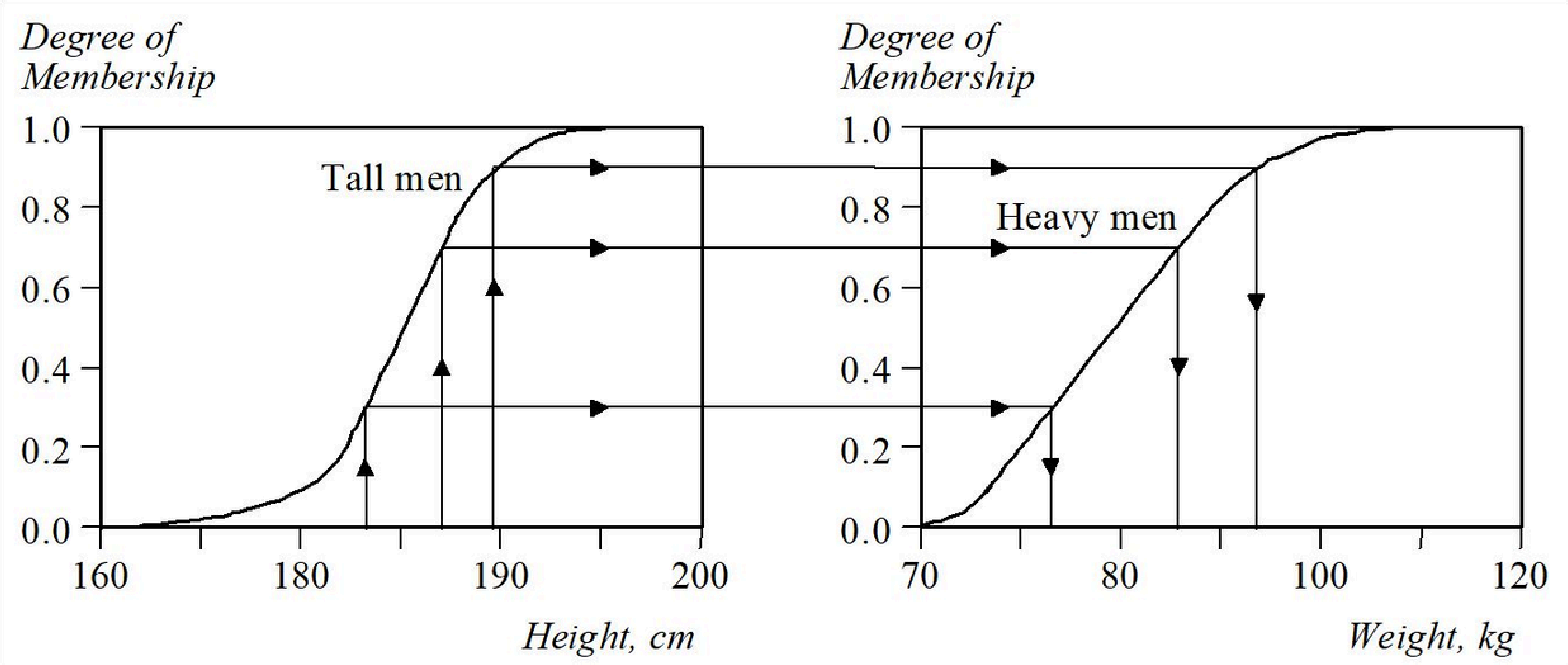




These fuzzy sets provide the basis for a weight estimation model. The model is based on a relationship between a man's height and his weight:

IF height is tall
THEN weight is heavy

FUZZY SETS OF TALL AND HEAVY MEN



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.**

FUZZY RULES

A fuzzy rule can have multiple antecedents, for example:

IF project_duration is long
AND project_staffing is large
AND project_funding is inadequate
THEN risk is high

IF service is excellent
OR food is delicious
THEN tip is generous

FUZZY RULES

The consequent of a fuzzy rule can also include multiple parts, for instance:

$$\begin{array}{rl} \textbf{IF} & \text{temperature is hot} \\ \textbf{THEN} & \text{hot\_water is reduced;} \\ & \text{cold\_water is increased} \end{array}$$