Artificial Intelligence

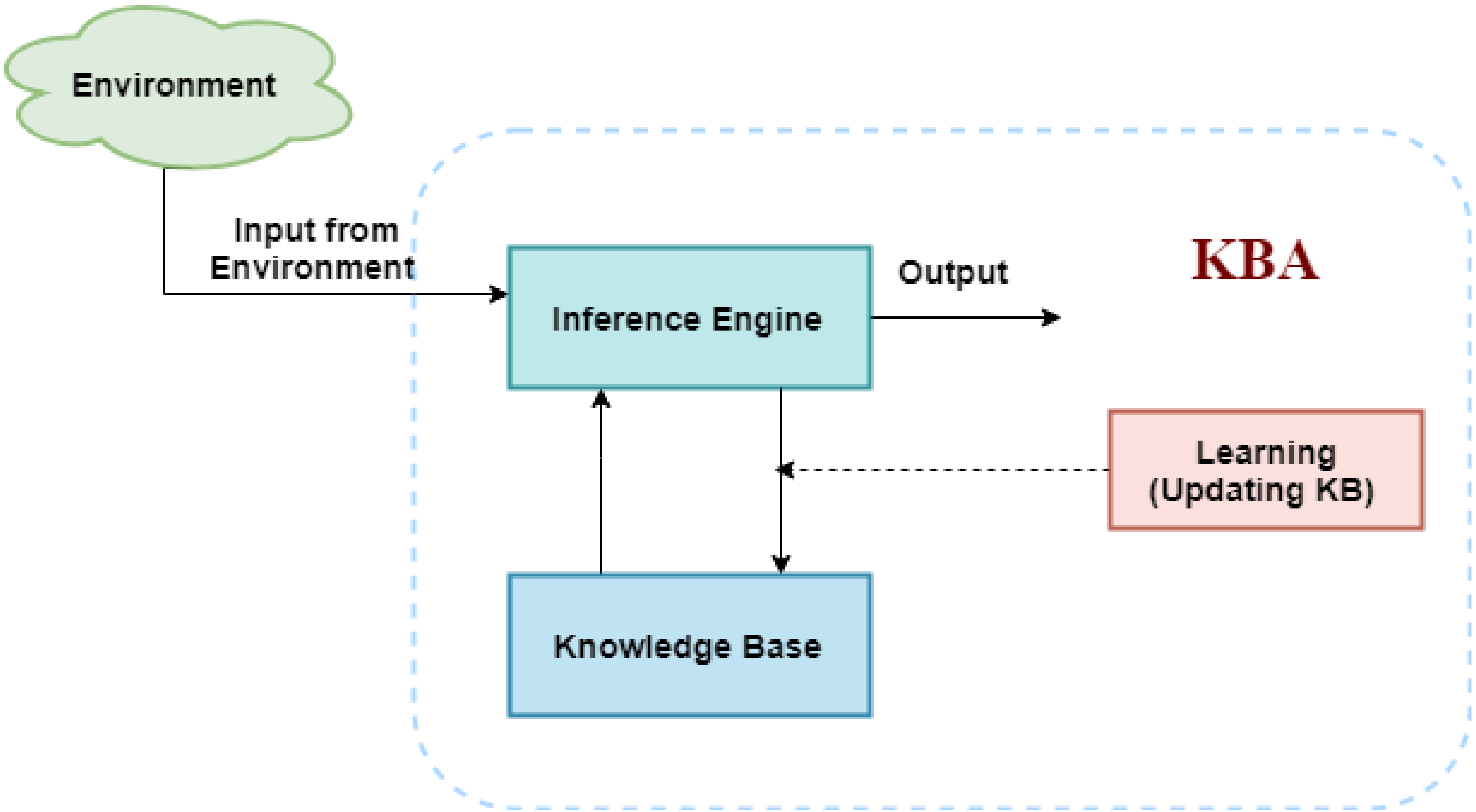
Introduction

Knowledge Based Agent

- An intelligent agent requires **knowledge** about the real world for taking decisions and **reasoning** to act efficiently.

- Knowledge-based agents have two main parts:
  - **Knowledge-base** and
  - **Inference system**

Knowledge Based Agent



Knowledge Based Agent

- **Knowledge-base** is a central component of a knowledge-based agent.
- It is a collection of sentences.
- The Knowledge-base stores **fact** about the world.
- **Inference** means deriving new sentences from old.
- A sentence is a proposition (**factual data**) about the world.
- **Inference system** applies logical rules to the KB to deduce new information.

Knowledge Based Agent

Knowledge Representation :

1. Relational / Frame Knowledge : Can be in **Tables**
2. Inheritable /Semantic Network Representation : In form of **Tree, Graphs Structures, Objects**
3. Production Rules - "**If condition then action**“
4. Propositional Logic : Factual Knowledge which is can be reasoned. (eg. Sun is green – **True/False**)

Probabilistic Reasoning

- Knowledge representation using first-order logic and propositional logic is with **certainty**.
- 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.

Probabilistic Reasoning

Causes of uncertainty

Following are some leading causes of uncertainty to occur in the real world.

1. Information occurred from unreliable sources.
2. Experimental Errors
3. Equipment fault
4. Environmental Factors

Probabilistic Reasoning

Probabilistic reasoning

- Probabilistic reasoning is a way of knowledge representation where we **apply the concept of probability** to indicate the uncertainty in knowledge.

- 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.

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**.

Probabilistic Reasoning

Two Schools of Probabilistic Thought

- There are two ways of interpreting probability:
- (1) **Frequentist probability,** which deals with actual outcomes from a sample space.
- (2) **Bayesian probability** which considers how strongly we believe that an event will occur.

Probabilistic Reasoning

Two Schools of Probabilistic Thought

- **Frequentist** probability includes techniques like **p-values** and **confidence intervals** used in statistical inference and **maximum likelihood** estimation for **parameter estimation**.
- **Frequentist** techniques are based on sampling – hence the frequency of occurrence of an event.
- **Frequentist** techniques are based on counts and
- **Bayesian** techniques are based on beliefs.
- **Bayesian techniques** are based on the Bayes' theorem.
- **Bayesian analysis** can be used to model events that have not occurred before or occur infrequently.

Probabilistic Reasoning

- **Probability quantifies uncertainty**.
- Probability is a **measure of uncertainty**.
- Probability applies to machine learning because in the real world, we need to **make decisions with incomplete information**.
- In contrast, in **traditional programming**, we work with **deterministic** problems i.e. **the solution is not affected by uncertainty**.
- Probability quantifies the likelihood or belief that an event will occur.

$$Probability = \frac{Occurrences}{Total\ no.\ of\ possible\ Outcomes}$$

Probabilistic Reasoning

- Probability theory has three important concepts:
- (1) **Event** – an outcome to which a probability is assigned. (I got Heads)
- (2) **Sample Space** which represents the set of possible outcomes for the events and the Probability Function which maps a probability to an event. {Heads, Tails}
- (3) **Probability Function** indicates the likelihood that the event being a part of the sample space is drawn.
- The **probability of an event** can be calculated directly by counting all the occurrences of the event and dividing them by the total possible outcomes of the event. **P(Tossing a Coin) = Getting a Head (1)/(2) {Heads, Tails}**
- Probability is a fractional value and has a value in the range between 0 and 1, where
- **0 indicates no probability** and
- **1 indicates full probability**.

Probabilistic Reasoning

- Probability is often written as a lowercase **p** and may be stated as a percentage by multiplying the value by 100.

- A probability of 0.3 can be stated as 30% (given 0.3 100)

Probabilistic Reasoning

$0 \leq P(A) \leq 1$, 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.

$$\textbf{Probability of occurrence} = \frac{\text{Number of desired outcomes}}{\text{Total number of outcomes}}$$

- $P(\neg A)$ = probability of a not happening event.
- $P(\neg A) + P(A) = 1$.

Probability

- Probability represents the certainty factor.
- For eg. Rolling a dice and you say that the certainty with which a 6 shows up on the dice is ⅙. It means there's a 16.67% chance that a 6 shows up on the dice. That's the certainty you allot to that particular event.
- This, in turn, is known as **probability**, or precisely, in our case, it's called **frequentist probability**.
- The frequentist probability denotes the frequency with which the event can happen amongst many trials/events. Rolling a dice is frequentist as ⅙ means that out of infinitely many trials of rolling a dice, there's a 1/6th chance that 6 is going to show up.

Probability

- Tossing a coin = Getting heads or tails is ½ = 50%
- Throwing a Dice = Getting 2 is 1/6, Getting 5 is 1/6, Getting odd or even number is 3/6=1/2
- Choosing an ace =4/52=1/13

$$\textbf{Probability of occurrence} = \frac{\text{Number of desired outcomes}}{\text{Total number of outcomes}}$$

Probability

Relative Frequency –

- England won 8 games from 12 in soccer.
- Hence Relative Frequency is 8/12= 66.66%
- All relative frequency add to 1
- Eg
- 15 students choose Java as elective
- 30 students choose python as elective
- 5 students choose Javascript programming as elective
- $15/50 + 30/50 + 5/50 = 0.3 + 0.6 + 0.1 = 1$

Probability

Eg.

- Tossing a Coin : Getting Heads (1/2)
- Throwing a Dice : Getting 4 (1/6)
- There are 9 balls in a bag (7 are Green, 2 are Red). What is the probability that Green is picked. (7/9)
- Choosing an ACE from a pack of cards.

Probability

Event – Event means one or more outcome/s.

Eg. Getting a Head while tossing a coin is an event.

Types of Event–

- 1. **Independent** – One event is not affect in any way by other event. They are isolated events.

  Eg Rolling a dice 3 times and getting event is totally isolated from previous output.
- 2. **Dependent** – One event is affected by other event. Also known as conditional.

  Eg. Drawing 2 cards from 52 without putting back and event of getting ACE both the times.
- 3. **Mutually Exclusive** – Events cannot happen at same time. Ie only one event can happen at a time. Eg Heads and Tails at one toss is mutually exclusive.

Probability

- **Joint probability** : Probability of event occurring simultaneously.
- **Marginal Probability** : Is the probability of event irrespective of the outcome of other variable
- **Conditional Probability** : is the probability of event occurring in the presence of a second event.

How to calculate Conditional/Likelihood Probability

$$P(A|B) = P(A \cap B) / P(B)$$

$P(A|B)$ = probability of A if event B has already occurred)

$P(A \cap B)$ = Count of common elements in A and B

$P(B)$ – Probability of event B (count of elements in B)

Probability

Dependent Events (Conditional Probability) : Eg.

Eg 1. Tossing a coin 2 times (Tree for Independent event)

Eg 2. Below (Tree for Dependent event)

- Imagine you want a DS job but that depends on Data Science certification.
- 1. With BTech course the getting a Data Science Job the probability is 0.6 over BCA course.
- 2. With DS certification given a BTech course the probability of getting a job is 0.6.
- 3. With DS certification given a BCA course the probability of getting a job is 0.4

Probability

Independent Event : Each Single event is not affected by any other event.

Dependent Event : First part of the event affects the second part.
Eg. Drawing an ACE from a deck of 52 cards. 4/52
Again drawing ACE from the same deck. 3/51

Mitually Exclusive Events : Events can’t happen at same time.
eg. While rolling a die, I cant get 1 and 5 simultaneously. Hence 1 and 5 are mutually exclusive events.

Probability

Bayes Theorem :

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

- Here P(A|B) is a conditional probability : The probability of event A occurring given that B is true. It is also called the posterior probability of A given B.
- P(B|A) is also a conditional probability : The probability of event B occurring given that A is true. It can also be interpreted as the likelihood of A given a fixed B because P(B|A)=L(A|B)
- P(A) and P(B) are the probabilities of observing A and B respectively without any given conditions, they are known as the marginal probability or prior probability.

Probability

Bayes Theorem : To calculate conditional probability. Widely used in ML.

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

Eg. $$P(Rain|Cloud) = \frac{P(Cloud|Rain) * P(Rain)}{P(Cloud)}$$

Here,

P(A|B) is a posterior probability (conditional probability)

P(A) is a prior probability (Marginal Probability – calculated irrespective of outcome of other variable)

P(B|A) is a likelihood (conditional probability)

P(B) Evidence (Marginal Probability)

Probability

Bayes Theorem : To calculate conditional probability. Widely used in ML.

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

Eg.

$$P(Rain|Cloud) = \frac{P(Rain)\ *P(Cloud|Rain)}{P(Cloud)}$$

- P(Rain)=30%, (rain in sept is around 9 days from old collected data)
- P(Cloud|Rain)=100%, (probability of cloud given rain is 100%)
- P(Cloud)=50% (In September 11-13 days are cloudy)

$P(Rain|Cloud) = \frac{.3*1}{.5}$ = .6 ie 60% possibility of rain on any particular day.

N-Gram Model

N-Gram is sequence of elements appearing contiguously.

The value of N can vary.

- If **N=1** it is **unigram**.
- If **N=2** It is **bigram** **….trigram…**

Eg Sentence : It is the best idea to start with excel data.

- [“It”, “is”, “the”, “best” ……] **N=1** Unigram
- [“It is”, “is the”, “the best”, “best idea” ……..] **N=2** Bigram
- [“It is the”, “is the best”, “the best idea”, “best idea to”] **N=3** Trigram

N-Gram Model

The gram can be character or word also depending on use-case.

Eg : Python

["P", "y", "t", "h" ......] **N=1** Unigram

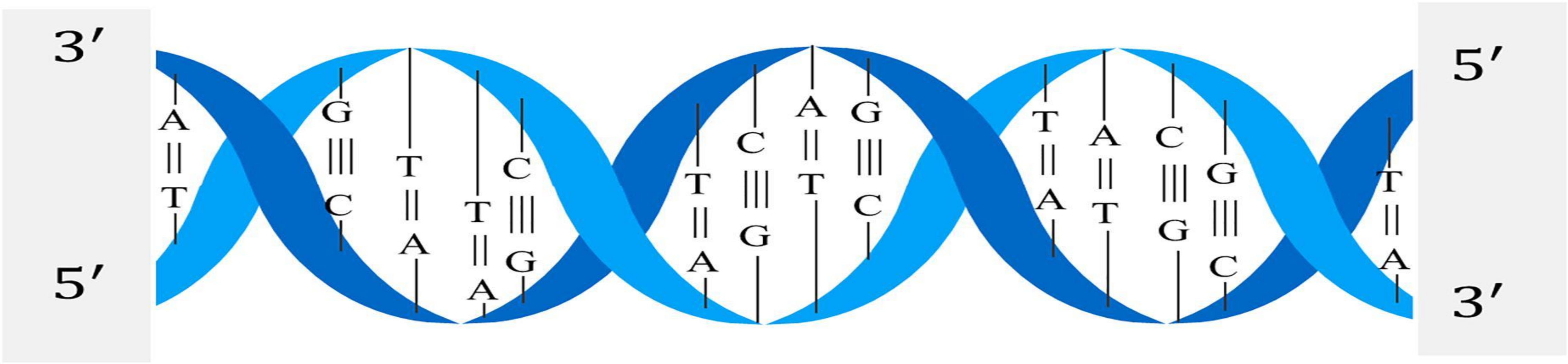
["Py", "yt", "th", "ho", "on"] **N=2** Bigram

["pyt", "yth", "tho", "hon"] **N=3** Trigram

N-Gram Model

Applications

- As it is a type of probabilistic model.
- It is **used for predicting** the **sequence of next character**, **next word**, **next POS** (part of speech) tag for the word. (Markov Model)
- Used of **biological DNA sequence prediction**.
- **Speech word prediction** etc.
- **Spell and Grammar verifying system.**



Applied View :

Conditional Probability : The occurrence of event A, where event B had already happened.

- $P(A|B) = P(A \cap B) \,/\, P(A)$

or

- $\mathbf{P(Wi \mid Wi\text{-}1) = C(Wi\text{-}1, Wi) \,/\, C(Wi\text{-}1)}$

N-Gram Model (Markov Model)

Conditional Probability

Eg. This is a best way to do work.

Please do work you like the most is to eat.

To do work is a best deed.

$P(W_i \mid W_{i-1}) = C(W_{i-1}, W_i) / C(W_{i-1})$

Where C=Count, P=Probability

Eg P(work | do)= Count("do work") / Count("do")

P(work | do) = 3/3 = 1.0 (Wi-1 = do, Wi=Work)

P(deed | best) = 1/2= 0.5 (Wi-1 = best, Wi=deed)

P(a | is) = 2/3 = 0.67

Till This Slide in Mid Sem

Bayesian Belief Network

- Deals with probabilistic events to solve problems with uncertainty.
- Is a **probabilistic graphical model** to represent set of variables and their conditional dependencies **using directed acyclic graph**.
- The **alternate names** for this concept are **Bayes Network, Belief Network, Decision Network or Bayesian model**.

Bayesian Belief Network

- Bayesian Belief Network consists of
- 1. **Directed Acyclic Graph**
- 2. **Table of Conditional Probabilities**.
- The node in a graph represents random variables (**discrete** or **continuous)**
- Arrows represent conditional probabilities between random variables.
- Each node in a BB network has conditional probability distribution

Eg P(Xi | Parent Xi). Ie The effect of current (child node) is dependent on the given parent node.

Bayesian Belief Network

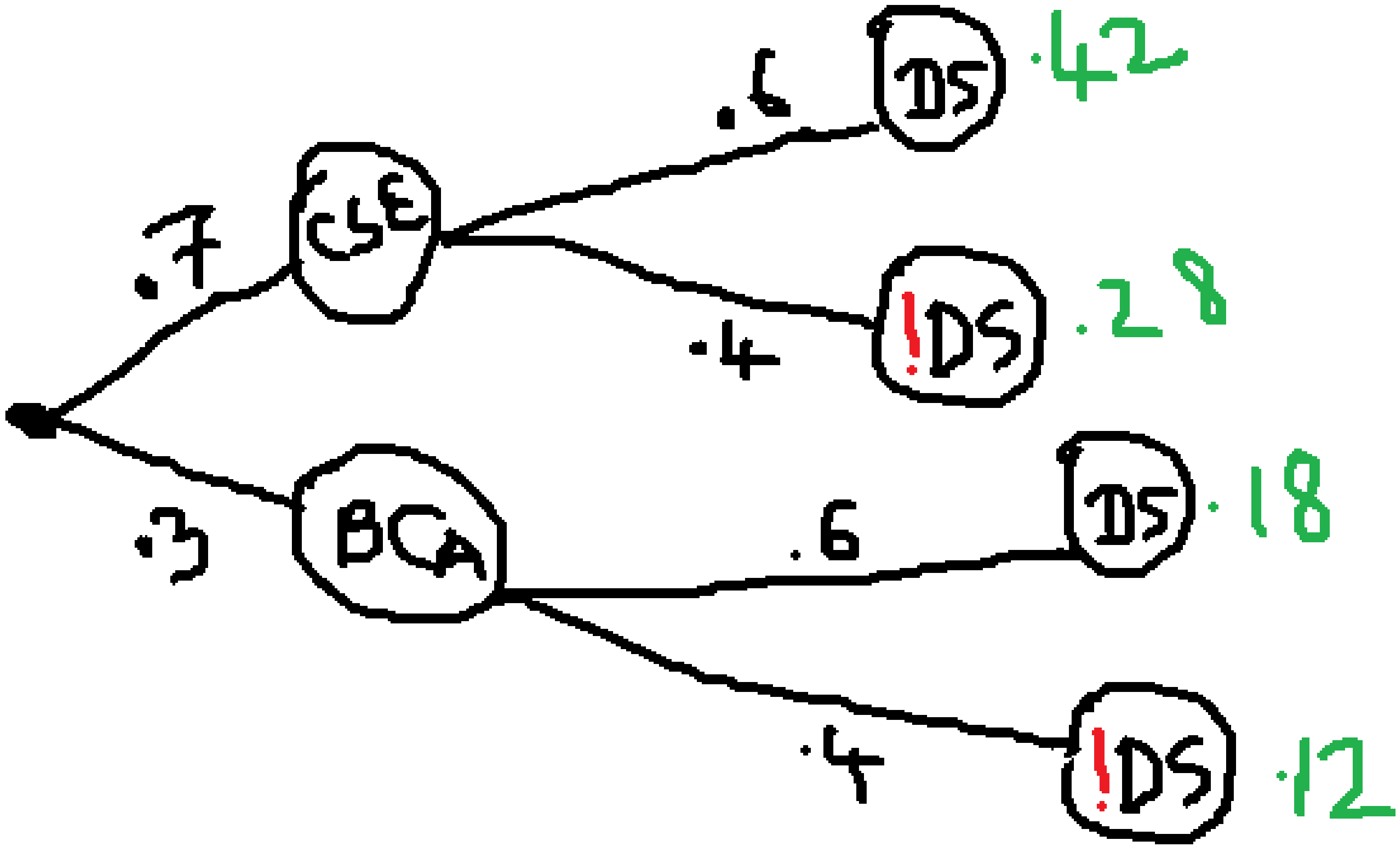
- BBN is based on Joint probability and Conditional probability distribution.
- In Joint Probability Distribution, if x1, x2, x3 are variables then different combination of x1, x2 and x3 are known as joint probabilities.
- Eg of Joint Prob
- Probability getting a card which is 8 and also red.
- **$P(8 \cap red)$=**$P(8) \times P(red) = (4/52) \times (26/52) = 1/26$
- Eg of Conditional Prob
- Probability getting a card which is 8 given that the card is red.
- **$P(8|red)$** = (Count of red 8's / count of reds) = 2/26 = 1/13

Bayesian Belief Network

- Hence for multiple variables
- Eg x1,x2,x3…..xn
- =P(x1 |x2,x3..xn) x P(x2,x3….xn)
- = P(x1 |x2,x3…xn) x P(x2 |x3,x4….xn)
- =P(xi | Parents(xi))

- Probability of getting a Job if students has done Btech-CSE or BCA course provided also passed certified DataScience course from IBM.

Bayesian Belief Network



Hidden Markov Model

Brief of Automata

1. Electric switch Ex   2. Fan Regulator ex

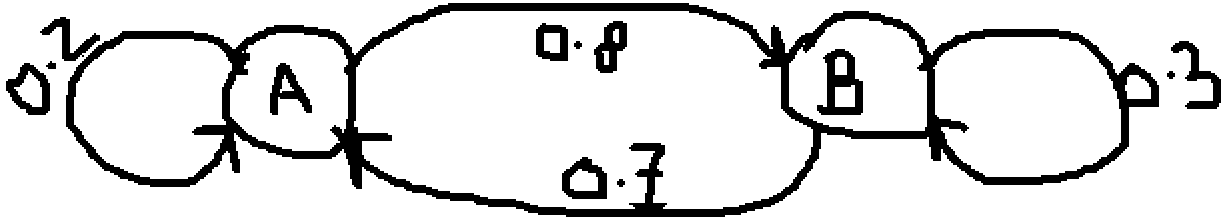
Markov Chain :

- It is used to model random/probabilistic process.
- A Markov chain is a set of transition between the states with some probability distribution that satisfy Markov property.
- Markov Property is a property where the current state depends on previous state only.
- It is an probabilistic automaton.

Hidden Markov Model

Markov Chain :

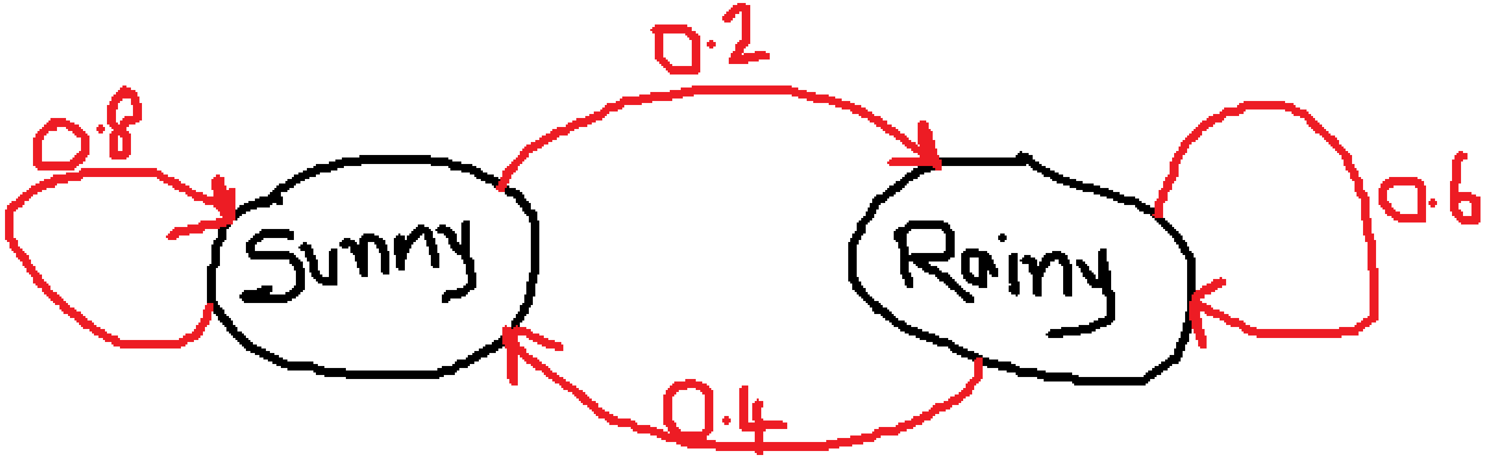
	A	B	
A	P(A|A)= 0.5	P(B|A)= 0.5	1
B	P(A|B)= 0.5	P(B|B)= 0.5	1



	A	B	
A	P(A|A)= 0.2	P(B|A)= 0.8	1
B	P(A|B)= 0.7	P(B|B)= 0.3	1

Hidden Markov Model

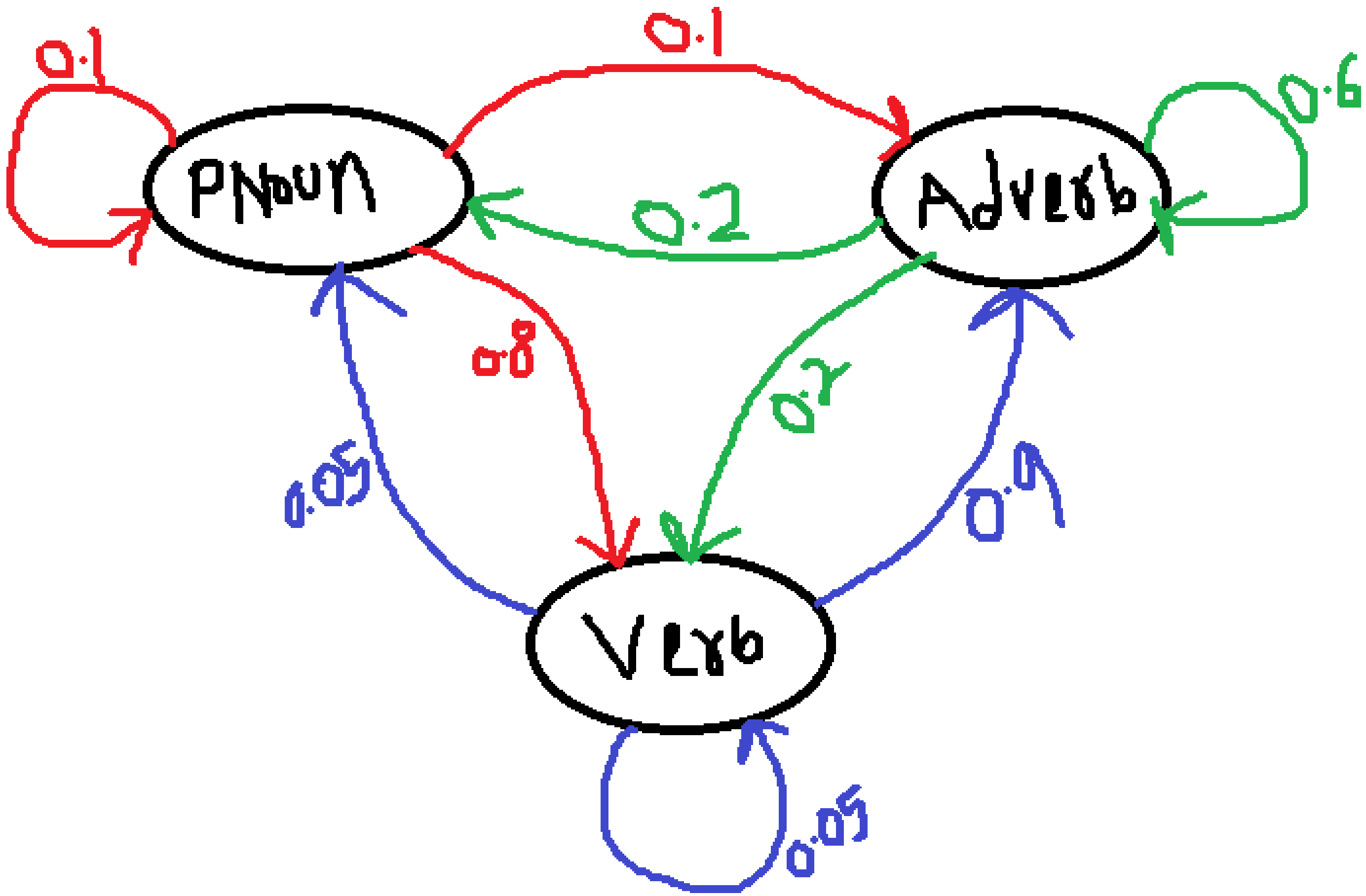
Markov Chain :



	Sunny	Rainy	
Sunny	P(S|S)=0.8	P(R|S)=0.2	1
Rainy	P(S|R)=0.4	P(R|R)=0.6	1

Hidden Markov Model

Markov Chain : Eg. Ram<Pnoun> runs<Verb> fast<Adverb>



	PNoun	Verb	Adverb	
PNoun	$P(PN \mid PN)=0.1$	$P(V \mid PN)=0.8$	$P(A \mid PN)=0.1$	1
Verb	$P(PN \mid V)=0.05$	$P(V \mid V)=0.05$	$P(A \mid V)=0.9$	1
Adverb	$P(PN \mid A)=0.2$	$P(V \mid A)=0.2$	$P(A \mid A)=0.6$	1

Applications

Applications

- **Pattern recognition**
- Probability theory is a key part of pattern recognition
- Pattern recognition is a key part of machine learning.
- We can approach machine learning as a pattern recognition problem from a Bayesian standpoint.
- because it helps to cater for noise / uncertainty
- **And if the dataset is small**

Application

- **Training – use in Maximum likelihood estimation**
- Many iterative machine learning techniques like Maximum likelihood estimation (MLE) are based on probability theory.
- MLE is used for training in models like linear regression, logistic regression and artificial neural networks.
- **Developing specific algorithms**
- Probability forms the basis of specific algorithms like Naive Bayes classifier
- **Hyperparameter optimization**
- In machine learning models such as neural networks, hyperparameters are tuned through techniques like grid search. Bayesian optimization can be also used for hyperparameter optimization.

References

1. Stuart Russel and Peter Norvig, “Artificial Intelligence – A Modern Approach”, 3rd edition, Pearson Education.