Unit-5:
Probabilistic **Reasoning**

Computer Engineering Department

Outline

- Approaches to Reasoning
- Probability And Bays’ Theorem
- Bayesian Networks
- Certainty Factors And Rule-Base Systems
- Dempster-Shafer Theory
- Fuzzy Logic

Approaches to Reasoning

- There are three different approaches to reasoning under uncertainties.
    1. Symbolic Reasoning
    2. Statistical Reasoning
    3. Fuzzy logic Reasoning

1. Symbolic Reasoning : Symbolic logic deals with how symbols relate to each other. It assigns symbols to verbal reasoning in order to be able to check the validity of the statements through a mathematical process.

- Propositions:

    A : All spiders have eight legs.

    B : Black widows are a type of spider.

    C : Black widows have eight legs.

    The Λ means "and," and the $\Rightarrow$ symbol means "implies."

- Conclusion: $A \Lambda B \Rightarrow C$

Symbolic Reasoning

- The reasoning is said to be symbolic when it can be performed by means of primitive operations manipulating elementary symbols.
- Usually, symbolic reasoning refers to mathematical logic, more precisely first-order (predicate) logic and sometimes higher orders.

Probability And Bays’ Theorem

Statistical Reasoning

- In the logic based approaches described, we have assumed that everything is either believed false or believed true.
- However, it is often useful to represent the fact that we believe, something is probably true, or true with probability 0.65.
- This is useful for dealing with problems where there is randomness and unpredictability (such as in games of chance) and also for dealing with problems where we could, if we had sufficient information, work out exactly what is true.
- To do all this in a principled way requires techniques for probabilistic reasoning.
- Probability quantifies the uncertainty of the outcomes of a random variable / event.
- Real world applications are probabilistic in nature, and to represent the relationship between multiple events, we need a Bayesian network.

Review of Probability Theory

- Marginal probability is the probability of an event, irrespective of other random variables.
  - Marginal Probability: The probability of an event irrespective of the outcomes of other random variables, e.g. P(A).
- The joint probability is the probability of two (or more) simultaneous events, often described in terms of events A and B from two dependent random variables, e.g. X and Y. The joint probability is often summarized as just the outcomes, e.g. A and B.
  - Joint Probability: Probability of two (or more) simultaneous events, e.g. P(A and B) or P(A, B).
- The conditional probability is the probability of one event given the occurrence of another event, often described in terms of events A and B from two dependent random variables e.g. X and Y.
  - Conditional Probability: Probability of one (or more) event given the occurrence of another event, e.g. P(A given B) or P(A | B).

Review of Probability Theory

- The joint probability can be calculated using the conditional probability :

$$\boxed{P(A, B) = P(A \mid B) * P(B)}$$

- The joint probability is symmetrical : $P(A, B) = P(B, A)$

- The conditional probability can be calculated using the joint probability:

$$\boxed{P(A \mid B) = P(A, B) / P(B)}$$

- The conditional probability is not symmetrical : $P(A \mid B) != P(B \mid A)$

Bayes' Theorem

- In statistics and probability theory, the Bayes' theorem (also known as the Bayes' rule) is a mathematical formula used to determine the conditional probability of events.
- Essentially, the Bayes' theorem describes the probability of an event based on prior knowledge of the conditions that might be relevant to the event.
- The Bayes' theorem is expressed in the following formula:

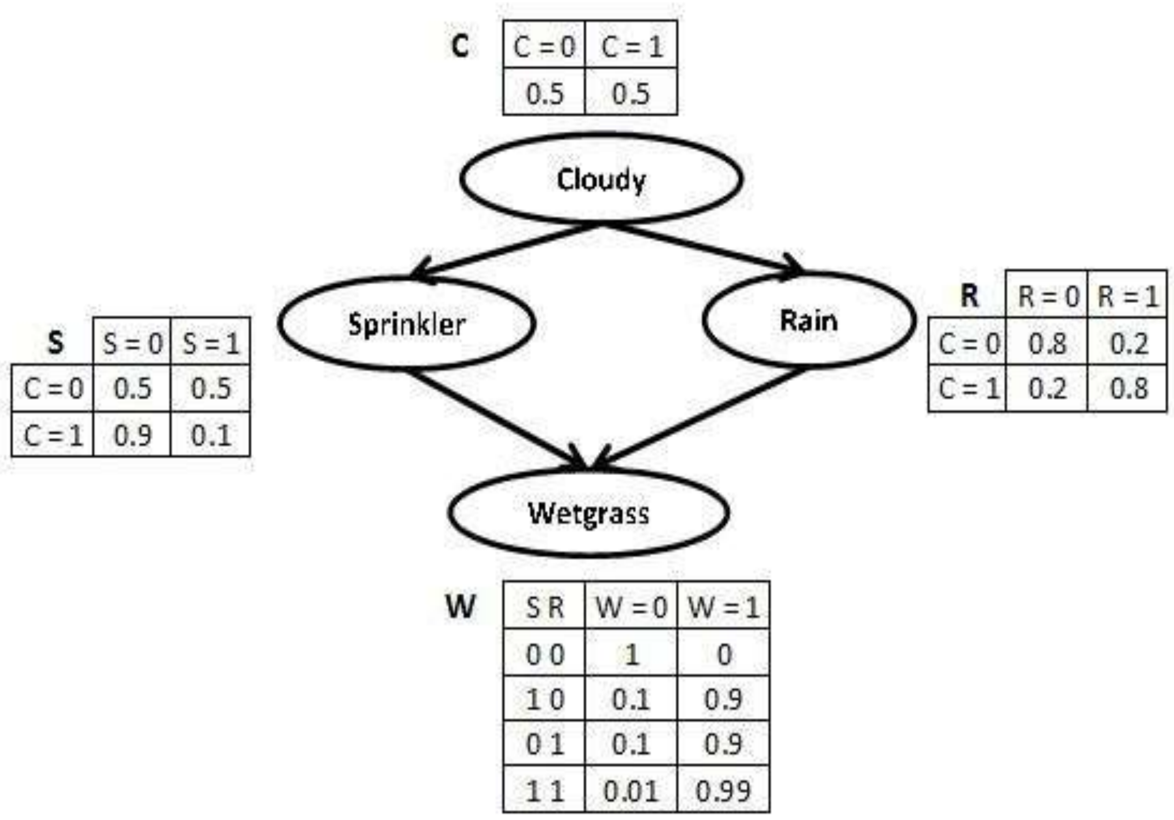
$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

  - Where:
  - $P(A|B)$ – the probability of event A occurring, given event B has occurred
  - $P(B|A)$ – the probability of event B occurring, given event A has occurred
  - $P(A)$ – the probability of event A
  - $P(B)$ – the probability of event B
  - Note that events A and B are independent events

Bayesian network

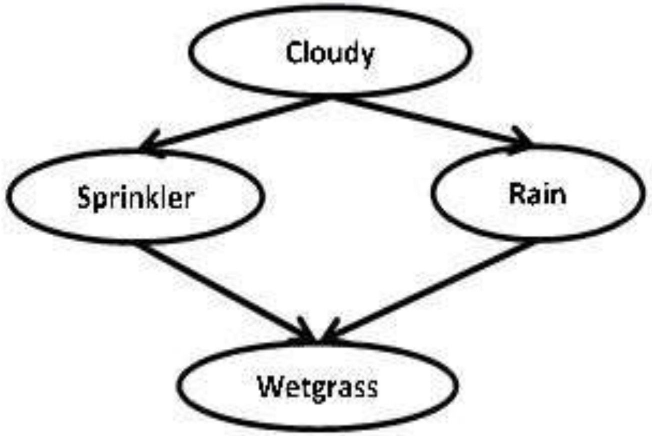
- A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph.
- In a directed acyclic graph, each edge corresponds to a conditional dependency, and each node corresponds to a unique random variable.
- It is also called a Bayes network, belief network, decision network, or Bayesian model.
- Bayesian Network represents the dependency among events and assigning probabilities to them.
- Thus ascertaining how probable or what is the change of occurrence of one event given the other.
- It can be used in various tasks including prediction, anomaly detection, diagnostics, automated insight, reasoning, time series prediction, and decision making under uncertainty.

Bayes Network - Example



- Whether the grass is wet, W, depends on whether the sprinkler has been used, S, or whether it has rained, R.
- Whether the sprinkler is used depends on whether it is cloudy, similarly for whether it has rained.
- The probability of the grass being wet is conditionally independent of it being cloudy, given information about the sprinklers and whether it has rained.
- This joint probability may be expressed as
  $P(C, S, R, W) = P(C)P(S|C)P(R|C,S)P(W|C,S, R)$

Bayes Network - Example



C = 0	C = 1
0.5	0.5

S	S = 0	S = 1
C = 0	0.5	0.5
C = 1	0.9	0.1

R	R = 0	R = 1
C = 0	0.8	0.2
C = 1	0.2	0.8

S R	W = 0	W = 1
0 0	1	0
1 0	0.1	0.9
0 1	0.1	0.9
1 1	0.01	0.99

- It is cloudy, what's the probability that the grass is wet?
- So, we want to compute P(W = T|C = T).
- By the chain rule of probability, the joint probability of all the nodes in the graph above is,

$$P(C,S,R,W) = P(C)*P(S|C)*P(R|C,S) * P(W|C,S,R)$$
$$P(C,S,R,W) = 0.99 \times 0.1 \times 0.8 + 0.90 \times 0.1 \times 0.2$$
$$+0.90 \times 0.9 \times 0.8 + 0.00 \times 0.9 \times 0.2$$
$$P(C,S,R,W) = 0.7452$$

Certainty Factor

- A Certainty Factor (CF) is a numerical estimates of the belief or disbelief on a conclusion in the presence of set of evidence. Different methods for adopting Certainty Factor have been adopted.
  1. Use a scale from 0 to 1, where 0 indicates certainly false (total disbelief), 1 indicates definitely true (total belief). Other values between 0 to 1 represents varying degrees of beliefs and disbeliefs.
  2. use a scale from −1 to +1 where -1 indicates certainly false, +1 indicates definitely true, and intermediate values represent varying degrees of certainty, with 0 meaning unknown.
- The weights express the perceived certainty of a fact being true.
- The use of certainty factors is similar to probabilistic reasoning but is less formally related to probability theory.
- There are many schemes for treating uncertainty in rule based systems. The most common are
  - Adding certainty factors.
  - Adoptions of Dempster-Shafer belief functions.
  - Inclusion of fuzzy logic.

Certainty Factor in a Rule based System

- In a rule based system, a rule is an expression of the form "if A then B" where A is an assertion and B can be either an action or another assertion.
- A problem with rule-based systems is that often the connections reflected by the rules are not absolutely certain or deterministic, and the gathered information is often subject to uncertainty.
- In such cases, a certainty measure is added to the premises as well as the conclusions in the rules of the system.
- A rule then provides a function that describes : how much a change in the certainty of the premise will change the certainty of the conclusion.
- In its simplest form, this looks like :

If A (with certainty x) then B (with certainty f(x))

Certainty Factor in a Rule based System

- Each rule has a certainty attached to it. Once the identities of the virus/bacteria are found, it then attempts to select a therapy by which the disease can be treated.
- A certainty factor (CF [h, e]) is defined in terms of two components:

1. MB[h, e] - a measure (between 0 and 1) of belief in hypothesis “h” given the evidence “e”.
   - MB measures the extent to which the evidence supports the hypothesis.
   - It is zero if the evidence fails to support the hypothesis.
2. MD[h, e] - a measure (between 0 and 1) of disbelief in hypothesis “h” given the evidence “e”.
   - MD measures the extent to which the evidence supports the negation of the hypothesis. It is zero if the evidence support the hypothesis.

$$CF[h, e] = MB[h, e] - MD[h, e]$$

Dempster – Shafer Theory

- In Dempster-Shafer Theory we consider sets of propositions and assign an interval to each of them in which the degree of belief must lie.

[Belief, Plausibility]

- Belief (denoted as Bel) measures the strength of the evidence in favor of a set of propositions.
- It ranges from 0 (no evidence) to 1 (definite certainty)
- Plausibility (Pl) is Pl(s) = 1- Bel(¬s)
- It also ranges from 0 to 1 and measures the extent to which evidence in favor of ¬s leaves room for belief in s.
- In short, if we have certain evidence in favor of (¬s), then Bel(not(s)) will be 1 and Pl(s) will be 0.
- This tells us that the only possible value for Bel(s) is also 0.
- The interval, also tells about the amount of information that we have.
- If we have no evidence we say that the hypothesis is in the range of [0, 1].

Dempster – Shafer Theory

- Let's take an example where we have some mutually exclusive hypothesis.
- Let the set {Allergy, Flu, Cold, Pneumonia} be denoted by **θ** and we want to attach some measure of belief to elements of **θ**.
- The key function we use here is a Probability Density Function, denoted by m.
- The function m, is not only defined for elements of **θ** but also all subsets of it.
- We must assign m so that the sum of all the m values assigned to subsets of **θ** is 1.
- At the beginning we have m as under **θ = (1.0)**
- If we get an evidence of 0.6 magnitude that the correct diagnosis is in the set {Flu, Cold, Pneu} then,

{Flu, Cold, Pneu} = (0.6)
θ = (0.4)

Dempster – Shafer Theory

- Now to move further, let's consider we have two belief function m1 and m2.
- Let X be the set of subsets of θ to which m1 assigns a nonzero value and let Y be the corresponding set for m2.
- We define m3, as a combination of the m1 and m2 to be,

$$\mathbf{m}_3(\mathbf{Z}) = \frac{\sum_{\mathbf{X} \cap \mathbf{Y} = \mathbf{Z}} \mathbf{m}_1(\mathbf{X}) \cdot \mathbf{m}_2(\mathbf{Y})}{1 - \sum_{\mathbf{X} \cap \mathbf{Y} = \emptyset} \mathbf{m}_1(\mathbf{X}) \cdot \mathbf{m}_2(\mathbf{Y})}$$

Dempster – Shafer Theory

- For example, suppose m1 corresponds to our belief after observing fever:

$$m1 = \{ F, C, P\} = 0.6 \text{ and } \theta = (0.4)$$

- suppose m2 corresponds to our belief after observing runny nose:

$$m2 = \{ A, F, C\} = 0.8 \text{ and } \theta = (0.2)$$

- Then we can compute their combination m3 using the following table.

	{A,F,C }	(0.8)	Θ	(0.2)
{F,C,P } (0.6)	{F,C }	(0.48)	{F,C,P}	(0.12)
Θ (0.4)	{A,F,C }	(0.32)	Θ	(0.08)

- So we produce a new, combined m3 as,

{Flu, Cold} (0.48)	{All, Flu, Cold} (0.32)	{Flu, Cold, Pneu} (0.12) Θ (0.08)

Fuzzy Logic

- Fuzzy logic is a set of mathematical principles for knowledge representation based on degrees of membership rather than on crisp membership of classical binary logic.
- Fuzzy logic is a form of many-valued logic in which the truth values of variables may be any real number between 0 and 1.
- By contrast, in Boolean logic, the truth values of variables may only be the integer values 0 or 1.
- Fuzzy logic has been employed to handle the concept of partial truth, where the truth value may range between completely true and completely false.
- Furthermore, when linguistic variables are used, these degrees may be managed by specific membership functions.
- Such methods have been used in control systems for devices like trains, AC, and washing machines.
- The concepts of Fuzzy Logic are extensively applied in business, finance, aerospace, defense, etc.

Fuzzy Sets and Membership function

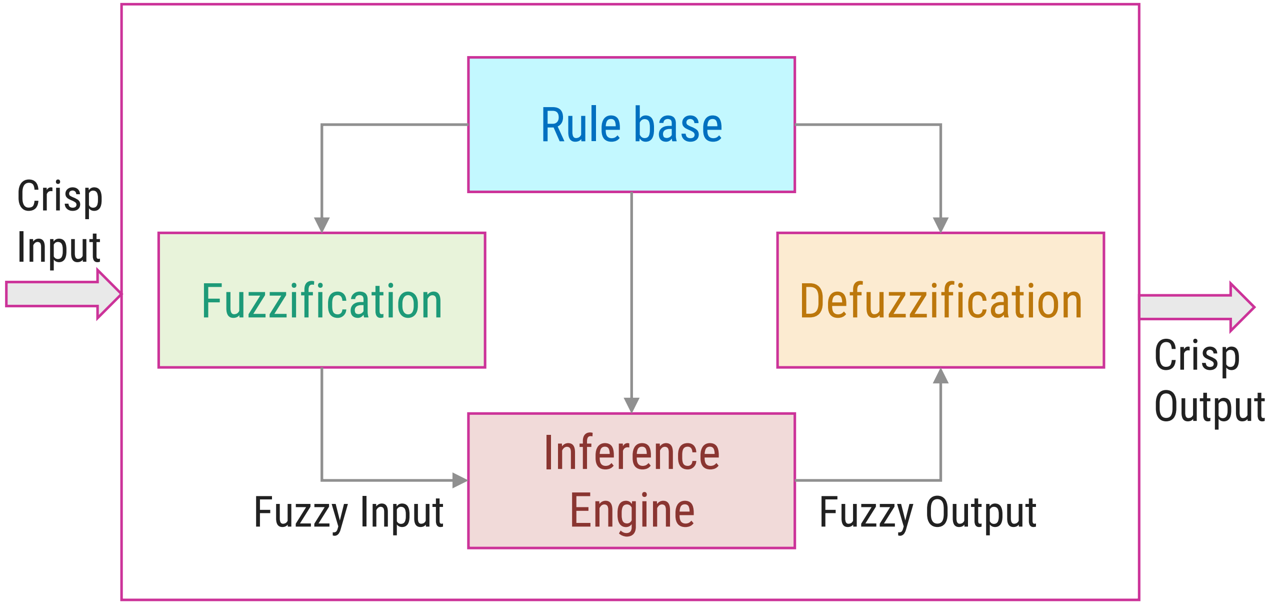
- The concept of a set is fundamental to mathematics. Crisp set theory is governed by a logic that uses one of only two values: true or false.
- This logic cannot represent vague concepts, and therefore fails to give the answers on the inconsistencies.
- In fuzzy set theory, an element is with a certain degree of membership. Thus, a proposition is not either true or false, but may be partly true (or partly false) to any degree.
- This degree is usually taken as a real number in the interval [0,1].

Fuzzy Sets and Membership function

- 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.

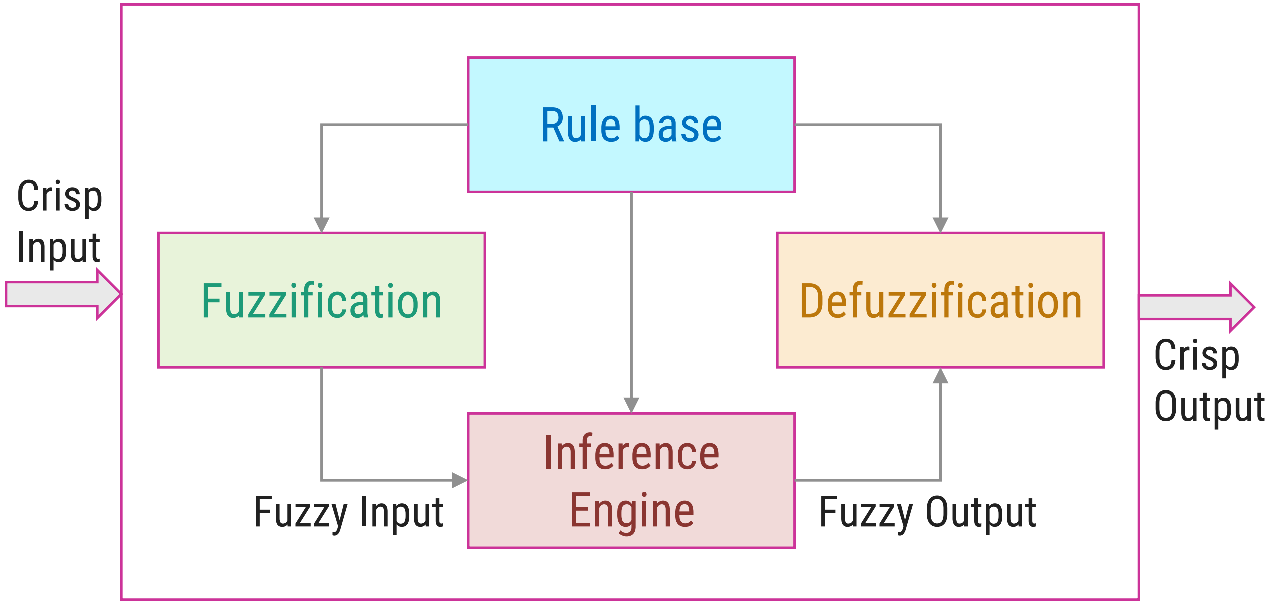
Name	Height in cm	Degree of Membership	
		Crisp	Fuzzy
John	208	1	1
Tom	181	1	0.8
Bob	152	0	0.0
Mike	198	1	0.9
Billy	158	0	0.4

Architecture of a Fuzzy Logic System



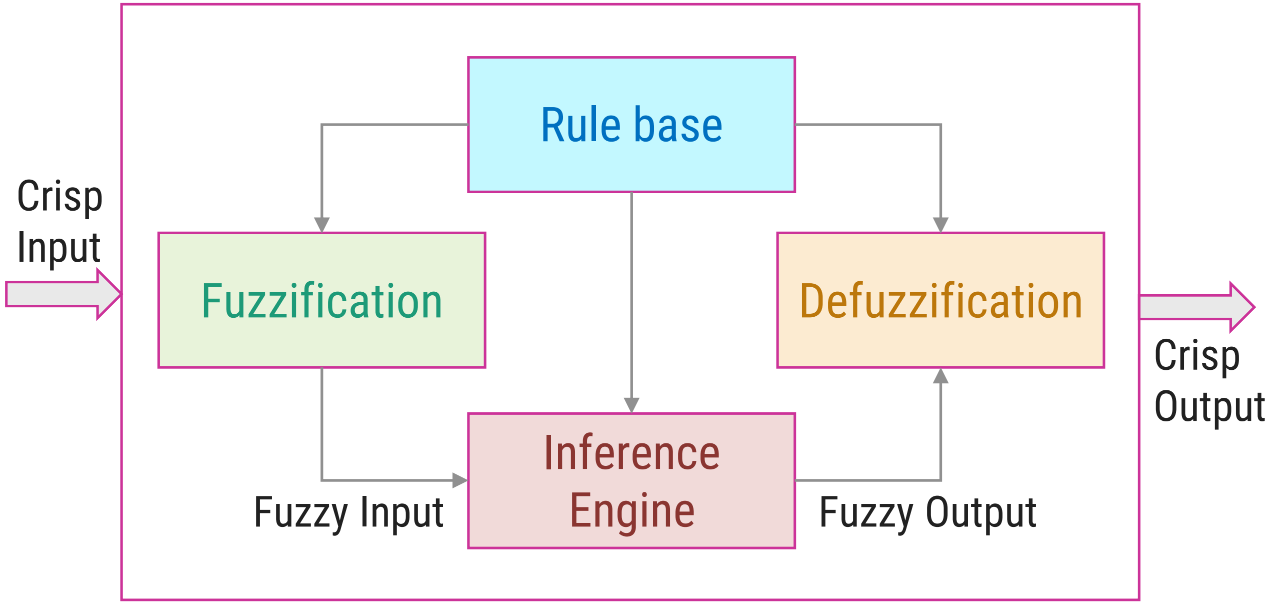
- In the architecture of the Fuzzy Logic system, each component plays an important role. The architecture consists of four different components.
  1. Rule Base
  2. Fuzzification
  3. Inference Engine
  4. Defuzzification

Architecture of a Fuzzy Logic System



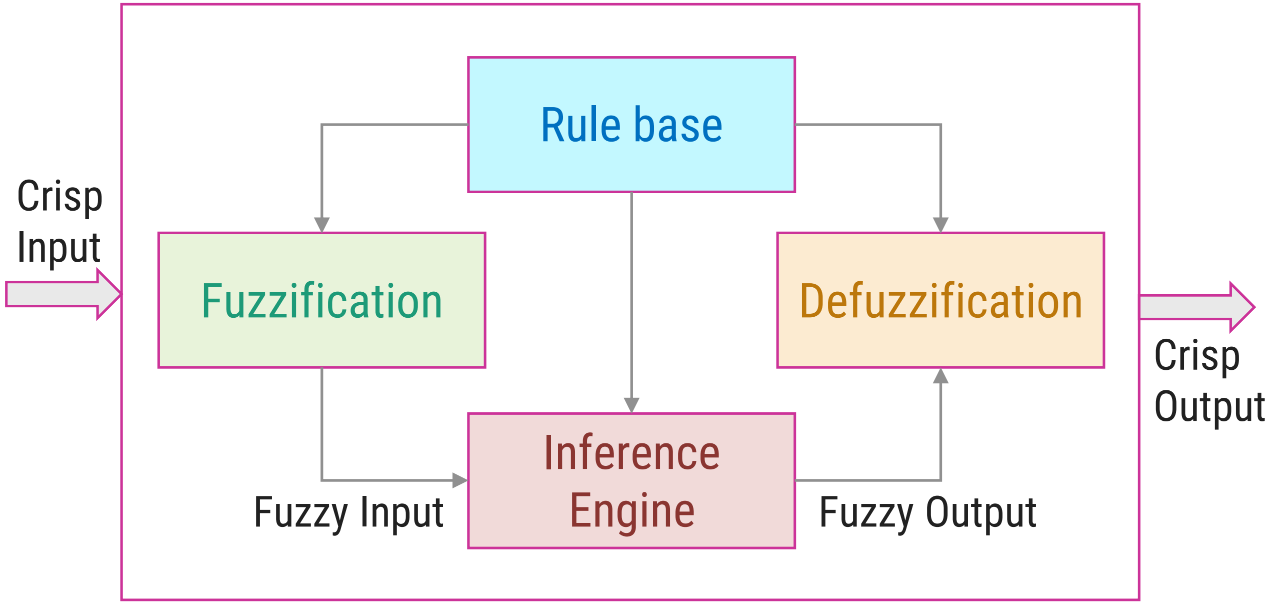
1. Rule Base :
   - Rule Base is a component used for storing the set of rules and the If-Then conditions given by the experts are used for controlling the decision-making systems.
   - There are so many functions which offer effective methods for designing and tuning of fuzzy controllers.
   - These updates or developments decreases the number of fuzzy set of rules.

Architecture of a Fuzzy Logic System



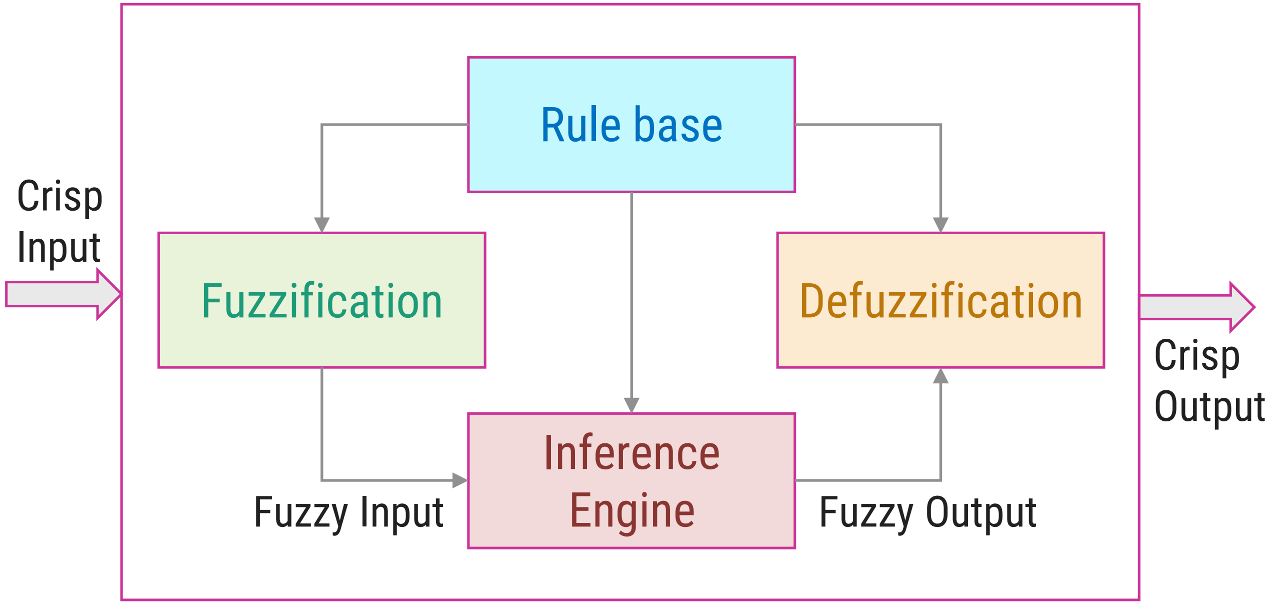
2. Fuzzification :
   - Fuzzification is a module or component for transforming the system inputs, i.e., it converts the crisp number into fuzzy steps.
   - The crisp numbers are those inputs which are measured by the sensors and then fuzzification passed them into the control systems for further processing.
   - This component divides the input signals into following five states in any Fuzzy Logic system:
     i. Large Positive (LP)
     ii. Medium Positive (MP)
     iii. Small (S)
     iv. Medium Negative (MN)
     v. Large negative (LN)

Architecture of a Fuzzy Logic System



3. Inference Engine :
   - This component is a main component in any Fuzzy Logic system (FLS), because all the information is processed in the Inference Engine.
   - It allows users to find the matching degree between the current fuzzy input and the rules.
   - After the matching degree, this system determines which rule is to be added according to the given input field. When all rules are fired, then they are combined for developing the control actions.

Architecture of a Fuzzy Logic System



4. Defuzzification
   - Defuzzification is a module or component, which takes the fuzzy set inputs generated by the Inference Engine, and then transforms them into a crisp value.
   - It is the last step in the process of a fuzzy logic system.
   - The crisp value is a type of value which is acceptable by the user.
   - Various techniques are present to do this, but the user has to select the best one for reducing the errors.

Thank You!