(Concepts of) Machine Learning

Lecture 2: Uncertainty modelling and fuzzy logic

George Magoulas
gmagoulas@dcs.bbk.ac.uk

Contents

- Reasoning under uncertainty
  - Probabilities, Bayesian rule and Bayesian reasoning
  - Certainty factors
- Reasoning with fuzzy logic
  - Fuzzy sets and fuzzy variables
  - Fuzzy rules
  - Fuzzy systems
  - Fuzzy clustering

Reasoning under uncertainty

Reasoning is associated with thinking, cognition, and intelligence.

➢ **What is "reasoning" in AI:** a mechanism for selecting the relevant facts and extracting conclusions from them in a logical way.

- The concept of probability is associated with words like "probably", "likely", "maybe", "perhaps", "possibly".
- The probability of an event is the proportion of cases in which the event occurs
- It can be expressed mathematically as a numerical index with a range between zero to unity (absolute certainty)

Bayesian rule

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}$$

where:

$p(A|B)$ is the conditional probability that event A occurs given that event B has occurred;

$p(B|A)$ is the conditional probability of event B occurring given that event A has occurred;

$p(A)$ is the probability of event A occurring;

$p(B)$ is the probability of event B occurring.

If the occurrence of event *A* depends on only two mutually exclusive events, *B* and NOT *B*, we obtain:

$$p(A) = p(A|B) \times p(B) + p(A|\neg B) \times p(\neg B)$$

where $\neg$ is the logical function NOT.

Similarly,

$$p(B) = p(B|A) \times p(A) + p(B|\neg A) \times p(\neg A)$$

Substituting this equation into the Bayesian rule yields:

$$p\left(A|B\right) = \frac{p\left(B|A\right) \times p\left(A\right)}{p\left(B|A\right) \times p\left(A\right) + p\left(B|\neg A\right) \times p\left(\neg A\right)}$$

The Bayesian rule expressed in terms of hypotheses and evidence looks like this:

$$p(H|E) = \frac{p(E|H) \times p(H)}{p(E|H) \times p(H) + p(E|\neg H) \times p(\neg H)}$$

where:

$p(H)$ is the prior probability of hypothesis H being true;

$p(E|H)$ is the probability that hypothesis H being true will result in evidence E;

$p(\neg H)$ is the prior probability of hypothesis H being false;

$p(E|\neg H)$ is the probability of finding evidence E even when hypothesis H is false.

- In practice, conditional independence among different evidences is assumed:

$$p\left(H_i \middle| E_1\, E_2 \ldots E_n\right) = \frac{p\left(E_1 \middle| H_i\right) \times p\left(E_2 \middle| H_i\right) \times \ldots \times p\left(E_n \middle| H_i\right) \times p\left(H_i\right)}{\sum_{k=1}^{m} p\left(E_1 \middle| H_k\right) \times p\left(E_2 \middle| H_k\right) \times \ldots \times p\left(E_n \middle| H_k\right) \times p\left(H_k\right)}$$

example

Let us consider a simple example.

Suppose an expert, given three conditionally independent evidences E_1, E_2 and E_3, creates three mutually exclusive and exhaustive hypotheses H_1, H_2 and H_3, and provides prior probabilities for these hypotheses – $p(H_1)$, $p(H_2)$ and $p(H_3)$, respectively. The expert also determines the conditional probabilities of observing each evidence for all possible hypotheses.

The prior and conditional probabilities

Probability	*Hypothesis*		
	$i = 1$	$i = 2$	$i = 3$
$p(H_i)$	0.40	0.35	0.25
$p(E_1 \| H_i)$	0.3	0.8	0.5
$p(E_2 \| H_i)$	0.9	0.0	0.7
$p(E_3 \| H_i)$	0.6	0.7	0.9

Assume that the AI system first observes evidence E_3. The system computes the posterior probabilities for all hypotheses as

$$p(H_i|E_3) = \frac{p(E_3|H_i) \times p(H_i)}{\sum_{k=1}^{3} p(E_3|H_k) \times p(H_k)}, \qquad i = 1, 2, 3$$

Thus, $p(H_1|E_3) = \dfrac{0.6 \times 0.40}{0.6 \times 0.40 + 0.7 \times 0.35 + 0.9 \times 0.25} = 0.34$

$$p(H_2|E_3) = \frac{0.7 \times 0.35}{0.6 \times 0.40 + 0.7 \times 0.35 + 0.9 \times 0.25} = 0.34$$

$$p(H_3|E_3) = \frac{0.9 \times 0.25}{0.6 \times 0.40 + 0.7 \times 0.35 + 0.9 \times 0.25} = 0.32$$

After evidence E_3 is observed, belief in hypothesis H_2 becomes equal to belief in hypothesis H_1. Belief in hypothesis H_3 also increases and even nearly reaches beliefs in hypotheses H_1 and H_2.

Suppose now the AI system gets evidence E_1. The posterior probabilities are calculated as

$$p(H_i|E_1E_3) = \frac{p(E_1|H_i) \times p(E_3|H_i) \times p(H_i)}{\sum_{k=1}^{3} p(E_1|H_k) \times p(E_3|H_k) \times p(H_k)}, \qquad i = 1, 2, 3$$

Hence,

$$p(H_1|E_1E_3) = \frac{0.3 \times 0.6 \times 0.40}{0.3 \times 0.6 \times 0.40 + 0.8 \times 0.7 \times 0.35 + 0.5 \times 0.9 \times 0.25} = 0.19$$

$$p(H_2|E_1E_3) = \frac{0.8 \times 0.7 \times 0.35}{0.3 \times 0.6 \times 0.40 + 0.8 \times 0.7 \times 0.35 + 0.5 \times 0.9 \times 0.25} = 0.52$$

$$p(H_3|E_1E_3) = \frac{0.5 \times 0.9 \times 0.25}{0.3 \times 0.6 \times 0.40 + 0.8 \times 0.7 \times 0.35 + 0.5 \times 0.9 \times 0.25} = 0.29$$

Hypothesis H_2 has now become the most likely one.

After observing evidence E_2, the final posterior probabilities for all hypotheses are calculated:

$$p\left(H_i \middle| E_1E_2E_3\right) = \frac{p\left(E_1 \middle| H_i\right) \times p\left(E_2 \middle| H_i\right) \times p\left(E_3 \middle| H_i\right) \times p\left(H_i\right)}{\sum_{k=1}^{3} p\left(E_1 \middle| H_k\right) \times p\left(E_2 \middle| H_k\right) \times p\left(E_3 \middle| H_k\right) \times p\left(H_k\right)}, \qquad i = 1, 2, 3$$

$$p\left(H_1 \middle| E_1E_2E_3\right) = \frac{0.3 \times 0.9 \times 0.6 \times 0.40}{0.3 \times 0.9 \times 0.6 \times 0.40 + 0.8 \times 0.0 \times 0.7 \times 0.35 + 0.5 \times 0.7 \times 0.9 \times 0.25} = 0.45$$

$$p\left(H_2 \middle| E_1E_2E_3\right) = \frac{0.8 \times 0.0 \times 0.7 \times 0.35}{0.3 \times 0.9 \times 0.6 \times 0.40 + 0.8 \times 0.0 \times 0.7 \times 0.35 + 0.5 \times 0.7 \times 0.9 \times 0.25} = 0$$

$$p\left(H_3 \middle| E_1E_2E_3\right) = \frac{0.5 \times 0.7 \times 0.9 \times 0.25}{0.3 \times 0.9 \times 0.6 \times 0.40 + 0.8 \times 0.0 \times 0.7 \times 0.35 + 0.5 \times 0.7 \times 0.9 \times 0.25} = 0.55$$

Although the initial ranking was H_1, H_2 and H_3, only hypotheses H_1 and H_3 remain under consideration after all evidences (E_1, E_2 and E_3) were observed.

Certainty factors theory and evidential reasoning

- Certainty factors theory is a popular alternative to Bayesian reasoning.
- A **certainty factor** (*cf*), a number to measure the expert's belief. The maximum value of the certainty factor is, say, +1.0 (definitely true) and the minimum –1.0 (definitely false). For example, if the expert states that some evidence is almost certainly true, a *cf* value of 0.8 would be assigned to this evidence.

- The certainty factors theory is based on two functions: measure of belief *MB*(*H*,*E*), and measure of disbelief *MD*(*H*,*E*).

$$MB(H,E) = \begin{cases} 1 & \text{if } p(H) = 1 \\ \dfrac{max\{p(H|E), p(H)\} - p(H)}{1 - p(H)} & \text{otherwise} \end{cases}$$

$$MD(H,E) = \begin{cases} 1 & \text{if } p(H) = 0 \\ \dfrac{p(H) - min\{p(H|E), p(H)\}}{p(H)} & \text{otherwise} \end{cases}$$

p(*H*) is the prior probability of hypothesis *H* being true;
p(*H*|*E*) is the probability that hypothesis *H* is true given evidence *E*.

- The values of *MB*(*H*, *E*) and *MD*(*H*, *E*) range between 0 and 1. The strength of belief or disbelief in hypothesis *H* depends on the kind of evidence *E* observed. Some facts may increase the strength of belief, but some increase the strength of disbelief.
- **The total strength of belief or disbelief in a hypothesis:**

$$cf = \frac{MB(H,E) - MD(H,E)}{1\text{-}\, min\{MB(H,E), MD(H,E)\}}$$

example

Consider an AI system that uses rules for reasoning:

IF *A* is *X*
THEN *B* is *Y*

If there is no absolute certain that this rule holds; suppose it has been observed that in some cases, even when the IF part of the rule is satisfied and *object A* takes on *value X*, object *B* can acquire some different value *Z*.

IF *A* is *X*
THEN *B* is *Y* {*cf* 0.7};
B is *Z* {*cf* 0.2}

- The certainty factor assigned by a rule is propagated through the reasoning chain. This involves establishing the net certainty of the rule consequent when the evidence in the rule antecedent is uncertain:

  $$cf(H,E) = cf(E) \times cf$$

  For example,

  IF sky is clear
  THEN the forecast is sunny {*cf* 0.8}

  and the current certainty factor of *sky is clear* is 0.5, then

  $$cf(H,E) = 0.5 \cdot 0.8 = 0.4$$

  This result can be interpreted as *"It may be sunny"*.

- For conjunctive rules such as

IF <evidence E_1>
⋮
AND <evidence E_n>
THEN <hypothesis H> {cf}

the certainty of hypothesis *H*, is established as follows:

$$\boldsymbol{cf(H, E_1 \cap E_2 \cap \ldots \cap E_n) = min\{cf(E_1), cf(E_2), \ldots, cf(E_n)\} \times cf}$$

For example,

IF sky is clear
AND the forecast is sunny
THEN the action is 'wear sunglasses' {*cf* 0.8}

and the certainty of *sky is clear* is 0.9 and the certainty of the *forecast of sunny* is 0.7, then

$cf(H, E_1 \cap E_2) = min\{0.9, 0.7\} \text{ x } 0.8 = 0.7 \text{ x } 0.8 = 0.56$

- For disjunctive rules such as

IF $\quad$ <evidence E_1>
$\vdots$
OR $\quad$ <evidence E_n>
THEN <hypothesis H> {cf}

the certainty of hypothesis *H*, is established as follows:

$$cf(H, E_1 \cup E_2 \cup \ldots \cup E_n) = max\{cf(E_1), cf(E_2), \ldots, cf(E_n)\} \times cf$$

For example,

IF sky is overcast
OR the forecast is rain
THEN the action is 'take an umbrella' {*cf* 0.9}

and the certainty of *sky is overcast* is 0.6 and the certainty of the *forecast of rain* is 0.8, then

$$cf(H, E_1 \cup E_2) = max\{0.6, 0.8\} \text{ x } 0.9 = 0.8 \text{ x } 0.9 = 0.72$$

- When the same consequent is obtained as a result of the execution of two or more rules, the individual certainty factors of these rules must be merged to give a combined certainty factor for a hypothesis.

Suppose the AI system uses the following rules:

Rule 1: IF *A* is *X*
THEN *C* is *Z* {*cf* 0.8}

Rule 2: IF *B* is *Y*
THEN *C* is *Z* {*cf* 0.6}

What certainty should be assigned to object *C* having value *Z* if both *Rule* 1 and *Rule* 2 are fired?

Common sense suggests that, if we have two pieces of evidence (*A* is *X* and *B* is *Y*) from different sources (*Rule* 1 and *Rule* 2) supporting the same hypothesis (*C* is *Z*), then the confidence in this hypothesis should increase and become stronger than if only one piece of evidence had been obtained.

To calculate a combined certainty factor we can use the following equation:

$$cf(cf_1, cf_2) = \begin{cases} cf_1 + cf_2 \times (1 - cf_1) & \text{if } cf_1 > 0 \text{ and } cf_2 > 0 \\[2ex] \dfrac{cf_1 + cf_2}{1 - min\left[|cf_1|,\, |cf_2|\right]} & \text{if } cf_1 < 0 \text{ or } cf_2 < 0 \\[2ex] cf_1 + cf_2 \times (1 + cf_1) & \text{if } cf_1 < 0 \text{ and } cf_2 < 0 \end{cases}$$

where:

cf_1 is the confidence in hypothesis H established by *Rule* 1;

cf_2 is the confidence in hypothesis H established by *Rule* 2;

$|cf_1|$ and $|cf_2|$ are absolute magnitudes of cf_1 and cf_2, respectively.

The certainty factors theory provides a ***practical*** alternative to Bayesian reasoning. The heuristic manner of combining certainty factors is different from the manner in which they would be combined if they were probabilities. The certainty theory is not “mathematically pure” but does mimic the thinking process of a human expert.

When reliable statistical information is not available or we cannot assume the conditional independence of evidence the **Bayesian** method becomes unsuitable. This issue motivated the development of the **certainty factors** theory.

- **Certainty factors** are used in cases where the probabilities are not known or are too difficult or expensive to obtain. The evidential reasoning mechanism can manage incrementally acquired evidence, the conjunction and disjunction of hypotheses, as well as evidences with different degrees of belief.

Reasoning with fuzzy logic

- **How can we represent 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.

- 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**.

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.

Examples of fuzzy logic applications

Fuzzy Logic: An Introduction
https://www.youtube.com/watch?v=P8wY6mi1vV8
(ABS) https://www.youtube.com/watch?v=rln_kZbYaWc
https://ieeetv.ieee.org/ieeetv-specials/the-sorites-paradox-introduction-to-fuzzy-logic?

An Egg-Boiling Fuzzy Logic Robot
https://www.youtube.com/watch?v=J_Q5X0nTmrA&list=WL&index=33

NASA | Fuzzy Logic Models for Real-Time Simulations
https://www.youtube.com/watch?v=GkoTFoWJvOg

Object Tracking with a 2 DOF Robot Arm Using Fuzzy Logic
https://www.youtube.com/watch?v=QsNcOCI0-Sc&index=36&list=WL

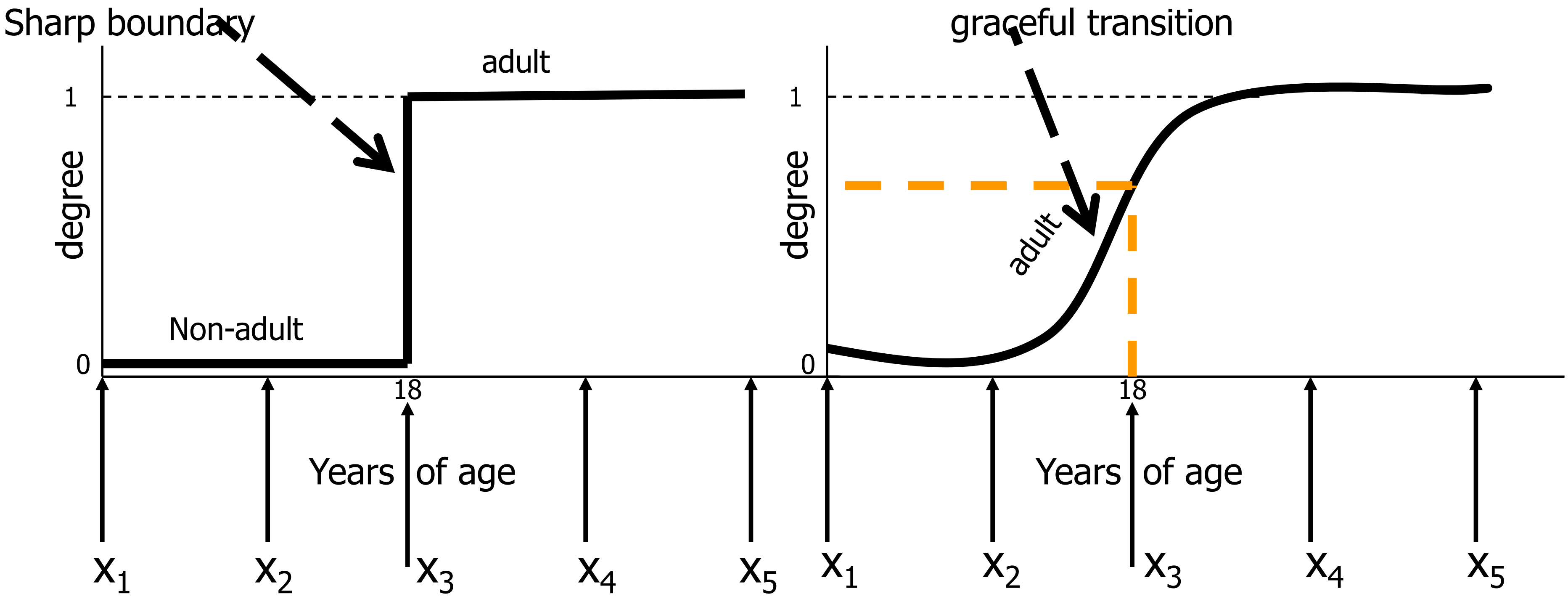
https://en.wikipedia.org/wiki/Sendai_Subway_1000_series

Crisp sets and fuzzy sets

Crisp set: an element x either belong to X or does not belong to X.

Fuzzy set: an element x belongs to a set with a certain degree of membership

$X = \{x1,x2,x3,x4,x5\}$



Operations on fuzzy sets

Intersection (symbols: ∩; AND; min)

Crisp sets: which element belongs to both sets?

$$\{1,2,3\} \cap \{3,4,5\} = \{3\}$$

Fuzzy sets: how much of the element is in both sets?

Tom is *tall* and *average*

tall men={0/165,0/175,0/180,0.25/182.5,0.5/185,1/190}

average men= {0/165,1/175,0.5/180,0.25/182.5,0/185,0/190}

$tall \cap average = min\,[\mu_{tall},\mu_{average}] = \mu_{tall} \cap \mu_{average} =$

= {0/165,0/175,0/180,0.25/182.5,0/185,0/190}

Operations on fuzzy sets

Union (symbols: U; OR; max)

Crisp sets: which element belongs to both sets?

{1,2,3} U {3,4,5} = {1,2,3,4,5}

Fuzzy sets: how much of the element is in both sets?

Tom is *tall* OR *average*

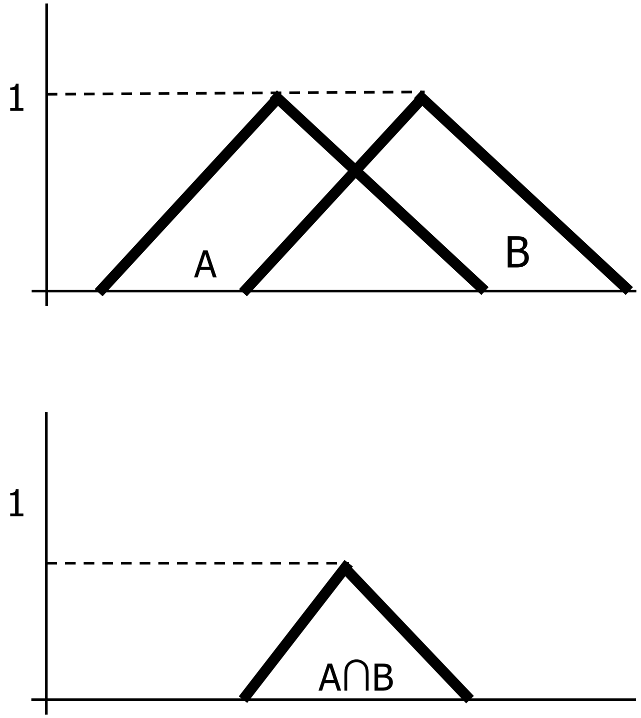
tall men={0/165,0/175,0/180,0.25/182.5,0.5/185,1/190}

average men= {0/165,1/175,0.5/180,0.25/182.5,0/185,0/190}

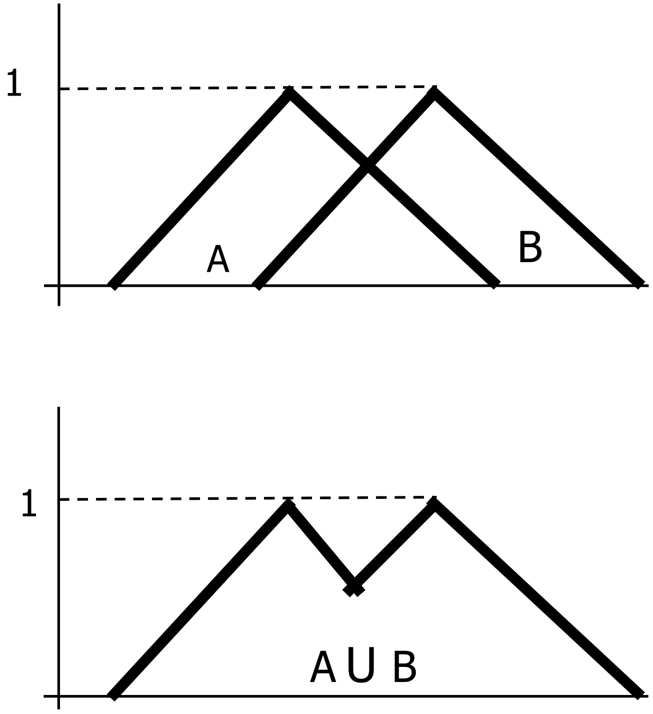
tall U *average = max* $[\mu_{tall},\mu_{average}] = \mu_{tall} \cup \mu_{average} =$

= {0/165,1/175,0.5/180,0.25/182.5,0.5/185,1/190}

Operations on fuzzy sets graphically



Intersection



Union

Linguistic variables

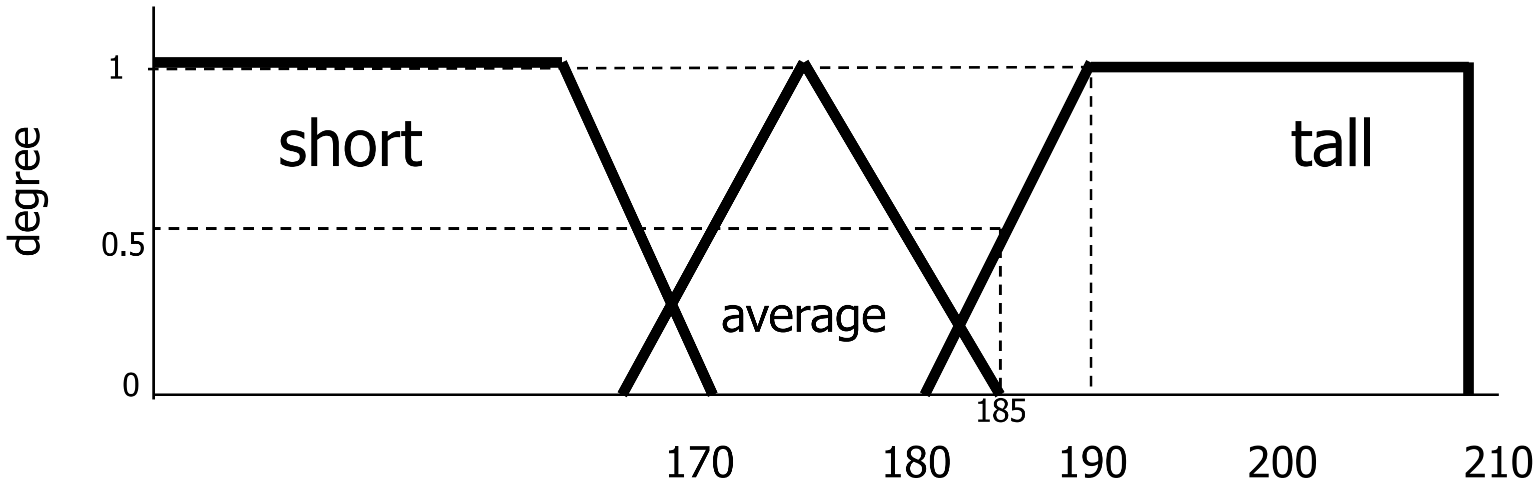
A LINGUISTIC VARIABLE is a fuzzy variable.

The range of all possible real values that a linguistic variable can take represents the **universe of discourse** of that variable.

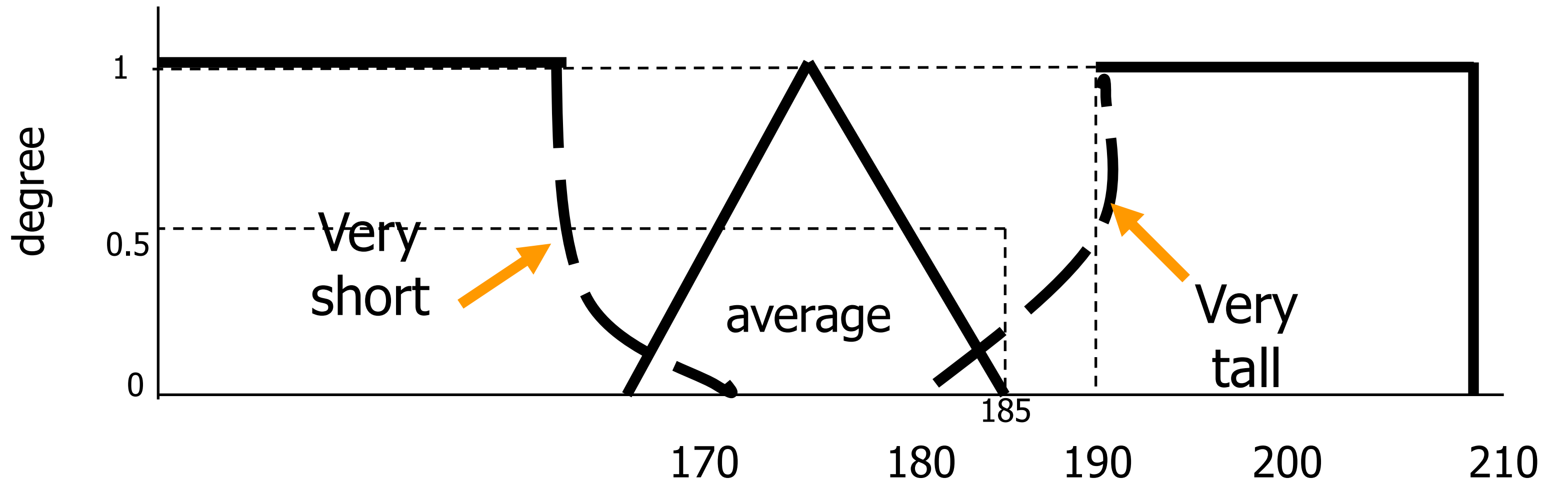
Linguistic variable *Speed*: might have a range 0-200 miles per hour. It might also have a set of **linguistic values**-*very slow, slow, medium, fast, very fast.* Each linguistic value is a fuzzy subset.

Hedges are terms that modify the shape of fuzzy sets: *very, more, quite, less, slightly, likely, very likely*

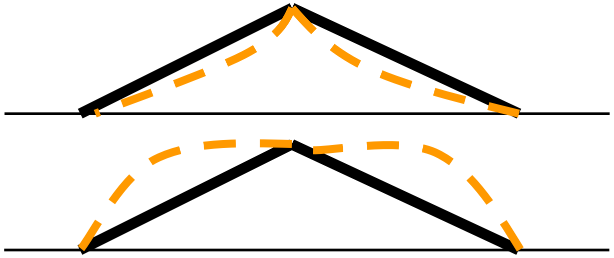
Linguistic variables



Linguistic variables



Mathematical formulation:



$\mu^{very}(x)=[\mu(x)]^2$

$\mu^{more\ or\ less}(x)=\sqrt{}[\mu(x)]$

$\mu^{extremely}(x) = [\mu(x)]^3$

$\mu^{slightly}(x) =[\mu(x)]^{1.7}$

Example:

If Tom has membership (0.86) in the set of TALL men, he will have a membership of $(0.86)^2=0.7396$ in the set of VERY TALL men and √(0.86)=0.92 in the set of MORE or LESS TALL MEN

Linguistic variables: some examples

HOW CAN WE DECIDE ON THE RANGE AND THE TERMS USED?

Reading speed*,* counted from user's self-paced reading, can be determined by computing the average reading time per 100 word.

Use universe of discourse that is from 15 to 65 sec.

It can be associated with linguistic values {Slow, Medium, Fast}. Medium speed is around 35 sec

Fuzzy rules

Linguistic variables are used in fuzzy rules.

Example: The rate of calories burn for someone taking exercise on a bike in a gym

Fuzzy rule

IF **speed**= *high* AND **load**= *high* THEN **burn** = *very fast*

IF speed= moderate AND load= high THEN burn= fast

- 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 systems

HOW TO BUILD IT

- Specify the problem and define linguistic variables;
- Determine fuzzy sets;
- Elicit and construct fuzzy rules;
- Encode fuzzy sets, rules and procedures;
- Evaluate and tune the fuzzy system

Fuzzy systems: first applications

In 1975, Professor Ebrahim Mamdani at the Electronic Engineering Department of Queen Mary & Westfield College of the University of London 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.

- *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.

- Later, Canon developed an auto focusing camera:
  - It used a charge-coupled device (CCD) to measure the clarity of the image in six regions of its field of view and use the information provided to determine if the image is in focus.
  - It tracked the rate of change of lens movement during focusing, and controlled its speed to prevent overshoot.

The camera's fuzzy control system used:

- 12 inputs: 6 to obtain the current clarity data provided by the CCD and 6 to measure the rate of change of lens movement.
- 1 output: the position of the lens.
- Rule-base: 13 rules
- Memory requirements: 1.1 kilobytes

- Mitsubishi: industrial air conditioner used 25 heating rules and 25 cooling rules.
  - A temperature sensor provides input, with control outputs fed to an inverter, a compressor valve, and a fan motor.
  - Compared to the previous design, the fuzzy controller heats and cools five times faster, reduces power consumption by 24%, increases temperature stability by a factor of two, and uses fewer sensors.

Fuzzy systems

HOW TO BUILD IT

- Specify the problem and define linguistic variables;
- Determine fuzzy sets;
- Elicit and construct fuzzy rules;
- Encode fuzzy sets, rules and procedures;
- Evaluate and tune the fuzzy system

Specify the problem and define linguistic variables

Describe the problem in terms of knowledge engineering, i.e. determine inputs and output variables and their ranges

Example

Problem: *automate the wash time when using a washing machine.*

Inputs: *dirt and grease of the clothes to be washed*

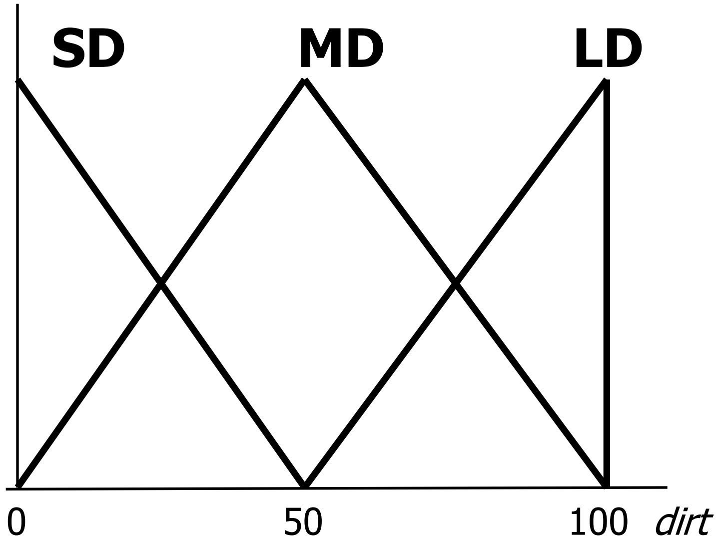
Output: *wash time*

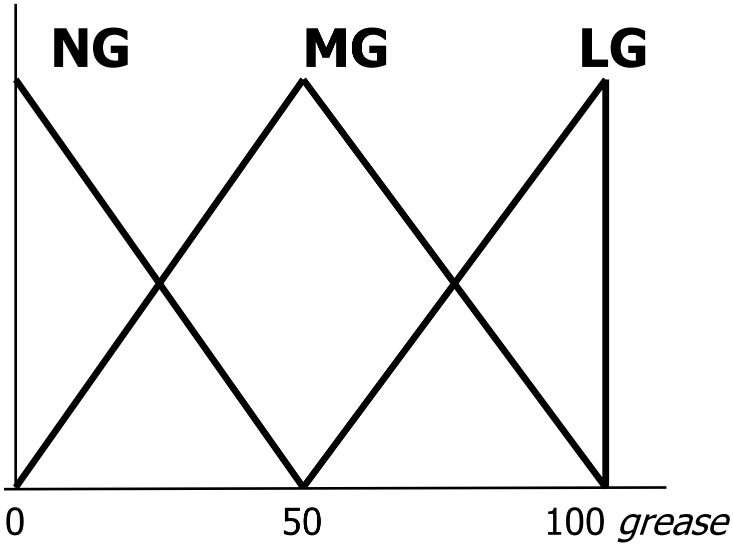
Determine fuzzy sets-1

dirt = {**SD** (small dirt), **MD** (medium dirt), **LD** (large dirt)}

grease={**NG** (no grease), **MG** (medium grease), **LG** (large grease)}

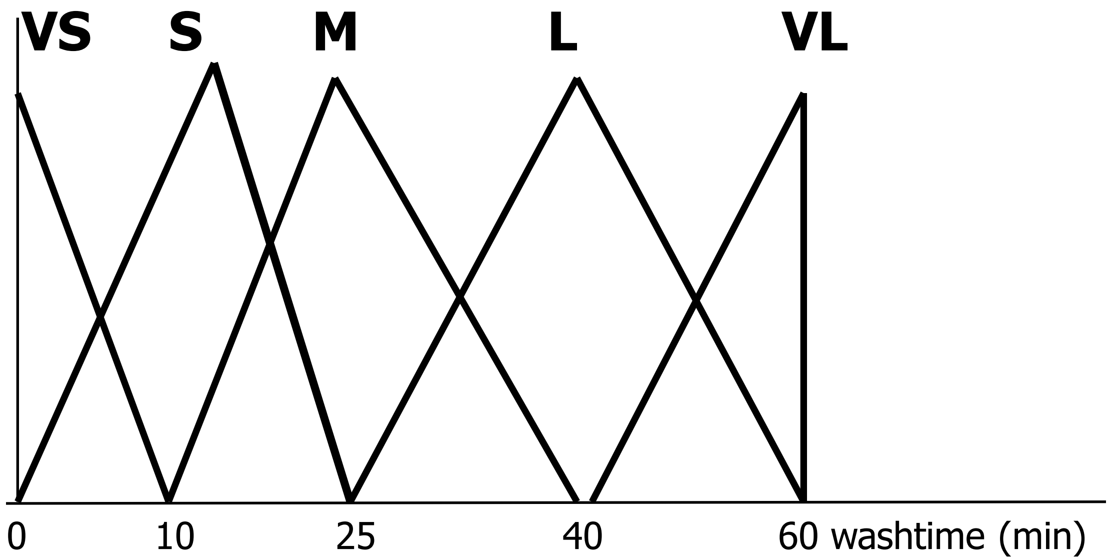
washtime={**VS**(very short),**S**(short),**M**(medium),**L**(long),**VL**(very long)}





Determine fuzzy sets-2

washtime={**VS**(very short),**S**(short),**M**(medium),**L**(long),**VL**(very long)}



Elicit and construct fuzzy rules

Experts supply the following rules for the washing machine

		grease		
		NG	**MG**	**LG**
dirt	**SD**	**VS**	**M**	**L**
	MD	**S**	**M**	**L**
	LD	**M**	**L**	**VL**

If dirt of clothes is **small** and there is **no grease** then washtime is **very short**

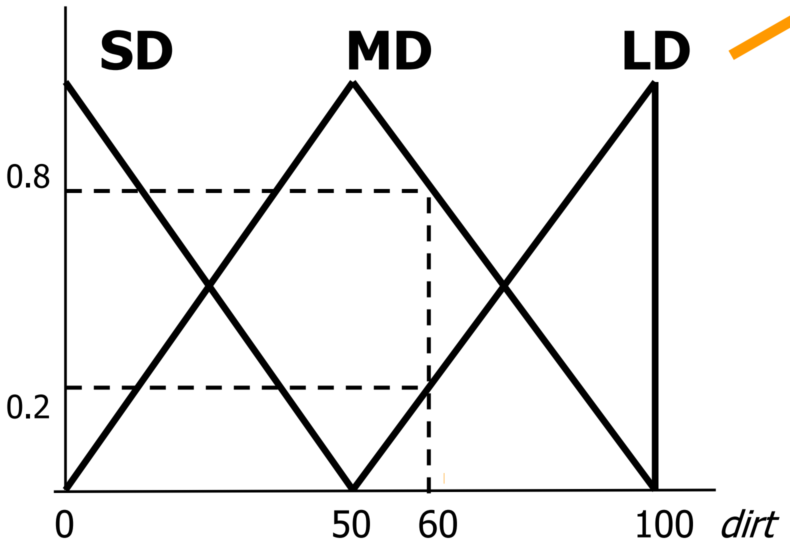
If dirt of clothes is **large** and there is **large grease** then washtime is **very long**

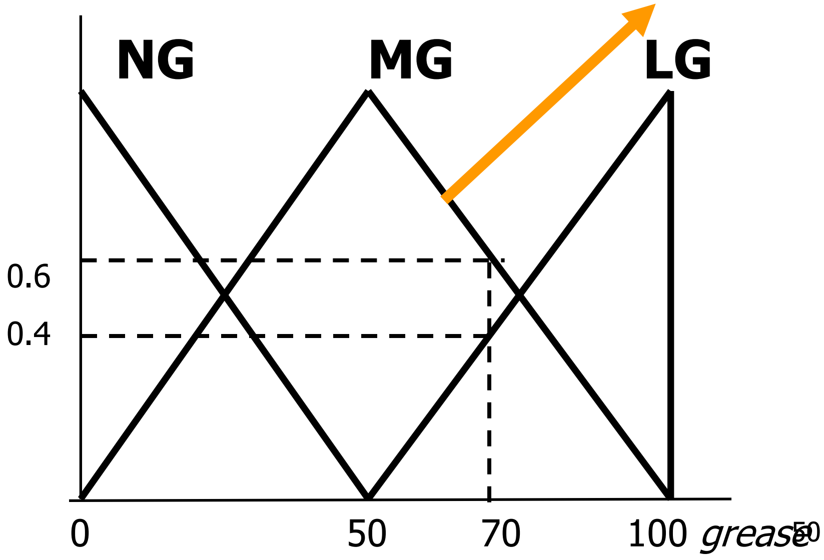
Encode sets, rules and procedures-1

1. Rule evaluation

Assume input is: *dirt* = 60, *grease* = 70

		grease		
		0	**MG**(70)	**LG**(70)
dirt	**0**	**0**	**0**	**0**
	MD (60)	**0**	**M**	**L**
	LD (60)	**0**	**L**	**VL**





Encode sets, rules and procedures-2

2. Conflict resolution

dirt	*grease*			
		0	**MG**(70)=0.6	**LG**(70)=0.4
	0	**0**	**0**	**0**
	MD (60)=0.8	**0**	**M**	**L**
	LD (60)=0.2	**0**	**L**	**VL**

If x is **MD** and y is **MG** then z is **M**

If x is **MD** and y is **LG** then z is **L**

If x is **LD** and y is **MG** then z is **L**

If x is **LD** and y is **LG** then z is **VL**

Encode sets, rules and procedures-3

Intersection (symbols: ∩; AND; min)

Crisp sets: which element belongs to both sets?

{1,2,3} ∩ {3,4,5} = {3}

Fuzzy sets: how much of the element is in both sets?

Tom is *tall* and *average*

tall men={0/165,0/175,0/180,0.25/182.5,0.5/185,1/190}

average men= {0/165,1/175,0.5/180,0.25/182.5,0/185,0/190}

tall ∩ *average* = *min* [μ_{tall},$\mu_{average}$]= $\mu_{tall} \cap \mu_{average}$=

= {0/165,0/175,0/180,0.25/182.5,0/185,0/190}

Encode sets, rules and procedures-4

If x is **MD** and y is **MG**

$\rightarrow \mathit{min}[\mu_{MD}(60), \mu_{MG}(70)]=\mathit{min}(0.8,0.6)=\mathbf{0.6}$

If x is **MD** and y is **LG**

$\rightarrow \mathit{min}[\mu_{MD}(60), \mu_{LG}(70)]=\mathit{min}(0.8,0.4)=\mathbf{0.4}$

If x is **LD** and y is **MG** → **0.2**
If x is **LD** and y is **LG** → **0.2**

Results shown are after calculating the if-part of the rules

		grease		
		0	**MG**(70)=0.6	**LG**(70)=0.4
dirt	**0**	**0**	**0**	**0**
	MD (60)=0.8	**0**	**0.6**	**0.4**
	LD (60)=0.2	**0**	**0.2**	**0.2**

Encode sets, rules and procedures-5

3. Aggregation of rule consequences (i.e. find the output of the four rules)

		grease		
dirt		0	**MG**(70)=0.6	**LG**(70)=0.4
	0	0	0	0
	MD (60)=0.8	0	0.6	0.4
	LD (60)=0.2	0	0.2	0.2

		grease		
dirt		0	**MG**(70)=0.6	**LG**(70)=0.4
	0	**0**	**0**	**0**
	MD (60)=0.8	**0**	**min (0.6,μ_M)**	**min (0.4,μ_L)**
	LD (60)=0.2	**0**	**min (0.2,μ_L)**	**min (0.2,μ_{VL})**

If x is **MD** and y is **MG** → then z is **M**

If x is **MD** and y is **LG** → then z is **L**

If x is **LD** and y is **MG** → then z is **L**

If x is **LD** and y is **LG** → then z is **VL**

Encode sets, rules and procedures-6

Union (symbols: U; OR; max)

Crisp sets: which element belongs to both sets?

{1,2,3} U {3,4,5} = {1,2,3,4,5}

Fuzzy sets: how much of the element is in both sets?

Tom is *tall* OR *average*

tall men={0/165,0/175,0/180,0.25/182.5,0.5/185,1/190}

average men= {0/165,1/175,0.5/180,0.25/182.5,0/185,0/190}

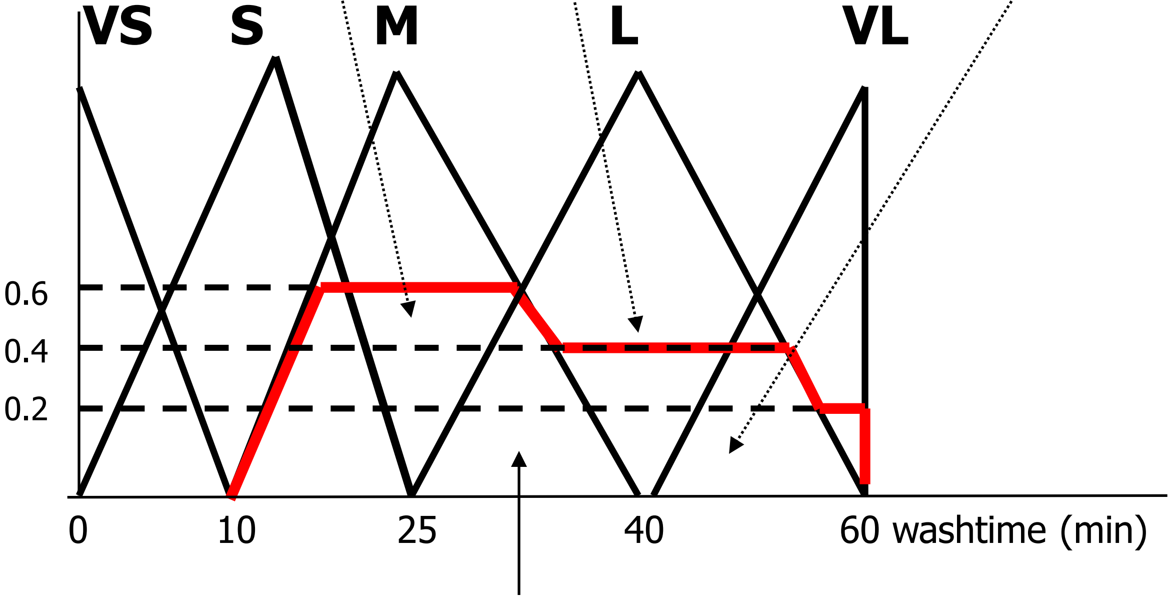
tall U *average* = *max* $[\mu_{tall},\mu_{average}] = \mu_{tall} \cup \mu_{average} =$

= {0/165,1/175,0.5/180,0.25/182.5,0.5/185,1/190}

Encode sets, rules and procedures-7

Union of three fuzzy numbers

$$\mu_{agg} = max\{ \min(0.6, \mu_M), \min(0.4, \mu_L), \min(0.2, \mu_{LV}) \}$$

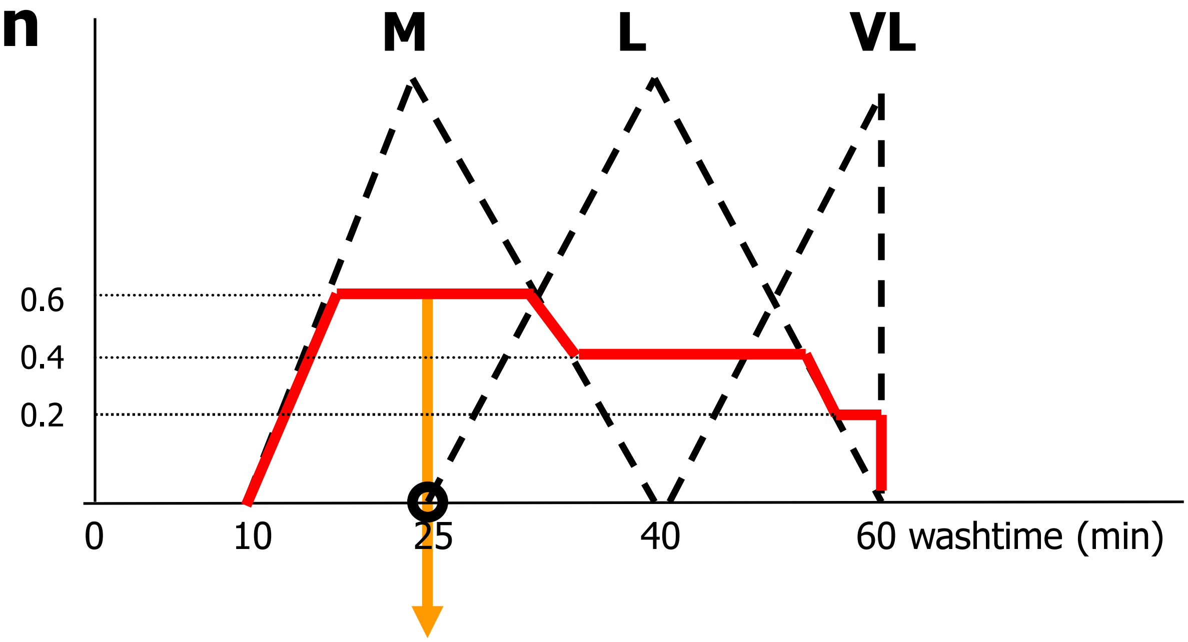


area of overlap

Encode sets, rules and procedures-8

4. Defuzzification

Fuzziness helps us to evaluate the rules but the final output of a fuzzy system has to be crisp number



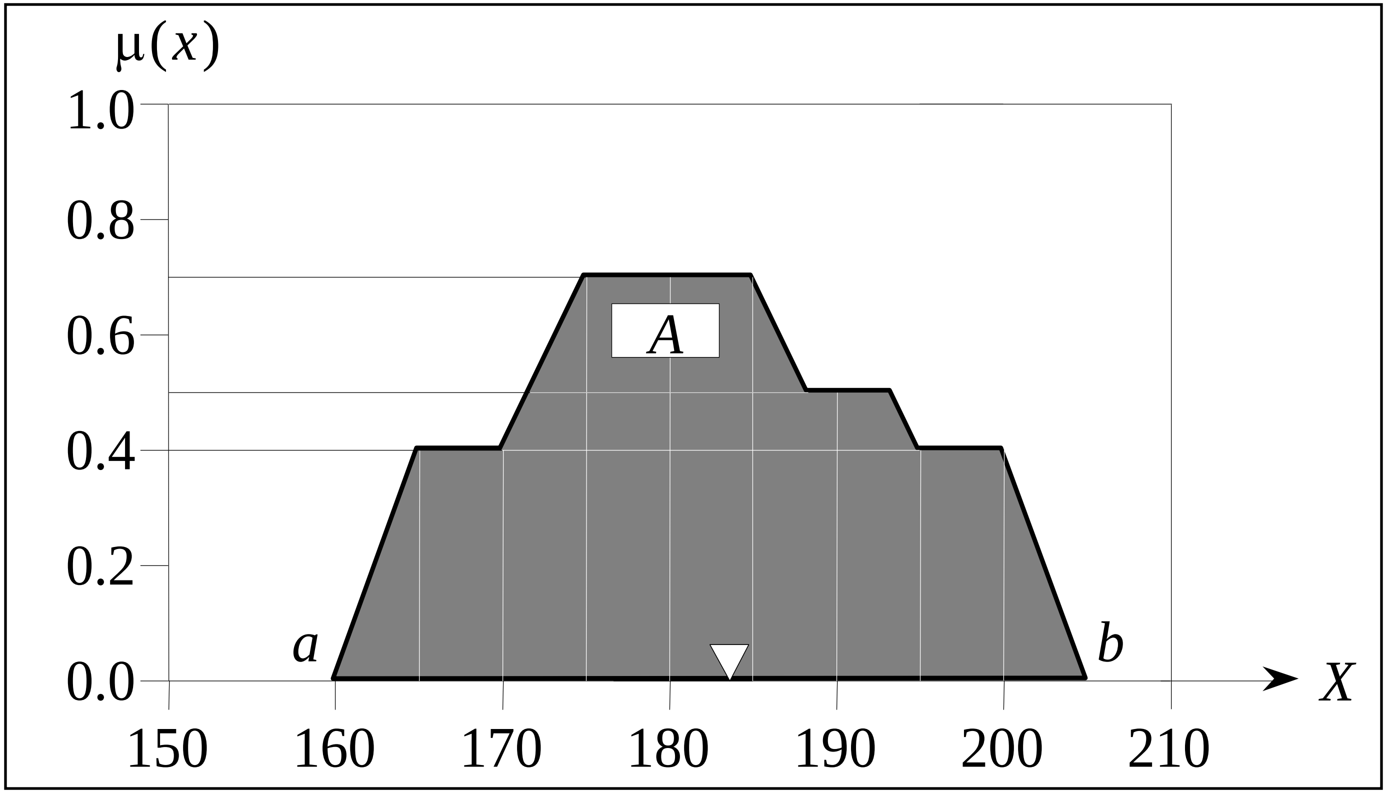
The output of the defuzzification process is the aggregate output fuzzy set and the output is a single number (mean of maximum method; centre of gravity method)

Centre of gravity defuzzification 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:

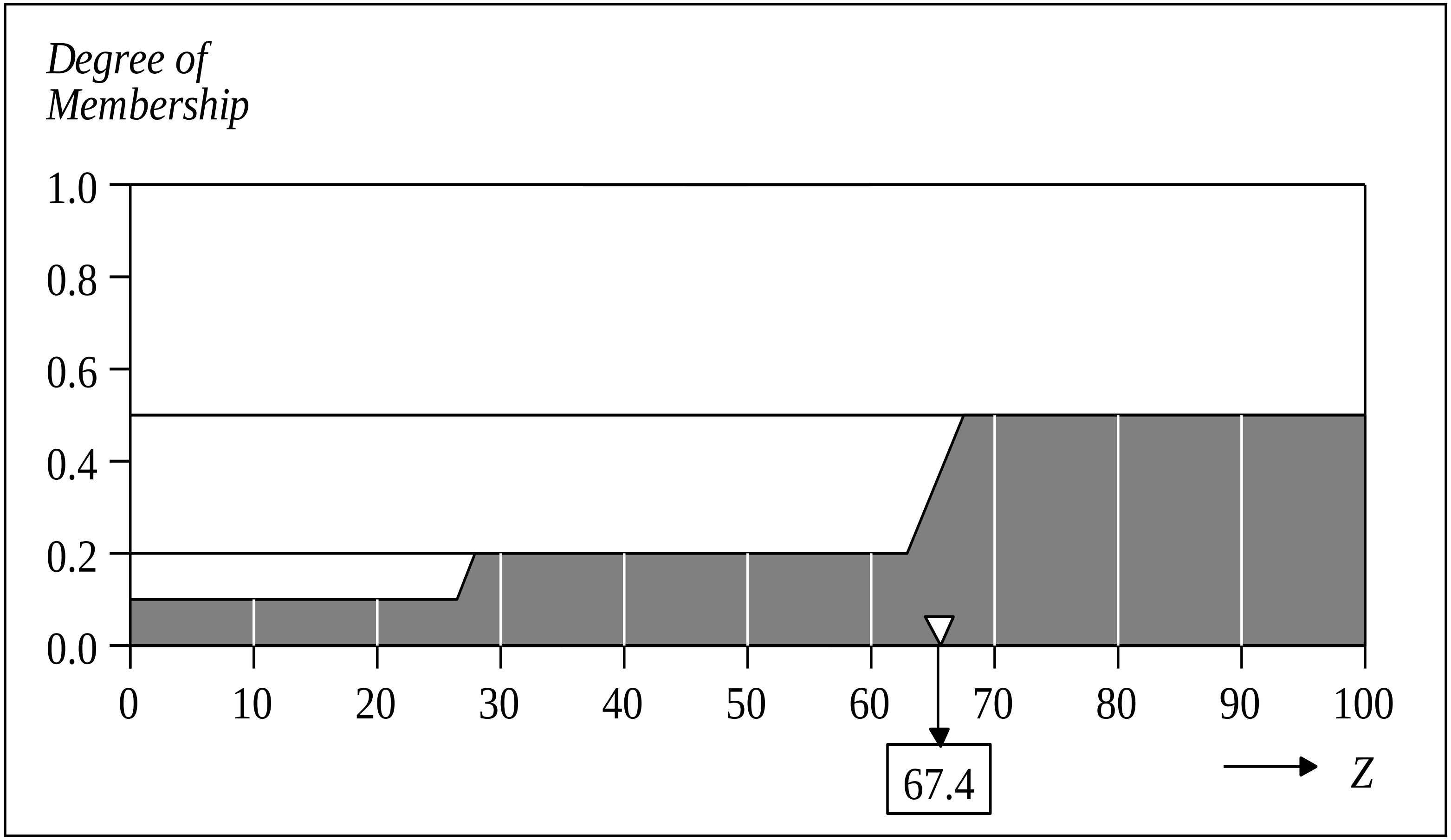
$$COG = \frac{\int_a^b \mu_A(x)\, x\, dx}{\int_a^b \mu_A(x)\, dx}$$

- Centroid defuzzification method (approximates COG) 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) using the centroid method: an example

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



Evaluate and tune the system

1. Review model input and output variables, and if required redefine their ranges.
2. Review the fuzzy sets, and if required define additional sets on the universe of discourse. The use of wide fuzzy sets may cause the fuzzy system to perform roughly.
3. Provide sufficient overlap between neighbouring sets. It is suggested that triangle-to-triangle and trapezoid-to-triangle fuzzy sets should overlap between 25% to 50% of their bases.

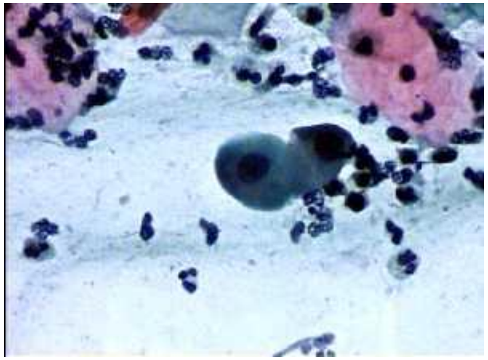
4. Review existing rules, and if required add new rules to the rule base.

5. Examine rule base for opportunities to write hedge rules to capture the pathological behaviour of the system.

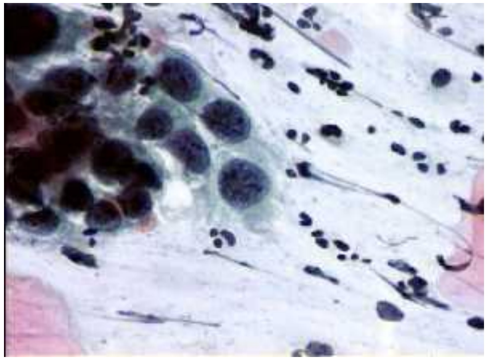
6. Adjust rule execution weights. Most fuzzy logic tools allow control of the importance of rules by changing a weight multiplier.

7. Revise shapes of fuzzy sets. In most cases, fuzzy systems are highly tolerant of a shape approximation.

Fuzzy clustering

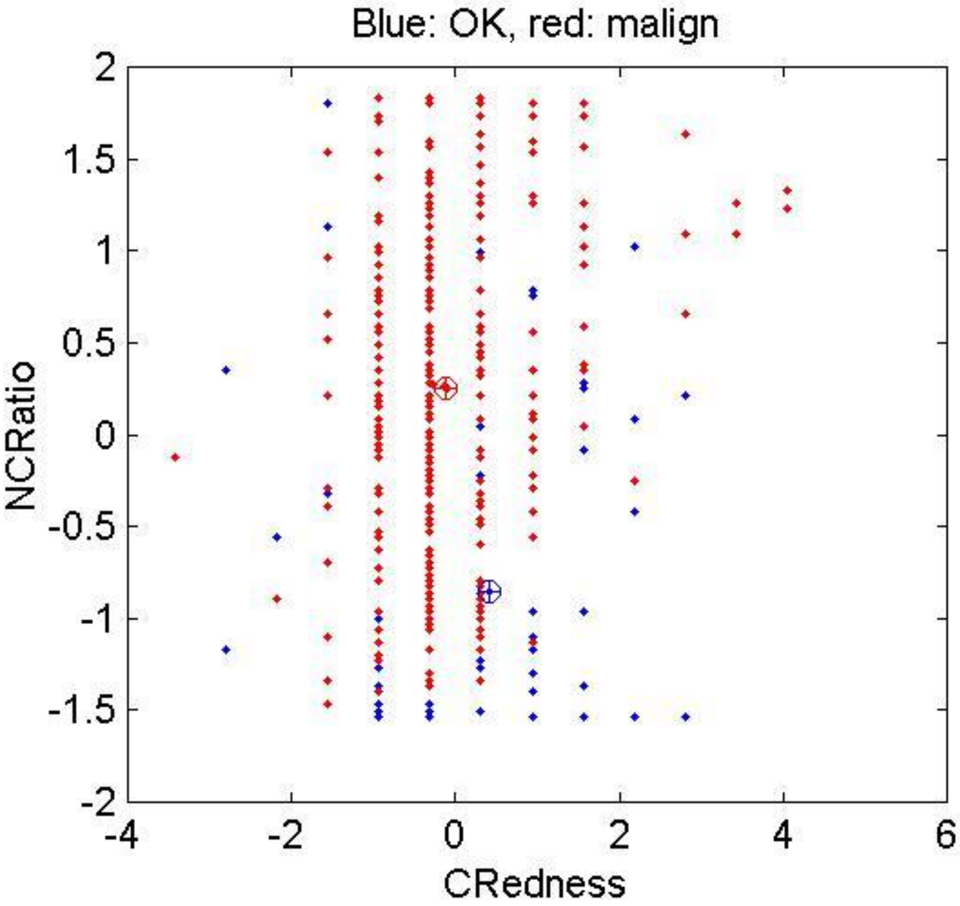


Normal smear



Severely dysplastic smear

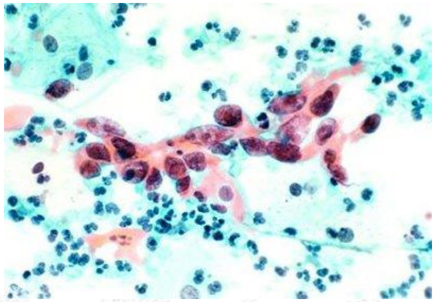
The term pap-smear refers to samples of human cells stained by the so-called Papanicolaou method. The purpose of the Papanicolaou method is to diagnose pre-cancerous cell changes before they progress to invasive carcinoma.



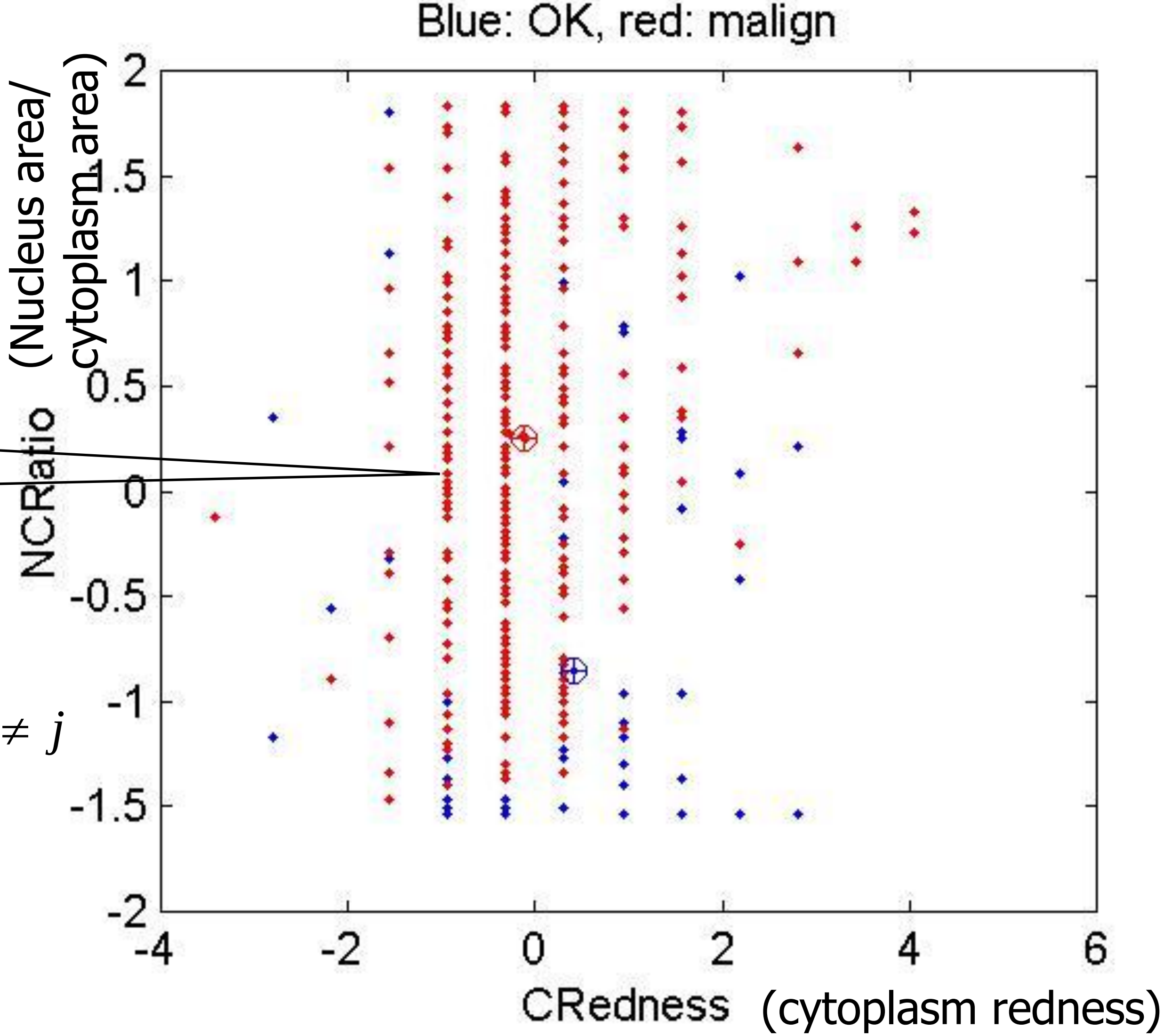
Nucleus and cytoplasm area
Nucleus and cyto brightness

What happens here?

Fuzzy Clustering



The papsmear problem



Looks like clusters overlap! A cell can belong to two classes to a degree

$$\bigcup_{i=1}^{c} C_i = U$$

$$C_i \xcancel{\cap C_j} = \emptyset \qquad for\ all\ i \neq j$$

$$\emptyset \subset C_i \subset U \qquad for\ all\ i$$

$$2 \leq c \leq K$$

Simple example: classify cracked tiles

Tiles are made from clay moulded into the right shape, brushed, glazed, and baked.

Baking may produce invisible cracks.

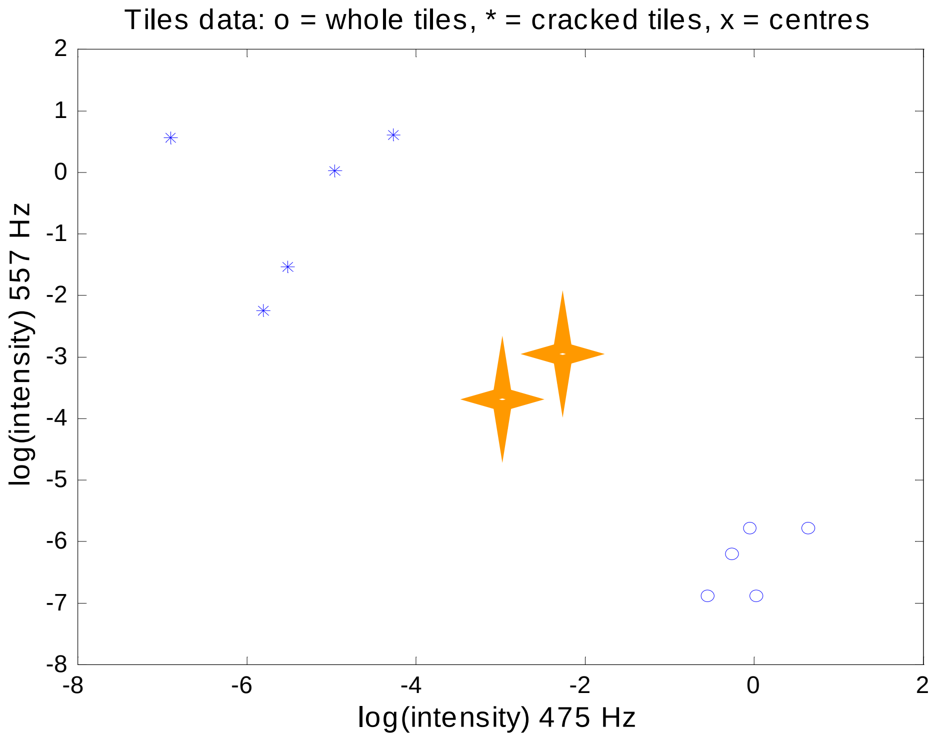
Operators can detect the cracks by hitting the tiles with a hammer, and in an automated system the response is recorded with a microphone, filtered, Fourier transformed, and normalised.

475Hz	557Hz	Ok?
0.958	0.003	Yes
1.043	0.001	Yes
1.907	0.003	Yes
0.780	0.002	Yes
0.579	0.001	Yes
0.003	0.105	No
0.001	1.748	No
0.014	1.839	No
0.007	1.021	No
0.004	0.214	No

Table 1: Sample of frequency intensities for ten tiles.

Fuzzy clustering

Points located between the two cluster centres have a gradual membership of both clusters.



The fuzzified *c-means* algorithm allows each data point to belong to a cluster to a degree specified by a membership grade, and thus each point may belong to several clusters.

Fuzzy clusters and clustering algorithms

Example: Classify cracked tiles

1. Place two cluster centres randomly
2. Assign a fuzzy membership to each data point depending on distance
3. Run Fuzzy C-means algorithm

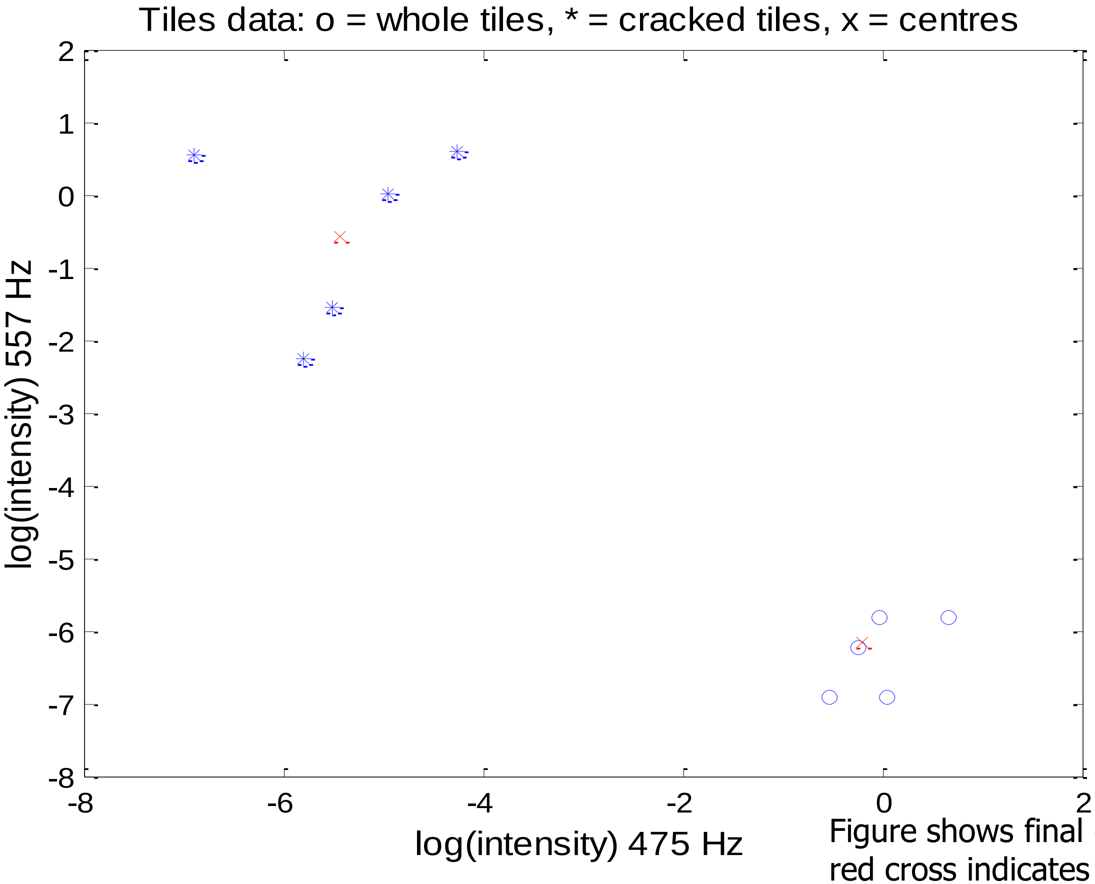


Figure shows final outcome: red cross indicates cluster centre.

Fuzzy clusters and clustering algorithms

The *fuzzy c-means clustering* (FCM) algorithm partitions a collection of N data points specified by m-dimensional vectors $\mathbf{u}_k$ (k=1,2,…, K) into c fuzzy clusters, and finds a cluster centre in each, minimising an objective function.

Fuzzy clusters and clustering algorithms

Fuzzy c-means Algorithm

1. Initialise the membership matrix **M** with random values between 0 and 1.

2. Calculate *c* cluster centres $\mathbf{c}_i$

3. Compute the objective function *J*. Stop if either it is below a certain threshold level or its improvement over the previous iteration is below a certain tolerance.

4. Compute a new **M**

5. Go to step 2.

Fuzzy clusters and clustering algorithms

```
M =

    0.0025    0.9975

    0.0091    0.9909

    0.0129    0.9871

    0.0001    0.9999

    0.0107    0.9893

    0.9393    0.0607

    0.9638    0.0362

    0.9574    0.0426

    0.9906    0.0094

    0.9807    0.0193
```

Example of membership matrix **M** one gets at the end:

1. The last five data points (rows) belong mostly to the first cluster (column)

2. The first five data points (rows) belong mostly to the second cluster (column)

Fuzzy clusters and clustering algorithms

hyper- parameter that controls how fuzzy the cluster will be. The higher it is, the fuzzier the cluster will be in the end.

STEP 2: Calculate cluster centre for *i* cluster

$$\mathbf{c}_i = \frac{\sum_{k=1}^{K} m_{ik}^{q} \mathbf{u}_k}{\sum_{k=1}^{K} m_{ik}^{q}}$$

STEP 3: Objective function

$$\min \boldsymbol{J}(\mathbf{M}, \mathbf{c}_1, \mathbf{c}_2, \mathrm{K}\ \mathbf{c}_c) = \min \sum_{\boldsymbol{i}=1}^{\boldsymbol{c}} \boldsymbol{J}_{\boldsymbol{i}} = \min \sum_{\boldsymbol{i}=1}^{\boldsymbol{c}} \sum_{\boldsymbol{k}=1}^{\boldsymbol{K}} \boldsymbol{m}_{\boldsymbol{ik}}^{\boldsymbol{q}} \boldsymbol{d}_{\boldsymbol{ik}}^{2}$$

Fuzzy clusters and clustering algorithms

STEP 4: Fuzzy membership matrix **M** with elements

Point *k*'s membership of cluster *i*

Fuzziness exponent

$$m_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(q-1)}}$$

Distance from point *k* to current cluster centre *i*

Distance from point *k* to other cluster centres *j*

$$d_{ik} = \left\| \mathbf{u}_k - \mathbf{c}_i \right\|$$

Fuzzy clusters and clustering algorithms

$$m_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{2/(q-1)}}$$

A large *q* results in smaller membership values and hence, fuzzier clusters. In the limit *q*=1, the membership degrees converge towards 0 or 1, which implies the partitioning approximates a crisp partitioning. It can be fine-tuned through experiments or by exploiting domain knowledge; otherwise *q* is set to 2.

$$= \frac{1}{\left(\frac{d_{ik}}{d_{1k}}\right)^{2/(q-1)} + \left(\frac{d_{ik}}{d_{2k}}\right)^{2/(q-1)} + \ldots + \left(\frac{d_{ik}}{d_{ck}}\right)^{2/(q-1)}}$$

$$= \frac{\frac{1}{d_{ik}{}^{2/(q-1)}}}{\frac{1}{d_{1k}{}^{2/(q-1)}} + \frac{1}{d_{2k}{}^{2/(q-1)}} + \ldots + \frac{1}{d_{ck}{}^{2/(q-1)}}}$$

Gravitation to cluster *i* relative to total gravitation

Fuzzy clusters and clustering algorithms

Fuzzy c-means Algorithm

1. Initialise the membership matrix **M** with random values between 0 and 1.

2. Calculate c cluster centres $\mathbf{c}_i$

3. Compute the objective function J. Stop if either it is below a certain threshold level or its improvement over the previous iteration is below a certain tolerance.

4. Compute a new **M** using relation for m_{ik}

5. Go to step 2.

Fuzzy clusters and clustering algorithms

Fuzzy c-partition

All clusters *C* together fill the whole universe *U*.
Remark: The sum of memberships for a data point is 1, and the total for all points is K

$$\bigcup_{i=1}^{c} C_i = U$$

Not valid: Clusters do overlap

$$C_i \cap C_j = \emptyset \quad for\ all\ i \neq j$$

A cluster *C* is never empty and it is smaller than the whole universe *U*

$$\emptyset \subset C_i \subset U \quad for\ all\ i$$

$$2 \leq c \leq K$$

There must be at least 2 clusters in a c-partition and at most as many as the number of data points *K*

Useful reading

- Negnevitsky, *Artificial Intelligence*, chapters 3 and 4.
- Jan Jantzen, Tutorial On Fuzzy Logic, http://webhome.csc.uvic.ca/~mcheng/460/notes/fuzzy_logic.pdf
- M. Hellmann, Fuzzy Logic Introduction, epsilon.nought.de/tutorials/fuzzy/fuzzy.pdf
- Fuzzy logic repository

http://uni-obuda.hu/users/fuller.robert/fuzs.html

- FuzzyLite is a free and open-source fuzzy logic control library programmed in C++ for multiple platforms (e.g., Windows, Linux, Mac, iOS). jfuzzylite is the equivalent library for Java and Android platforms.

https://www.fuzzylite.com/

- FisPro (Fuzzy Inference System Professional) allows to create fuzzy inference systems and to use them for reasoning purposes, especially for simulating a physical or biological system.

http://www7.inra.fr/mia/M/fispro/fispro2013_en.html

Useful reading

- Jan Jantzen et al., Pap-smear Benchmark Data For Pattern Classification, https://www.researchgate.net/publication/265873515_Pap-smear_Benchmark_Data_For_Pattern_Classification
- A Tutorial on Clustering Algorithms- Fuzzy C-Means Clustering http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/cmeans.html
- Rui Xu, Survey of Clustering Algorithms. (2005), IEEE Transactions on Neural Networks, vol. 16, no. 3, 645-678.

https://drive.google.com/open?id=0By995HEqDrWQVlVnaVBNUGlVV1E

Next week

Feature engineering and learning paradigms in machine learning