Artificial Intelligence

Unit-3

Syllabus

- **UNIT – III**
- Advanced Knowledge Representation and Reasoning: Knowledge Representation Issues, Non-monotonic Reasoning, Other Knowledge Representation Schemes Reasoning Under Uncertainty: Basic probability, Acting Under Uncertainty, Bayes' Rule, Representing Knowledge in an Uncertain Domain, Bayesian Networks

Knowledge Representation Issues

- The fundamental goal of Knowledge Representation is to facilitate inferencing (conclusions) from knowledge. The issues that arise while using KR techniques are many. Some of these are explained below.

- **Important Attributes** : Any attribute of objects so basic that they occur in almost every problem domain ?

- **Relationship among attributes**: Any important relationship that exists among object attributes ?

- **Choosing Granularity** : At what level of detail should the knowledge be represented ?

- **Set of objects** : How set of objects be represented ?

- **Finding Right structure** : Given a large amount of knowledge stored, how can relevant parts be accessed ?

Reasoning

- The reasoning is the mental process of deriving logical conclusion and making predictions from available knowledge, facts, and beliefs.

or

- we can say, "**Reasoning is a way to infer facts from existing data**." It is a general process of thinking rationally, to find valid conclusions.

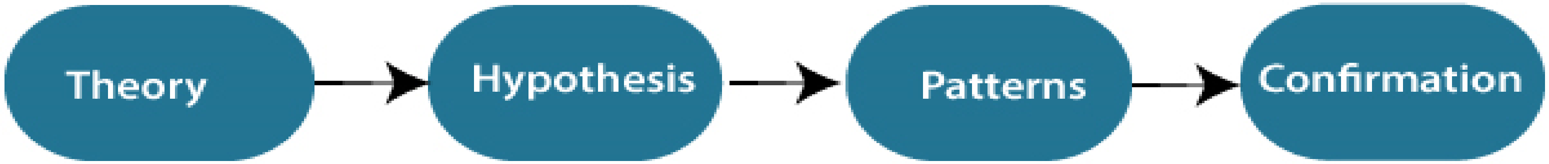
- In artificial intelligence, the reasoning is essential so that the machine can also think rationally as a human brain, and can perform like a human.

- **Types of Reasoning**

In artificial intelligence, reasoning can be divided into the following categories:

- Deductive reasoning
- Inductive reasoning
- Abductive reasoning
- Common Sense Reasoning
- Monotonic Reasoning
- Non-monotonic Reasoning

- **1. Deductive reasoning:**
- It is deriving new information from logically related known information. It is the form of valid reasoning, which means the argument's conclusion must be true when the premises are true.
- Deductive reasoning is a type of propositional logic in AI, and it requires various rules and facts. It is sometimes referred to as top-down reasoning, and contradictory to inductive reasoning.
- In deductive reasoning, the truth of the premises guarantees the truth of the conclusion.
- Deductive reasoning mostly starts from the general premises to the specific conclusion, which can be explained as below example.
- **Example:**
- **Premise-1: All the human eats veggies**
- **Premise-2: Suresh is human.**
- **Conclusion: Suresh eats veggies.**
- The general process of deductive reasoning is given below:



Deductive Reasoning

- **<u>Deductive Reasoning</u>** – A type of logic in which one goes from a general statement to a specific instance.
- The classic example

  All men are mortal. (major premise)

  <u>Socrates is a man.</u> (minor premise)

  Therefore, Socrates is mortal. (conclusion)

The above is an example of a ***syllogism***.

2. Inductive Reasoning:

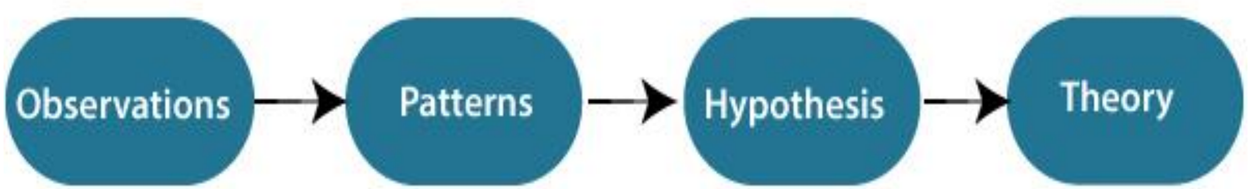
Inductive reasoning is a form of reasoning to arrive at a conclusion using limited sets of facts by the process of generalization. It starts with the series of specific facts or data and reaches to a general statement or conclusion.

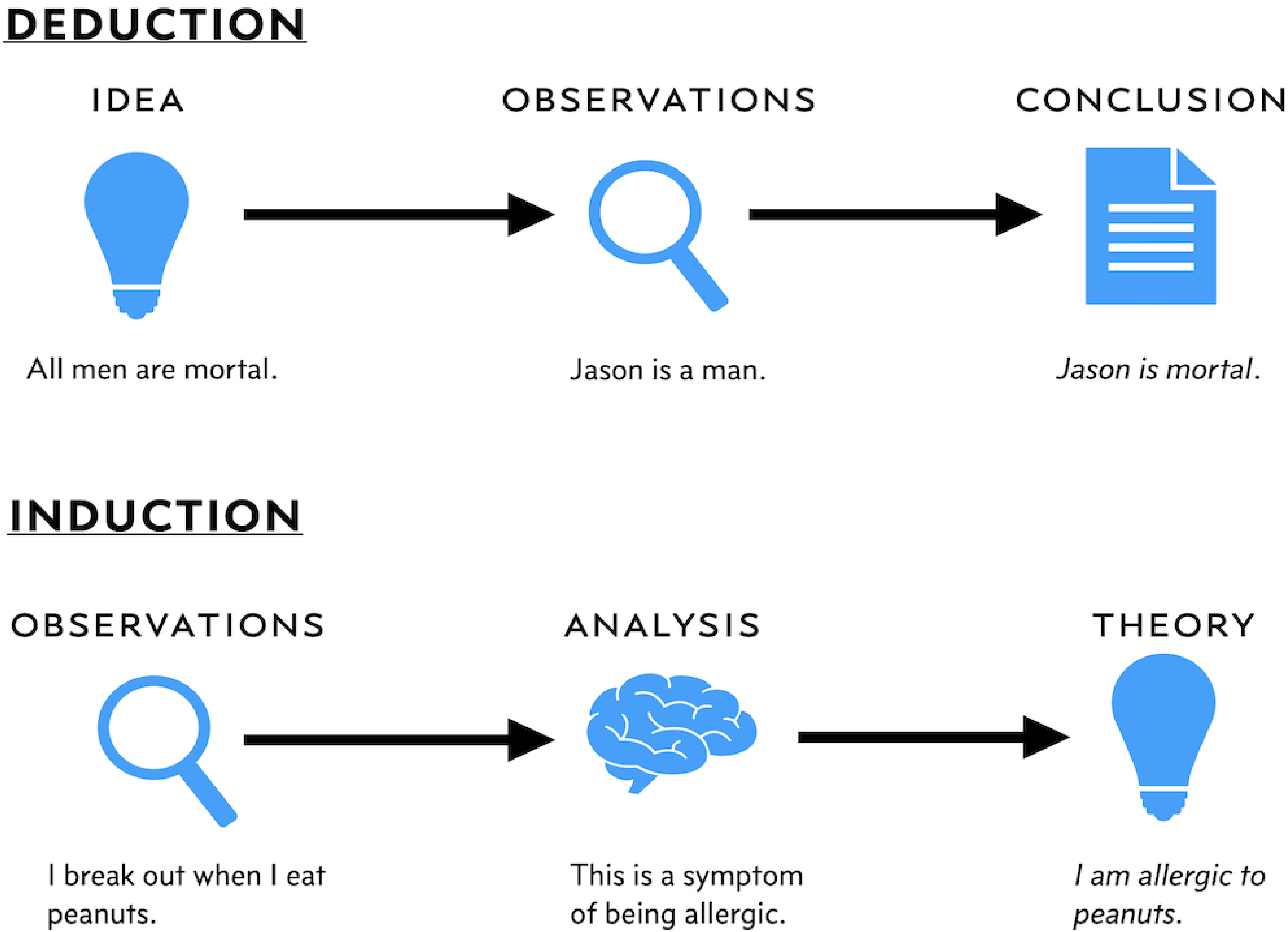
Inductive reasoning is a type of propositional logic, which is also known as cause-effect reasoning or bottom-up reasoning.

In inductive reasoning, we use historical data or various premises to generate a generic rule, for which premises support the conclusion.

In inductive reasoning, premises provide probable supports to the conclusion, so the truth of premises does not guarantee the truth of the conclusion.

Example:
Premise: All of the pigeons we have seen in the zoo are white.
Conclusion: Therefore, we can expect all the pigeons to be white.




DEDUCTION
IDEA
OBSERVATIONS
CONCLUSION
All men are mortal.
Jason is a man.
Jason is mortal.
INDUCTION
OBSERVATIONS
ANALYSIS
THEORY
I break out when I eat peanuts.
This is a symptom of being allergic.
I am allergic to peanuts.

- **3. Abductive reasoning:**
- Abductive reasoning is a form of logical reasoning which starts with single or multiple observations then seeks to find the most likely explanation or conclusion for the observation.
- Abductive reasoning is an extension of deductive reasoning, but in abductive reasoning, the premises do not guarantee the conclusion.
- **Example:**
- **Implication:** Cricket ground is wet if it is raining
- **Axiom:** Cricket ground is wet.
- Conclusion It is raining.

- **4. Common Sense Reasoning**
- Common sense reasoning is an informal form of reasoning, which can be gained through experiences.
- Common Sense reasoning simulates the human ability to make presumptions about events which occurs on every day.
- It relies on good judgment rather than exact logic and operates on **heuristic knowledge** and **heuristic rules**.
- **Example:**
- **One person can be at one place at a time.**
- **If I put my hand in a fire, then it will burn.**
- The above two statements are the examples of common sense reasoning which a human mind can easily understand and assume.

5. Monotonic Reasoning: In this, once the conclusion is taken, then it will remain the same even if we add some other information to existing information in our knowledge base. In monotonic reasoning, adding knowledge does not decrease the set of prepositions that can be derived.

•To solve monotonic problems, we can derive the valid conclusion from the available facts only, and it will not be affected by new facts.

•Monotonic reasoning is not useful for the real-time systems, as in real time, facts get changed, so we cannot use monotonic reasoning.

•Any theorem proving is an example of monotonic reasoning.

Example: Earth revolves around the Sun.

It is a true fact, and it cannot be changed even if we add another sentence in knowledge base like, "The moon revolves around the earth" Or "Earth is not round," etc.

Advantages of Monotonic Reasoning:

•In monotonic reasoning, each old proof will always remain valid.

•If we deduce some facts from available facts, then it will remain valid for always.

Disadvantages of Monotonic Reasoning:

•We cannot represent the real world scenarios using Monotonic reasoning.

•Since we can only derive conclusions from the old proofs, so new knowledge from the real world cannot be added.

6. Non-monotonic Reasoning When new information is added to the system and if the truthfulness of a conclusion remains same , then the system is referred to as non-monotonic.

•Non-monotonic reasoning deals with incomplete and uncertain models.

•"Human perceptions for various things in daily life, "is a general example of non-monotonic reasoning.

Example: Let suppose the knowledge base contains the following knowledge:

•**Birds can fly**
•**Pitty is a bird**
•**Penguins cannot fly**

So from the above sentences, we can conclude that **Pitty can fly**.

However, if we add one another sentence into knowledge base "**Pitty is a penguin**", which concludes "**Pitty cannot fly**", so it invalidates the above conclusion.

Advantages of Non-monotonic reasoning:

- For real-world systems such as Robot navigation, we can use non-monotonic reasoning.

Disadvantages of Non-monotonic Reasoning:

- In non-monotonic reasoning, the old facts may be invalidated by adding new sentences.

- It cannot be used for theorem **proving**.

Other Knowledge Representation Schemes

- **According to Mylopoulos and Levesque (1984) they have been classified into four categories:**

1. Logical Representation Scheme:

2. Procedural Representation Scheme:

3. Network Representation Scheme:

4. Structured Representation Scheme:

- **1. Logical Representation Scheme:**

- This class of representation **uses expressions** in formal logic to represent a knowledge base.

- Inference rules and proof procedures apply this knowledge to problem solving.

- First order predicate calculus is the most widely used logical representation scheme, and PROLOG is an ideal programming language for implementing logical representation schemes.

- **2. Procedural Representation Scheme:**

- Procedural scheme represents knowledge as **a set of instructions** for solving a problem.

- In a rule-based system, for example, an if then rule may be interpreted as a procedure for searching a goal in a problem domain: to arrive at the conclusion, solve the premises in order.

- Production systems are examples of a procedural representation scheme.

3. Network Representation Scheme:

- Network representation captures knowledge as a graph in which the nodes represent objects or concepts in the problem domain and the arcs represent relations or associations between them.

- Examples of network representations include semantic network, conceptual dependencies and conceptual graphs.

4. Structured Representation Scheme:

- Structured representation languages extend networks by allowing each node to be a complex data structure consisting of named slots with attached values.

- These values may be simple numeric or complex data, such as pointers to other frames, or even procedures.

Reasoning Under Uncertainty

- Till now, we have learned knowledge representation using first-order logic and propositional logic with certainty, which means we were sure about the predicates.

- With this knowledge representation, we might write A→B, which means if A is true then B is true, but consider a situation where we are not sure about whether A is true or not then we cannot express this statement, this situation is called uncertainty.

- So to represent uncertain knowledge, where we are not sure about the predicates, we need uncertain reasoning or probabilistic reasoning.

Causes of uncertainty:

- Information occurred from unreliable sources.
- Experimental Errors
- Equipment fault
- Temperature variation
- Climate change.

Basic probability

- Probabilistic reasoning is a way of knowledge representation where we apply the concept of probability to indicate the uncertainty in knowledge. In probabilistic reasoning, we combine probability theory with logic to handle the uncertainty.

- We use probability in probabilistic reasoning because it provides a way to handle the uncertainty that is the result of someone's laziness and ignorance.

- In the real world, there are lots of scenarios, where the certainty of something is not confirmed, such as "It will rain today," "behavior of someone for some situations," "A match between two teams or two players."

- These are probable sentences for which we can assume that it will happen but not sure about it, so here we use probabilistic reasoning.

- **Need of probabilistic reasoning in AI:**

- When there are unpredictable outcomes.
- When specifications or possibilities of predicates becomes too large to handle.
- When an unknown error occurs during an experiment.
- In probabilistic reasoning, there are two ways to solve problems with uncertain knowledge:
- **Bayes' rule**
- **Bayesian Statistics**

- As probabilistic reasoning uses probability and related terms, so before understanding probabilistic reasoning, let's understand some common terms:

- **Probability:** Probability can be defined as a chance that an uncertain event will occur. It is the numerical measure of the likelihood that an event will occur. The value of probability always remains between 0 and 1 that represent ideal uncertainties.

- $0 \leq P(A) \leq 1$, where P(A) is the probability of an event A.
- $P(A) = 0$, indicates total uncertainty in an event A.
- $P(A) = 1$, indicates total certainty in an event A.
-
- We can find the probability of an uncertain event by using the below formula.
- $P(\neg A)$ = probability of a not happening event.
- **$P(\neg A) + P(A) = 1$.**

- **Imp Terms**

- **Event:** Each possible outcome of a variable is called an event.

- **Sample space:** The collection of all possible events is called sample space.

- **Random variables:** Random variables are used to represent the events and objects in the real world.

- **Prior probability:** The prior probability of an event is probability computed before observing new information.

- **Posterior Probability:** The probability that is calculated after all evidence or information has taken into account. It is a combination of prior probability and new information.

Bayes Theorem

- Bayes' theorem is also known as **Bayes' rule, Bayes' law**, or **Bayesian reasoning**, which determines the probability of an event with uncertain knowledge.

- In probability theory, it relates the conditional probability and marginal probabilities of two random events.

- Bayes' theorem was named after the British mathematician **Thomas Bayes**. The **Bayesian inference** is an application of Bayes' theorem, which is fundamental to Bayesian statistics.

Bayes Theorem

- It is a way to calculate the value of P(B|A) with the knowledge of P(A|B).
- Bayes' theorem allows updating the probability prediction of an event by observing new information of the real world.
- **Example**: If cancer corresponds to one's age then by using Bayes' theorem, we can determine the probability of cancer more accurately with the help of age.

- Bayes' theorem can be derived using product rule and conditional probability of event A with known event B:

- As from product rule we can write:
- $P(A \land B) = P(A|B)\, P(B)$ or
- Similarly, the probability of event B with known event A:

- $P(A \land B) = P(B|A)\, P(A)$

- Equating right hand side of both the equations, we will get:
- The above equation (a) is called as **Bayes' rule** or **Bayes' theorem**. This equation is basic of most modern AI systems for **probabilistic inference**.

Example-1:

Question: what is the probability that a patient has diseases meningitis with a stiff neck?

Given Data:

A doctor is aware that disease meningitis causes a patient to have a stiff neck, and it occurs 80% of the time. He is also aware of some more facts, which are given as follows:

- The Known probability that a patient has meningitis disease is 1/30,000.
- The Known probability that a patient has a stiff neck is 2%.

Let a be the proposition that patient has stiff neck and b be the proposition that patient has meningitis. , so we can calculate the following as:

P(a|b) = 0.8

P(b) = 1/30000

P(a)= .02

$$\mathbf{P(b|a)} = \frac{P(a|b)P(b)}{P(a)} = \frac{0.8*(\frac{1}{30000})}{0.02} = 0.001333333.$$

Hence, we can assume that 1 patient out of 750 patients has meningitis disease with a stiff neck.

Example-2:

Question: From a standard deck of playing cards, a single card is drawn. The probability that the card is king is 4/52, then calculate posterior probability P(King|Face), which means the drawn face card is a king card.

Solution:

$$\mathbf{P(king|face)} = \frac{P(Face|king) * P(King)}{P(Face)} \quad \ldots\ldots(i)$$

P(king): probability that the card is King= 4/52= 1/13

P(face): probability that a card is a face card= 3/13

P(Face|King): probability of face card when we assume it is a king = 1

Putting all values in equation (i) we will get:

$$P(king|face) = \frac{1 * (\frac{1}{13})}{(\frac{3}{13})}$$ **= 1/3, it is a probability that a face card is a king card.**

Application of Bayes' theorem in Artificial intelligence:

- It is used to calculate the next step of the robot when the already executed step is given.
-
- Bayes' theorem is helpful in weather forecasting.

Representing Knowledge in an Uncertain Domain

- In real life, it is not always possible to determine the state of the environment as it might not be clear.

- Due to partially observable or non-deterministic environments, agents may need to handle uncertainty and deal with

- **Uncertain data**: Data that is missing, unreliable, inconsistent or noisy

- **Uncertain knowledge:** When the available knowledge has multiple causes leading to multiple effects or incomplete knowledge

- **Uncertain knowledge representation:** The representations which provides a restricted model of the real system, or has limited expressiveness

- **Inference:** In case of incomplete or default reasoning methods, conclusions drawn might not be completely accurate.

Such uncertain situations can be dealt with using

- Probability theory
- Truth Maintenance systems
- Fuzzy logic.

Probability

- Probability is the degree of likeliness that an event will occur.
- It provides a certain degree of belief in case of uncertain situations.
- It is defined over a set of events U and assigns value P(e) i.e. probability of occurrence of event e in the range [0,1].
- Here each sentence is labeled with a real number in the range of 0 to 1, 0 means the sentence is false and 1 means it is true.
- Conditional Probability or Posterior Probability is the probability of event A given that B has already occurred.
- $P(A|B) = (P(B|A) * P(A)) / P(B)$

- For example, P(It will rain tomorrow| It is raining today) represents conditional probability of it raining tomorrow as it is raining today.
- $P(A|B) + P(NOT(A)|B) = 1$

Joint Probability

- Joint probability is the probability of 2 independent events happening simultaneously like rolling two dice or tossing two coins together.

- Joint probability has a wide use in various fields such as physics, astronomy, and comes into play when there are two independent events.

Fuzzy logic:

- Fuzzy Logic (FL) is a method of reasoning that resembles human reasoning.

- The approach of FL imitates the way of decision making in humans that involves all intermediate possibilities between digital values YES and NO.

- The conventional logic block that a computer can understand takes precise input and produces a definite output as TRUE or FALSE, which is equivalent to human's YES or NO.

- The inventor of fuzzy logic, Lotfi Zadeh, observed that unlike computers, the human decision making includes a range of possibilities between YES and NO, such as

CERTAINLY YES
POSSIBLY YES
CANNOT SAY
POSSIBLY NO
CERTAINLY NO

The fuzzy logic works on the levels of possibilities of input to achieve the definite output.

Fuzzy Logic Systems Architecture

- **Fuzzification Module** − It transforms the system inputs, which are crisp numbers, into fuzzy sets. It splits the input signal into five steps such as −

LP	x is Large Positive
MP	x is Medium Positive
S	x is Small
MN	x is Medium Negative
LN	x is Large Negative

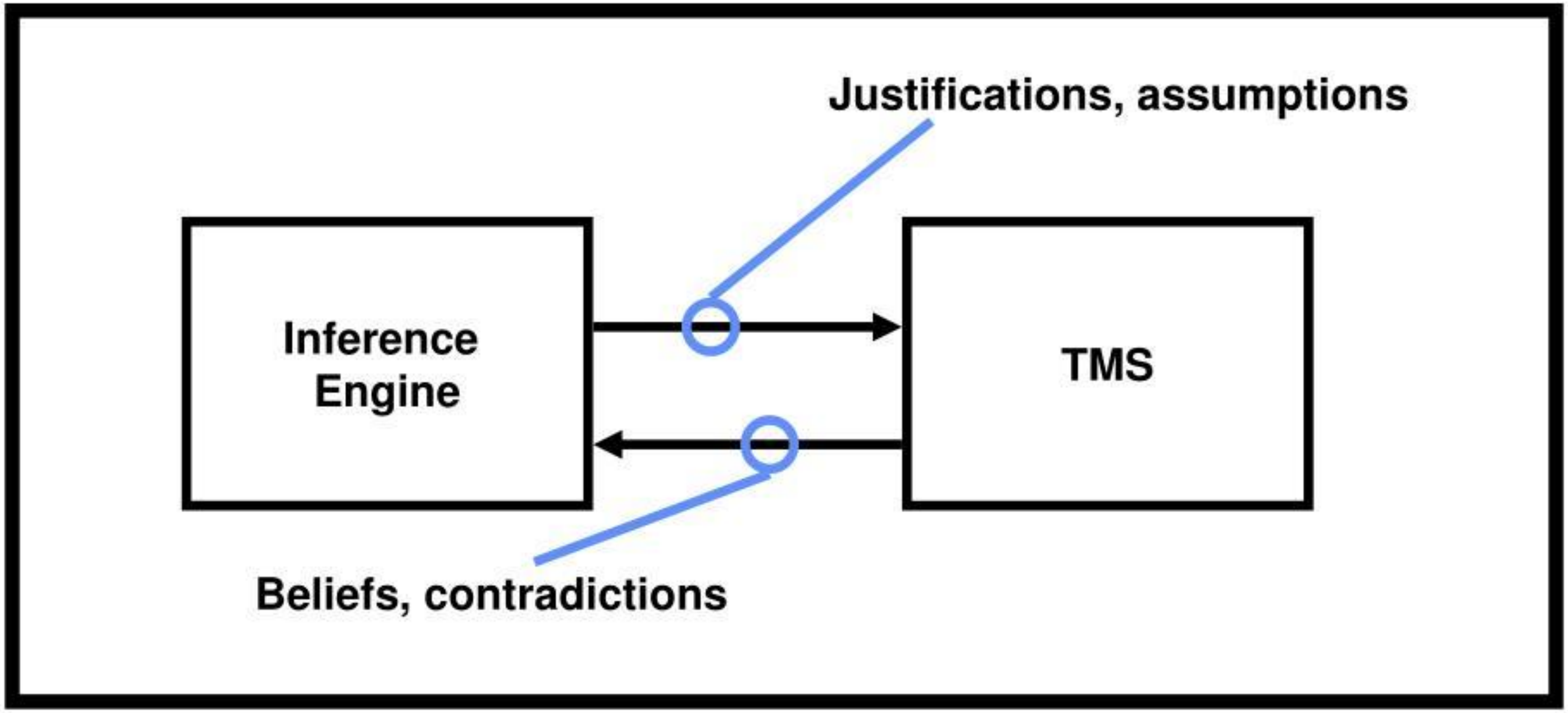
What is TMS?

A Truth Maintenance System (TMS) is a Problem Solver module responsible for:

- Enforcing logical relations among beliefs.
- Generating explanations for conclusions.
- Finding solutions to search problems
- Supporting default reasoning.
- Identifying causes for failure and recover from inconsistencies.

What is a TMS?

- A useful problem-solver module



Truth Maintenance System:

- To choose their actions, reasoning programs must be able to make assumptions and subsequently revise their beliefs when discoveries contradict these assumptions.

- The Truth Maintenance System (TMS) is a problem solver subsystem for performing these functions by recording and maintaining the reasons for program beliefs.

- Such recorded reasons are useful in constructing explanations of program actions and in guiding the course of action of a problem solver.

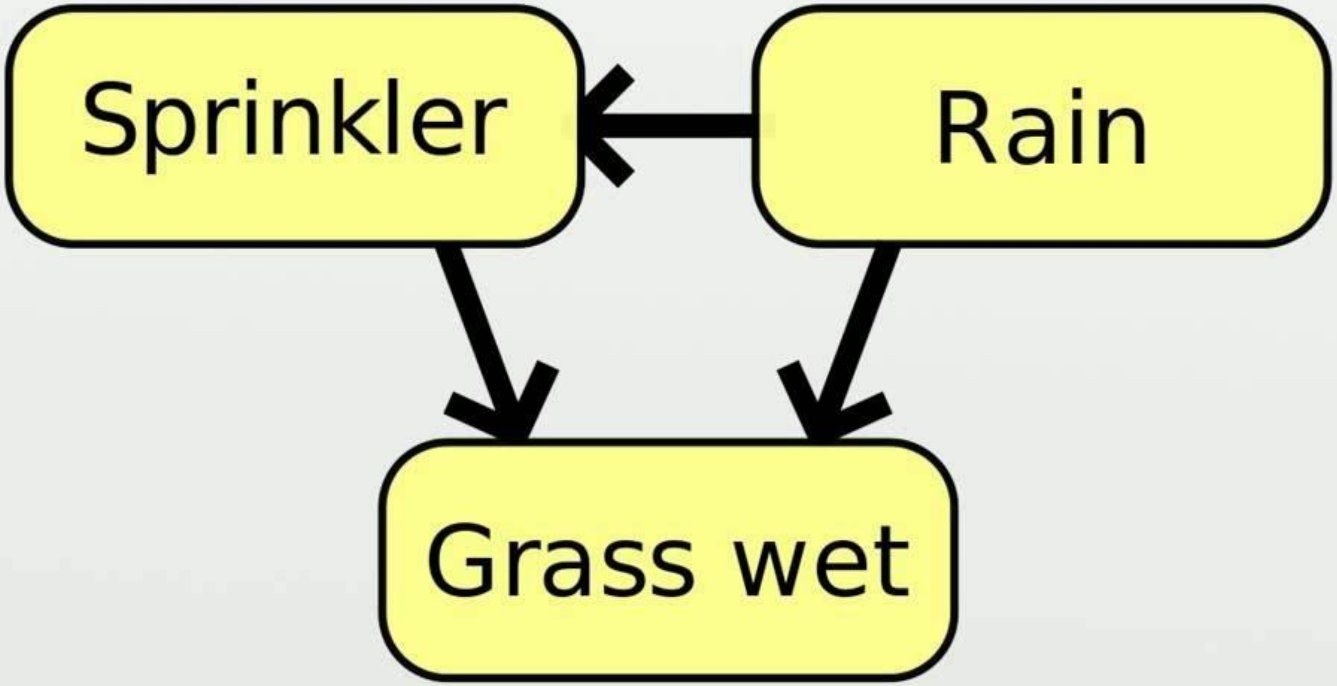
Truth Maintenance System:

- TMS are another form of knowledge representation which is best visualized in terms of graphs.
- It stores the latest truth value of any predicate. The system is developed with the idea that truthfulness of a predicate can change with time, as new knowledge is added or exiting knowledge is updated.
- It keeps a record showing which items of knowledge is currently believed or disbelieved.

Bayesian Networks

- "A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph.“
- It is also called a **Bayes network, belief network, decision network**, or **Bayesian model**.
- Bayesian networks are probabilistic, because these networks are built from a **probability distribution**, and also use probability theory for prediction and anomaly detection.

Bayesian network



Bayesian Networks

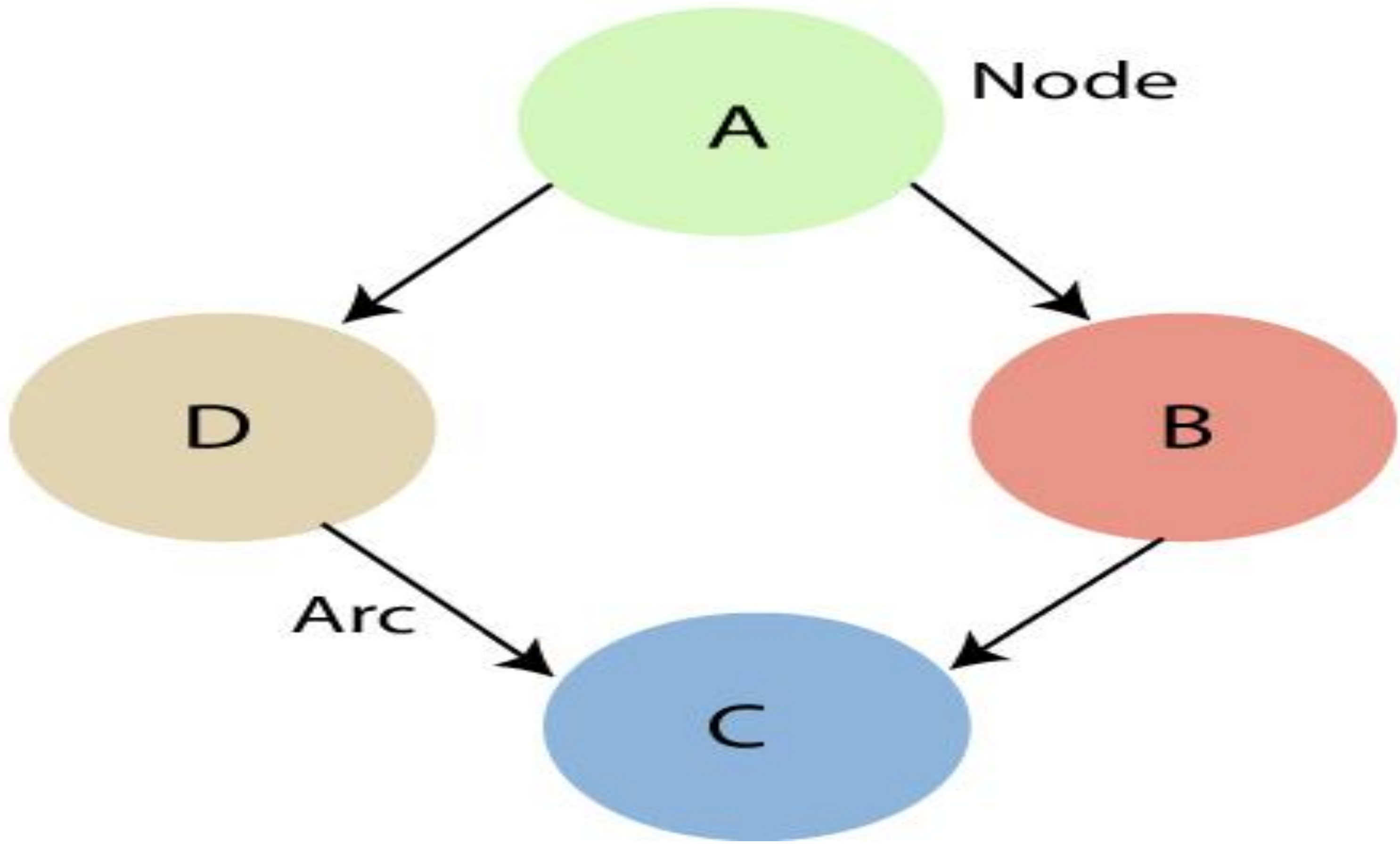
- Real world applications are probabilistic in nature, and to represent the relationship between multiple events, we need a Bayesian network.

- It can also be used in various tasks including **prediction, anomaly detection, diagnostics, automated insight, reasoning, time series prediction**, and **decision making under uncertainty**.

- Bayesian Network can be used for building models from data and experts opinions, and it consists of two parts:
- **Directed Acyclic Graph**
- **Table of conditional probabilities.**

- The generalized form of Bayesian network that represents and solve decision problems under uncertain knowledge is known as an **Influence diagram**.

- **A Bayesian network graph is made up of nodes and Arcs (directed links)**


Node
A
D
B
Arc
C

- Each **node** corresponds to the random variables, and a variable can be **continuous** or **discrete**.

- **Arc or directed arrows** represent the causal relationship or conditional probabilities between random variables. These directed links or arrows connect the pair of nodes in the graph. These links represent that one node directly influence the other node, and if there is no directed link that means that nodes are independent with each other .

  – **In the given diagram, A, B, C, and D are random variables represented by the nodes of the network graph.**
  – **If we are considering node B, which is connected with node A by a directed arrow, then node A is called the parent of Node B.**
  – **Node C is independent of node A.**

- The Bayesian network has mainly two components:
- **Causal Component**
- **Actual numbers**

- Each node in the Bayesian network has condition probability distribution $\mathbf{P(X_i|Parent(X_i))}$, which determines the effect of the parent on that node.

- Bayesian network is based on Joint probability distribution and conditional probability. So let's first understand the joint probability distribution:

- **Joint probability distribution:**
- If we have variables x1, x2, x3,....., xn, then the probabilities of a different combination of x1, x2, x3.. xn, are known as Joint probability distribution.

- $\mathbf{P[x_1, x_2, x_3,....., x_n]}$, it can be written as the following way in terms of the joint probability distribution.
- $\mathbf{= P[x_1 | x_2, x_3,....., x_n]P[x_2, x_3,....., x_n]}$
- $\mathbf{= P[x_1 | x_2, x_3,....., x_n]P[x_2|x_3,....., x_n]....P[x_{n-1}|x_n]P[x_n].}$
- In general for each variable Xi, we can write the equation as:
- $P(X_i|X_{i-1},........., X_1) = P(X_i |Parents(X_i))$

- **Example :**
- Harry installed a new burglar alarm at his home to detect burglary. The alarm reliably responds at detecting a burglary but also responds for minor earthquakes.

- Harry has two neighbors David and Sophia, who have taken a responsibility to inform Harry at work when they hear the alarm.
- David always calls Harry when he hears the alarm, but sometimes he got confused with the phone ringing and calls at that time too.
- On the other hand, Sophia likes to listen to high music, so sometimes she misses to hear the alarm. Here we would like to compute the probability of Burglary Alarm.

- **Problem:**
- **Calculate the probability that alarm has sounded, but there is neither a burglary, nor an earthquake occurred, and David and Sophia both called the Harry.**

- **Solution:**
- The Bayesian network for the above problem is given below. The network structure is showing that burglary and earthquake is the parent node of the alarm and directly affecting the probability of alarm's going off, but David and Sophia's calls depend on alarm probability.

- The network is representing that our assumptions do not directly perceive the burglary and also do not notice the minor earthquake, and they also not confer before calling.
- The conditional distributions for each node are given as conditional probabilities table or CPT.
- Each row in the CPT must be sum to 1 because all the entries in the table represent an exhaustive set of cases for the variable.
- In CPT, a boolean variable with k boolean parents contains 2^K probabilities. Hence, if there are two parents, then CPT will contain 4 probability values

- **List of all events occurring in this network:**

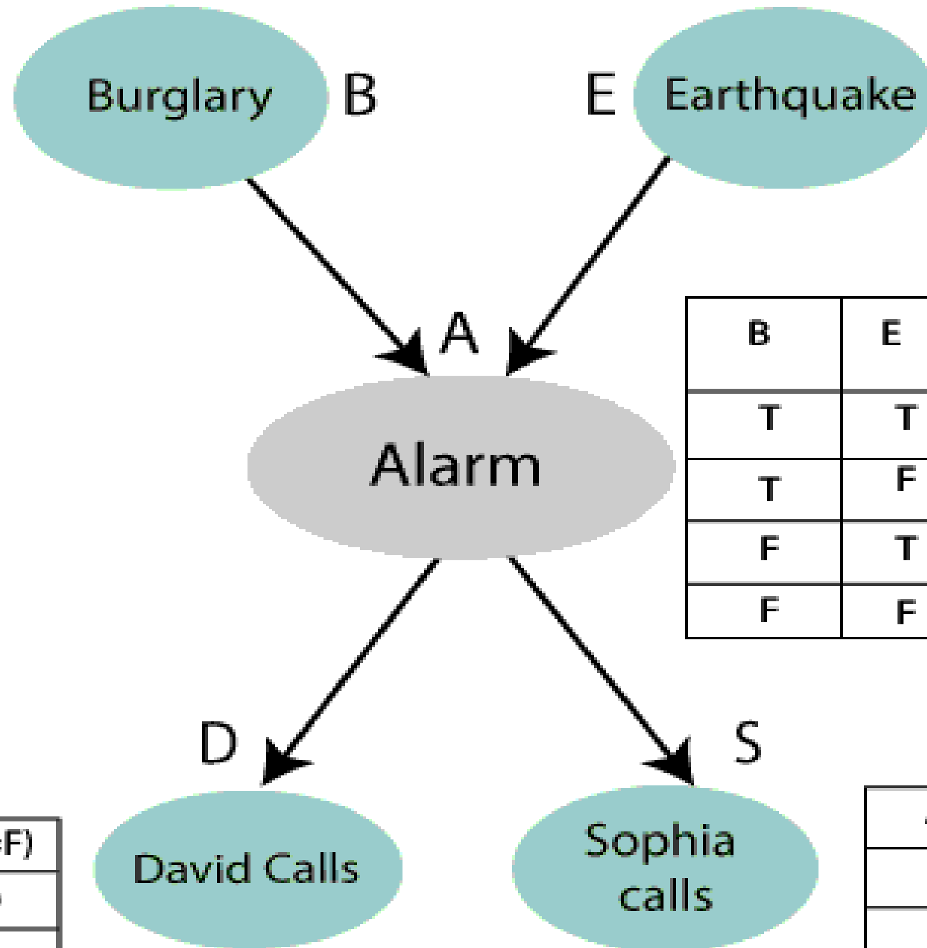
- **Burglary (B)**
- **Earthquake(E)**
- **Alarm(A)**
- **David Calls(D)**
- **Sophia calls(S)**

- We can write the events of problem statement in the form of probability: **P[D, S, A, B, E]**, can rewrite the above probability statement using joint probability distribution:

- **P[D, S, A, B, E]= P[D | S, A, B, E]. P[S, A, B, E]**
- **=P[D | S, A, B, E]. P[S | A, B, E]. P[A, B, E]**
- **= P [D| A]. P [S| A, B, E]. P[A, B, E]**
- **= P[D | A]. P[S | A]. P[A| B, E]. P[B, E]**
- **= P[D | A]. P[S | A]. P[A| B, E]. P[B |E]. P[E]**

T	0.002
F	0.998

T	0.001
F	0.999



B	E	P(A=T)	P(A=F)
T	T	0.94	0.06
T	F	0.95	0.04
F	T	0.69	0.69
F	F	0.999	0.999

A	P (D=T)	P (D=F)
T	0.91	0.09
F	0.05	0.95

A	P (S=T)	P (S=F)
T	0.75	0.25
F	0.02	0.98

- Let's take the observed probability for the Burglary and earthquake component:

- P(B= True) = 0.002, which is the probability of burglary.
- P(B= False)= 0.998, which is the probability of no burglary.
- P(E= True)= 0.001, which is the probability of a minor earthquake
- P(E= False)= 0.999, Which is the probability that an earthquake not occurred.

- We can provide the conditional probabilities as per the below tables:

- **Conditional probability table for Alarm A:**
- The Conditional probability of Alarm A depends on Burglar and earthquake:

Conditional probability table for Alarm A:

The Conditional probability of Alarm A depends on Burglar and earthquake:

B	E	P(A= True)	P(A= False)
True	True	0.94	0.06
True	False	0.95	0.04
False	True	0.31	0.69
False	False	0.001	0.999

Conditional probability table for David Calls:

The Conditional probability of David that he will call depends on the probability of Alarm.

A	P(D= True)	P(D= False)
True	0.91	0.09
False	0.05	0.95

Conditional probability table for Sophia Calls:

The Conditional probability of Sophia that she calls is depending on its Parent Node "Alarm."

A	P(S= True)	P(S= False)
True	0.75	0.25
False	0.02	0.98

- From the formula of joint distribution, we can write the problem statement in the form of probability distribution:
- **P(S, D, A, ¬B, ¬E) = P (S|A) *P (D|A)*P (A|¬B ^ ¬E) *P (¬B) *P (¬E).**
- = 0.75* 0.91* 0.001* 0.998*0.999
- **= 0.00068045.**

- **Hence, a Bayesian network can answer any query about the domain by using Joint distribution.**

- **The semantics of Bayesian Network:**
- There are two ways to understand the semantics of the Bayesian network, which is given below:

- **1. To understand the network as the representation of the Joint probability distribution.**
- It is helpful to understand how to construct the network.

- **2. To understand the network as an encoding of a collection of conditional independence statements.**
- It is helpful in designing inference procedure.