Probability reasoning

DR. VENKATARAMANA VEERAMSETTY

Probability

Probability implies 'likelihood' or 'chance' for an event occurrence

It states that if there are n mutually exclusive and equally likely cases out of which m cases are favorable to the happening of event A, Then the probabilities of event A is defined as given by the following probability function:

$$P(A) = \frac{Number\ of\ favourable\ cases}{Total\ number\ of\ equally\ likely\ cases} = \frac{m}{n}$$

Problem Statement

A coin is tossed. What is the probability of getting a head?

Solution:

Total number of equally likely outcomes (n) = 2 (i.e. head or tail)

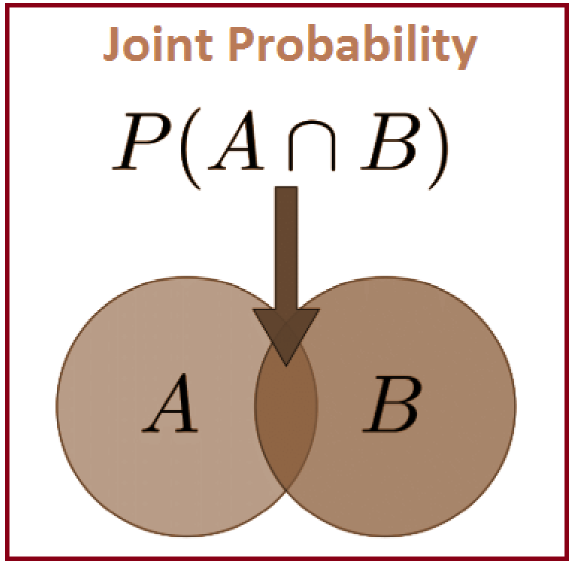
Number of outcomes favorable to head (m) = 1

P(head)= ½ =0.5

Joint Probability

A statistical measure that calculates the likelihood of two events occurring together and at the same point in time is called Joint probability.

Let A and B be the two events, joint probability is the probability of event B occurring at the same time that event A occurs. It is represented as $P(A \cap B)$



Problem Statement

Find the probability that the number three will occur twice when two dice are rolled at the same time.

Solution:

Let A be the event of occurring 3 on first die and B be the event of occurring 3 on the second die.

$P(A) = 1/6$

$P(B)=1/6$

$P(A \cap B) = P(A) * P(B) = 1/6 * 1/6 = 1/36$

Mutually Exclusive Events

When the occurrence is not simultaneous for two events then they are termed as **Mutually exclusive events.**

For example, when a coin is tossed then the result will be either head or tail, but we cannot get both the results. Such events are also called **disjoint events** since they do not happen simultaneously.

The mathematical formula for mutually exclusive events can be represented as $P(X \text{ and } Y) = 0$

Real time examples

When tossing a coin, the event of getting head and tail are mutually exclusive. Because the probability of getting head and tail simultaneously is 0.

In a six-sided die, the events “2” and “5” are mutually exclusive. We cannot get both the events 2 and 5 at the same time when we threw one die.

In a deck of 52 cards, drawing a red card and drawing a club are mutually exclusive events because all the clubs are black.

Additive Theorem - I

The additive theorem of probability states if A and B are two mutually exclusive events then the probability of either A or B is given by

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$$

If the **two or more events that cannot happen simultaneously,** such type of events are called mutually exclusive events

Problem Statement

A card is drawn from a pack of 52, what is the probability that it is a king or a queen?

- Event (A) = Draw of a card of king
- Event (B) Draw of a card of queen
- Probability to draw king P(A)=4/52
- Probability to draw queen P(B)=4/52
- P (card draw is king or queen) = P (card is king) + P (card is queen)
- P(A or B) = P(A U B) = P(A) + P(B) = 4/52 + 4/52 =2/13

Additive Theorem - II

The additive theorem of probability states if A and B are two non-mutually exclusive events then the probability of either A or B is given by

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \text{ and } B)$$

Non-mutually exclusive events are **events that can happen at the same time**

Problem Statement

A shooter is known to hit a target 3 out of 7 shots; whereas another shooter is known to hit the target 2 out of 5 shots. Find the probability of the target being hit at all when both of them try.

Solution

- Probability of first shooter hitting the target P (A) = 3/7
- Probability of second shooter hitting the target P (B) = 2/5
- Event A and B are not mutually exclusive as both the shooters may hit target.
-

$$
\begin{aligned}
&P(A \cup B) = P(A) + P(B) - P(A \cap B) \\
&= \tfrac{3}{7} + \tfrac{2}{5} - \left(\tfrac{3}{7} \times \tfrac{2}{5}\right) \\
&= \tfrac{29}{35} - \tfrac{6}{35} \\
&= \tfrac{23}{35}
\end{aligned}
$$

Independent Events

Independent events are those events whose occurrence is not dependent on any other event.

If the probability of occurrence of an event A is not affected by the occurrence of another event B, then A and B are said to be independent events.

The mathematical formula for independent events can be represented as P(A and B) = P(A)*P(B) and $P(A|B) = P(A)$

An example of a mutually exclusive event is when a coin is a tossed and there are two events that can occur, either it will be a head or a tail. Hence, both the events here are mutually exclusive. But if we take two separate coins and flip them, then the occurrence of Head or Tail on both the coins are independent to each other.

Problem Statement

Consider an example of rolling a die. If A is the event ‘the number appearing is odd’ and B be the event ‘the number appearing is a multiple of 3’, then what is the probability for the event ‘the number appearing is odd and a multiple of 3'

- $P(A) = 3/6 = 1/2$ and $P(B) = 2/6 = 1/3$
- $P(A \text{ and } B) = P(A \cap B) = P(A)*P(B) = 1/6$
- $P(A|B) = P(A \cap B)/ P(B) = (1/6)/(1/3) = ½ = P(A)$
- $P(B|A) = P(A \cap B)/ P(A) = (1/6)/(1/2) = 1/3 = P(B)$
- $P(A|B) * P(B) = P(A \cap B)$
- $P(B|A) * P(A) = P(A \cap B)$

Problem Statement

Let X and Y are two independent events such that P(X) = 0.3 and P(Y) = 0.7. Find P(X and Y), P(X or Y), P(Y not X), and P(neither X nor Y).

- P(X and Y) = $P(X \cap Y) = P(X) * P(Y) = 0.3*0.7=0.21$
- P(X or Y) = $P(X \cup Y) = P(X) + P(Y) - P(X \cap Y) = 0.3+ 0.7 - 0.21 = 0.79$
- P(Y not X) = P(Y and X')= $P(Y \cap X') = P(Y)*(1-P(X)) = 0.7*0.7=0.49$
- P(neither X nor Y) = P(Y' and X')= $P(Y' \cap X') =(1-P(Y))*(1-P(X)) = 0.3*0.7=0.21$

Problem Statement

A college has to appoint a lecturer who must be B.Com., MBA, and Ph. D, the probability of which is 1/20, 1/25 and 1/40 respectively. Find the probability of getting such a person to be appointed by the college.

$P(B.Com. \cap MBA \cap Ph.\ D)=P(B.Com.)*P(MBA)*P(Ph.\ D)=1/20 * 1/25 * 1/40 = 0.00005$

Dependent Events & Conditional Probability

Dependent events are the events where the occurrences or nonoccurrence of one event effects the outcome of next event.

$P(A \cap B) \mathrel{!}= P(A)*P(B)$

The probability associated with such events is called as conditional probability and is given by

- $P(A/B) = P(A \cap B)/P(B)$ [Read $P(A/B)$ as the probability of occurrence of event A when event B has already occurred.]
- $P(B/A) = P(A \cap B)/P(A)$ [Read $P(B/A)$ as the probability of occurrence of event B when event A has already occurred.]

Problem Statement

A coin is tossed 2 times. The toss resulted in one head and one tail. What is the probability that the first throw resulted in a tail?

Solution

- Sample space={HH,HT,TH,TT}
- Event A be the first throw resulting in a tail.
- Event B be that one tail and one head occurred.

$$P(A) = \frac{P(TH,TT)}{P(HH,HT,TH,TT)} = \frac{2}{4} = \frac{1}{2}$$

$$P(A \cap B) = \frac{P(TH)}{P(HH,HT,TH,TT)} = \frac{1}{4}$$

$$So\ P(A/B) = \frac{P(A \cap B)}{P(A)}$$

$$= \frac{\frac{1}{4}}{\frac{1}{2}}$$

$$= \frac{1}{2} = 0.5$$

Probability Bayes Theorem

The Bayes Theorem was developed by a British Mathematician Rev. Thomas Bayes.

The probability given under Bayes theorem is also known by the name of inverse probability, posterior probability or revised probability.

The bayes theorem is based on the formula of conditional probability.

- P(A/B)=[P(B/A)*P(A)]/P(B)

Proof:

- P(A/B) = P(A ∩ B)/P(B) ..... (1)
- P(B/A) = P(A ∩ B)/P(A) .....(2)
- Eq.1 /eq.2 leads bayes rule

$$\underbrace{\boxed{P(A|B)}}_{\text{posterior}} = \underbrace{\boxed{P(A)}}_{\text{prior}} \times \frac{\overbrace{\boxed{P(B|A)}}^{\text{likelihood}}}{\underbrace{\boxed{P(B)}}_{\text{marginal}}}$$

Posterior probability (P(A/B)) (updated probability after the evidence (B) is considered)

Prior probability (P(A)) (the probability before the evidence is considered)

Likelihood (P(B/A)) (probability of the evidence, given the belief is true)

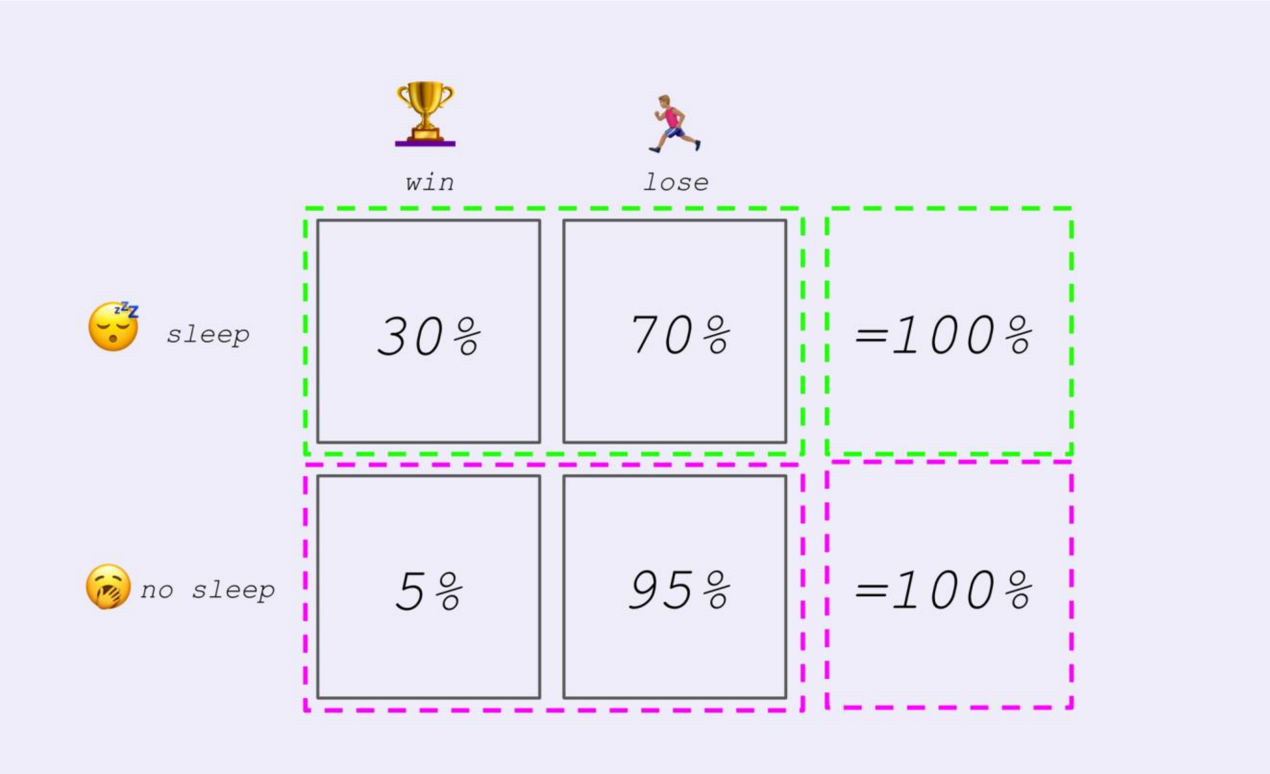
Marginal probability (P(B)) (probability of the evidence, under any circumstance)

Bayes' Rule lets you calculate the **posterior (or "updated") probability**. This is a conditional probability. It is the probability of the hypothesis being true, if the evidence is present.

prior (or "previous") probability as your belief in the hypothesis before seeing the new evidence. If you had a strong belief in the hypothesis already, the prior probability will be large.

The numerator is the **likelihood**. This is another conditional probability. It is the probability of the evidence being present, given the hypothesis is true.

Denominator is the **marginal probability** of the evidence. That is, it is the probability of the evidence being present, whether the hypothesis is true or false. The smaller the denominator, the more "convincing" the evidence.



Bayes theorem developed based on conditional probability.

Problem Statement

~~You might be interested in finding out a patient's probability of having liver disease if they are an alcoholic.~~ "Being an alcoholic" is the **test** (kind of like a litmus test) for liver disease.

A could mean the event "Patient has liver disease." Past data tells you that 10% of patients entering your clinic have liver disease. P(A) = 0.10.

B could mean the litmus test that "Patient is an alcoholic." Five percent of the clinic's patients are alcoholics. P(B) = 0.05.

You might also know that among those patients diagnosed with liver disease, 7% are alcoholics. This is your **B|A:** the probability that a patient is alcoholic, given that they have liver disease, is 7%.

Bayes' theorem tells you:
$P(A|B) = (0.07 * 0.1)/0.05 = 0.14$

In other words, if the patient is an alcoholic, their chances of having liver disease is 0.14 (14%). This is a large increase from the 10% suggested by past data.

Problem Statement

In a particular pain clinic, 10% of patients are prescribed narcotic pain killers. Overall, five percent of the clinic's patients are addicted to narcotics (including pain killers and illegal substances). Out of all the people prescribed pain pills, 8% are addicts. *If a patient is an addict, what is the probability that they will be prescribed pain pills?*

- *P(Pain Pills) = 0.1*
- *P(Addicts)= 0.05*
- *P(Addict|Painpills)=0.08*

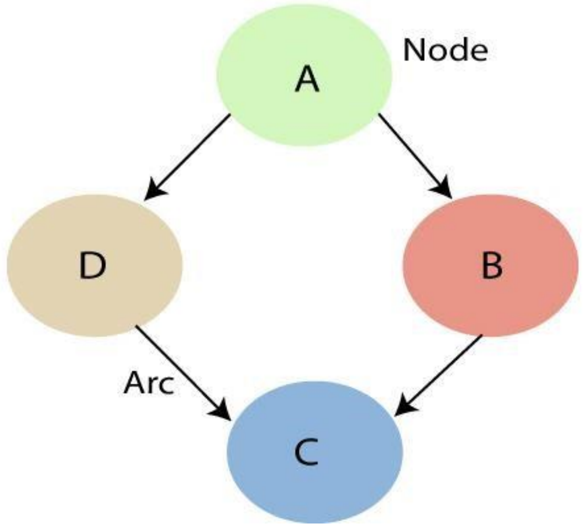
- *P(Painpills|Addict)= (P(Addict|Painpills)*P(Pain Pills))/P(Addicts) = 0.08*0.1/0.05=0.16*

Local Markov Property

It states that a node is conditionally independent of its non-descendants, given its parents.

In the given example, $P(D|A, B) = P(D|A)$

because D is independent of its non-descendent, B.

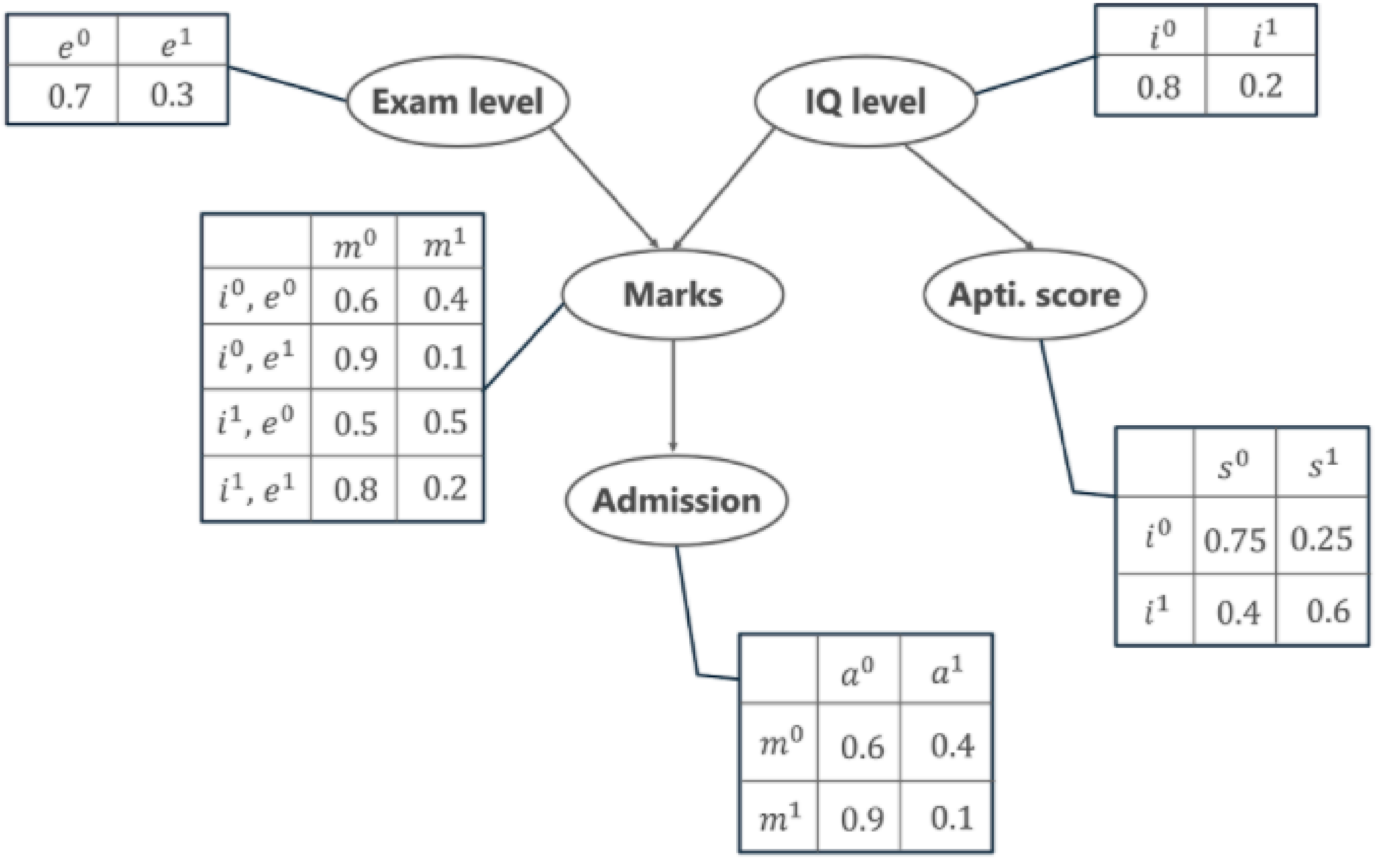


Bayesian Network

Bayesian Networks are a type of Probabilistic Graphical Model that uses the Bayesian inferences for probability computations.

It represents a set of variables and its conditional probabilities with a Directed Acyclic Graph (DAG).

Example



e^0	e^1
0.7	0.3

i^0	i^1
0.8	0.2

	m^0	m^1
i^0, e^0	0.6	0.4
i^0, e^1	0.9	0.1
i^1, e^0	0.5	0.5
i^1, e^1	0.8	0.2

	s^0	s^1
i^0	0.75	0.25
i^1	0.4	0.6

	a^0	a^1
m^0	0.6	0.4
m^1	0.9	0.1

Variables

- Exam Level (e)
- IQ Level (i)
- Aptitude Score (s)
- Marks (m)
- Admission (a)

Joint Probability Distribution

Five variables are represented in the form of a Directed Acyclic Graph (DAG) in a Bayesian Network format with their Conditional Probability tables.

- Exam Level (e)
- IQ Level (i)
- Aptitude Score (s)
- Marks (m)
- Admission (a)

Joint Probability Distribution

Now, to calculate the Joint Probability Distribution of the 5 variables the formula is given by,

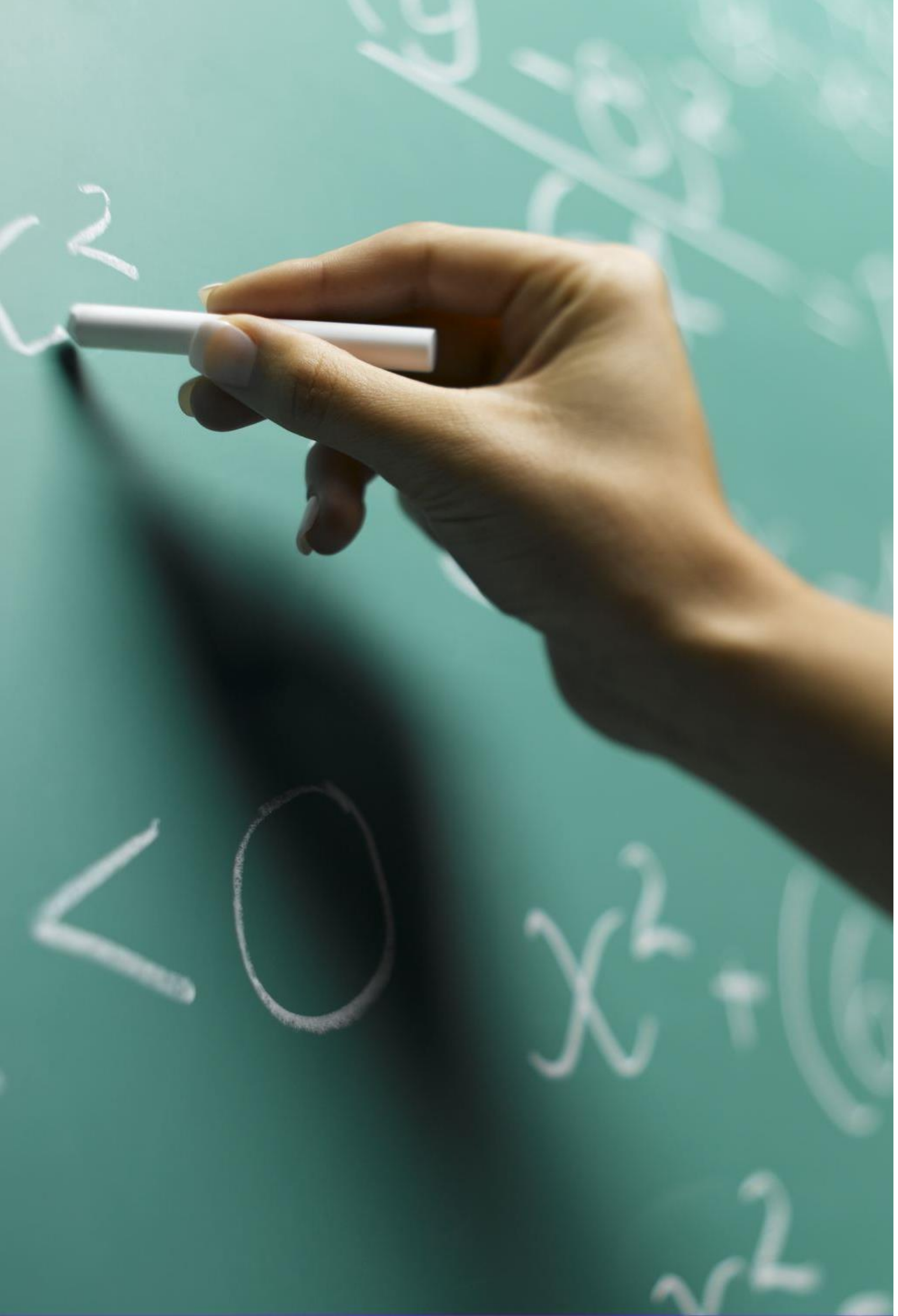
P[a, m, i, e, s]= P(a | m) . P(m | i, e) . P(i) . P(e) . P(s | i)

P(a | m) denotes the conditional probability of the student getting admission based on the marks he has scored in the examination.

P(m | i, e) represents the marks that the student will score given his IQ level and difficulty of the Exam Level.

P(i) and P(e) represent the probability of the IQ Level and the Exam Level.

P(s | i) is the conditional probability of the student's Aptitude Score, given his IQ Level.



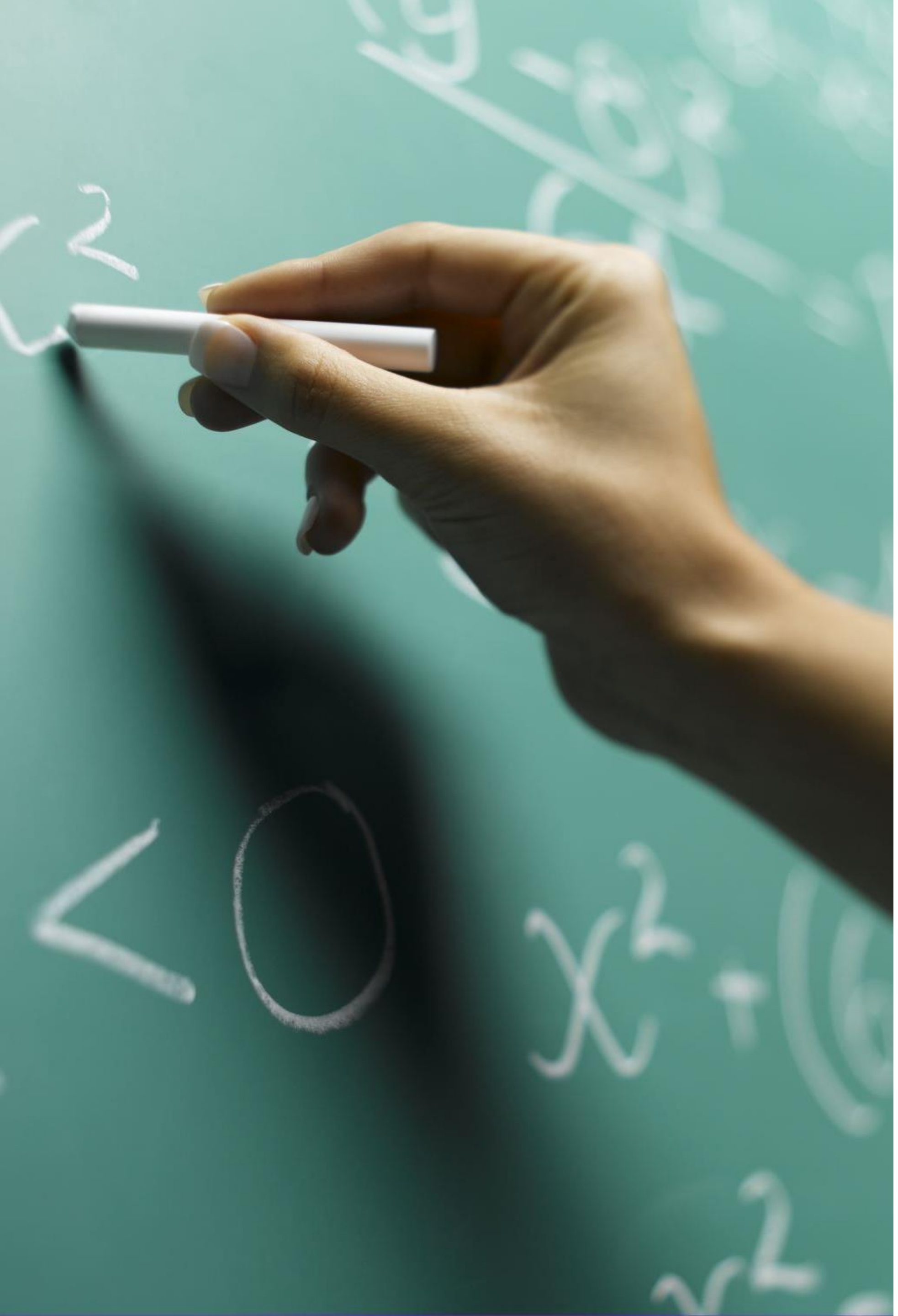
Case-1

Calculate the probability that in spite of the exam level being difficult, the student having a low IQ level and a low Aptitude Score, manages to pass the exam and secure admission to the university.

I.e $P[a=1, m=1, i=0, e=1, s=0]$

$P(a=1|m=1)*P(m=1|e=1,I=0)*P(e=1)*P(S=0|i=0)*P(i=0)$

$=0.1*0.1*0.3*0.75*0.8 = 0.0018$



Case-2

calculate the probability that the student has a High IQ level and Aptitude Score, the exam being easy yet fails to pass and does not secure admission to the university.

$P(I=1,s=1,m=0,e=0,a=0)$

$P(a=0|m=0)*P(m=0|e=0,I=1)*P(e=0)*P(I=1)*P(S=1|I=1)$

$0.6*0.5*0.7*0.2*0.6= 0.0252$

Temporal Models

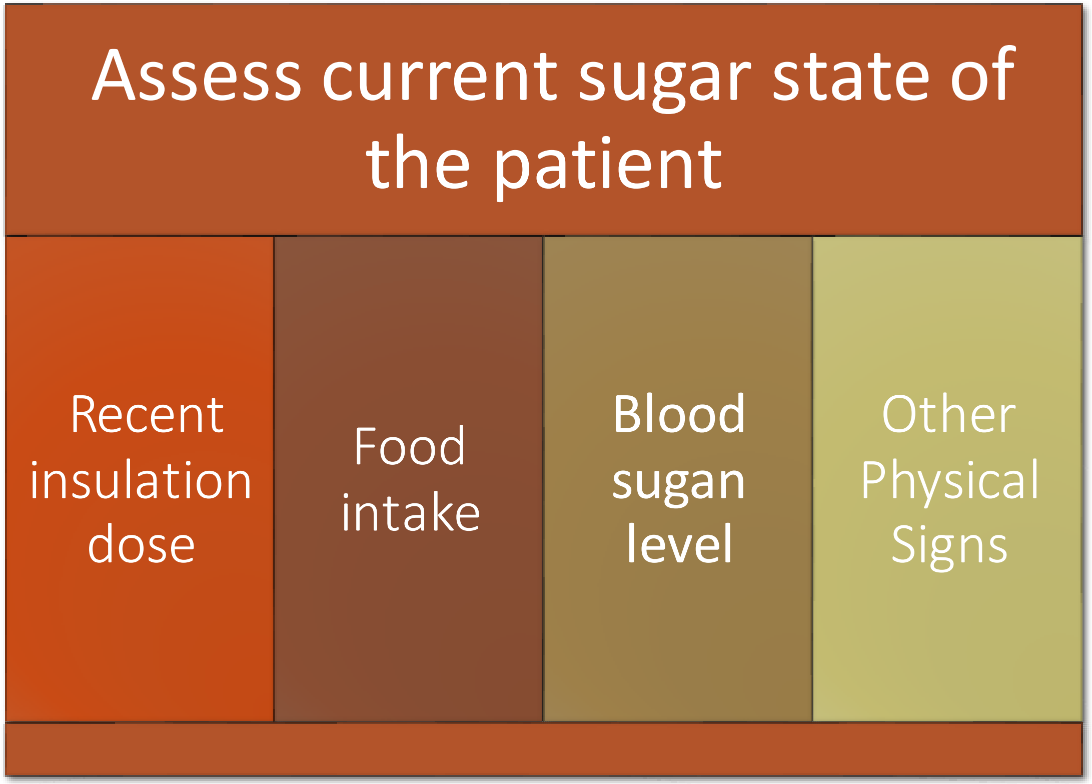
- Agents in uncertain environment must able to track current state of the environment
- This is difficult with partial or noisy data as environment is uncertain over time
- Hence agent will able to obtain probabilistic assessment of the current state

Temporal Models – Time and Uncertainty

A changing environment is modeled using a random variable for each aspect of the environment state at each time

The relation between these variables will tell how state evolves

Case study – Treating Diabetic Patient



Evidence
- Recent insulin doses
- Food intake
- Blood sugar measurement
- Other Physical signs

Task
- Assess current sugar state of the patient

Need
- This information will help doctor and patient makes decision about food intake and insulin dose

States and Observations

The process of change can be viewed as series of snapshots, describes the state of the environment at a particular time

Each snapshot or time slice has a set of variables some of which are observable and some are not.

Observations
- Recent insulin doses
- Food intake
- Blood sugar measurement
- Other Physical signs

State of the environment
- Assess current sugar state of the patient

Case Study – Umbrella and Rain

You are underground security and wants to know every day whether there is rain or not outside based on whether your director coming with umbrella or not.

Et : Observable evidence

Xt: Unobservable state

Et: Ut (Observable evidence is director coming with umbrella (Ut=1) or without umbrella (Ut=0))

Xt: Rt (Unobservable state, whether it is raining or not)

If(Ut=1) then Rt=1

If(Ut=0) then Rt=0

States and Observations

Time slice depends on the problem

for diabetic patient may be 1 hour and for security guard 1 day

$E_{1:3}$ – E_1, E_2, E_3

$X_{1:3}$ – X_1, X_2,X_3

Stationary Process

We need to specify the dependencies among set of state and evidence variables

Since cause usually precedes effect so we need add the variable in causal order

Set of variables are unbounded as it includes the state and evidence variables for every time slice

Creates two problems

- Unbounded number of conditional probability table
- Unbounded number of parents

Solution – Stationary Process

Stationary Process means if there is no change in environment then there is no change in variables

This leads fixed variables for an state in a particular environment

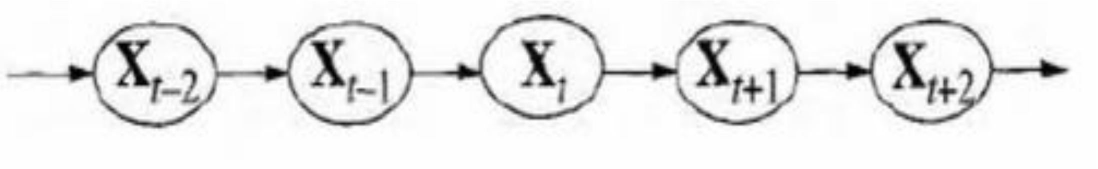
In umbrella environment problem the conditional probability that the umbrella appears P(Ut | Parents(Ut)) is same for all time slice

Markov Assumption

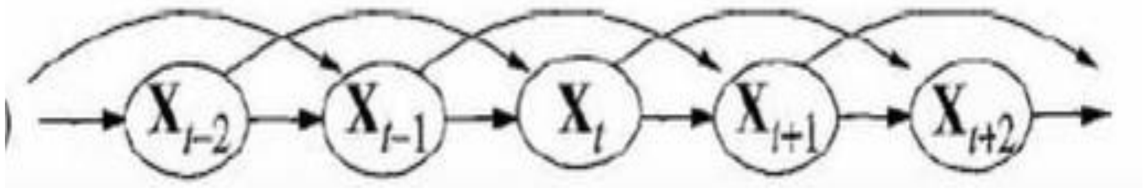
The current state only depends on only finite history of previous states

First order Markov process says current state depends on only on the previous state and not on earlier states

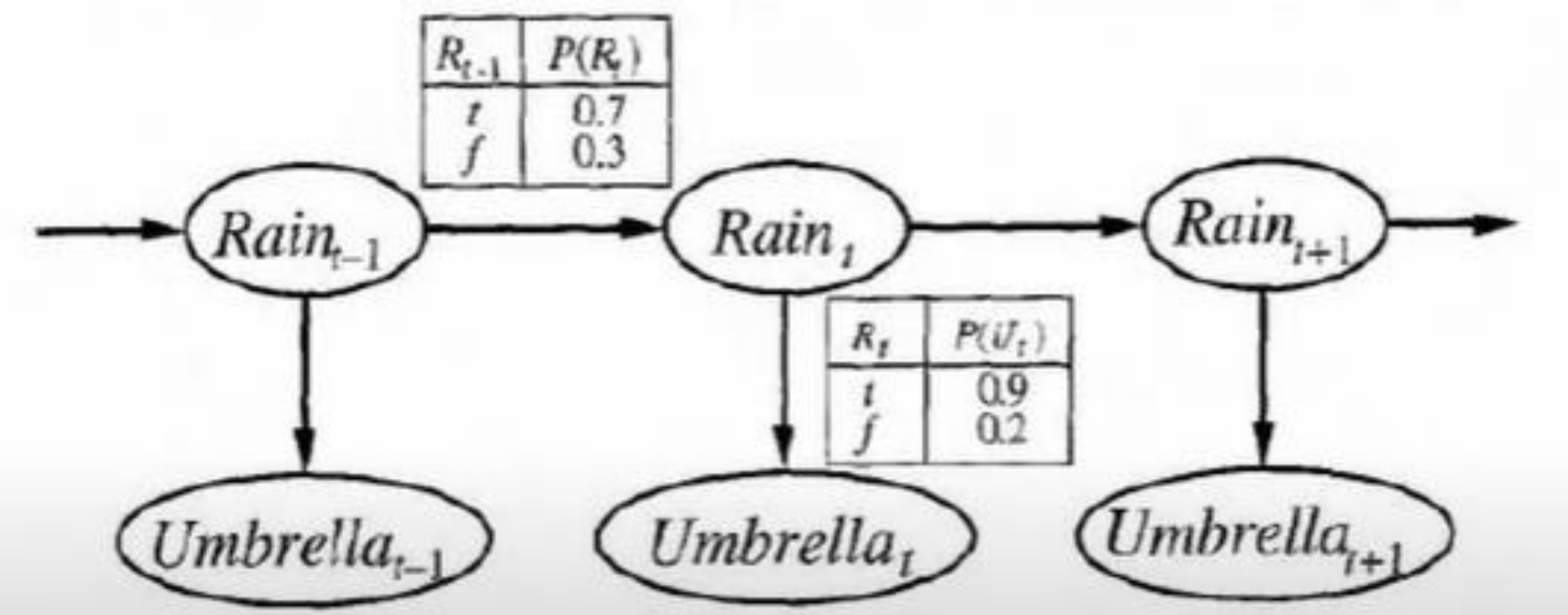
P(Xt | Xo:t-1) = P(Xt | Xt-1)



Second order Markov process says c [illegible] tates and not on earlier states



P(Xt | Xo:t-1) = P(Xt | Xt-1,Xt-2)



Bayesian Network – Umbrella Problem

Hidden Markov Model

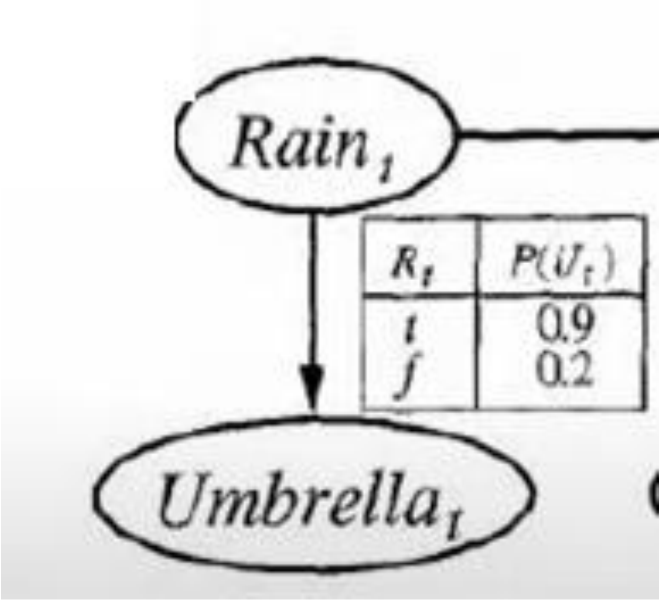
Hidden Markov model is statistical model in which the model states are hidden

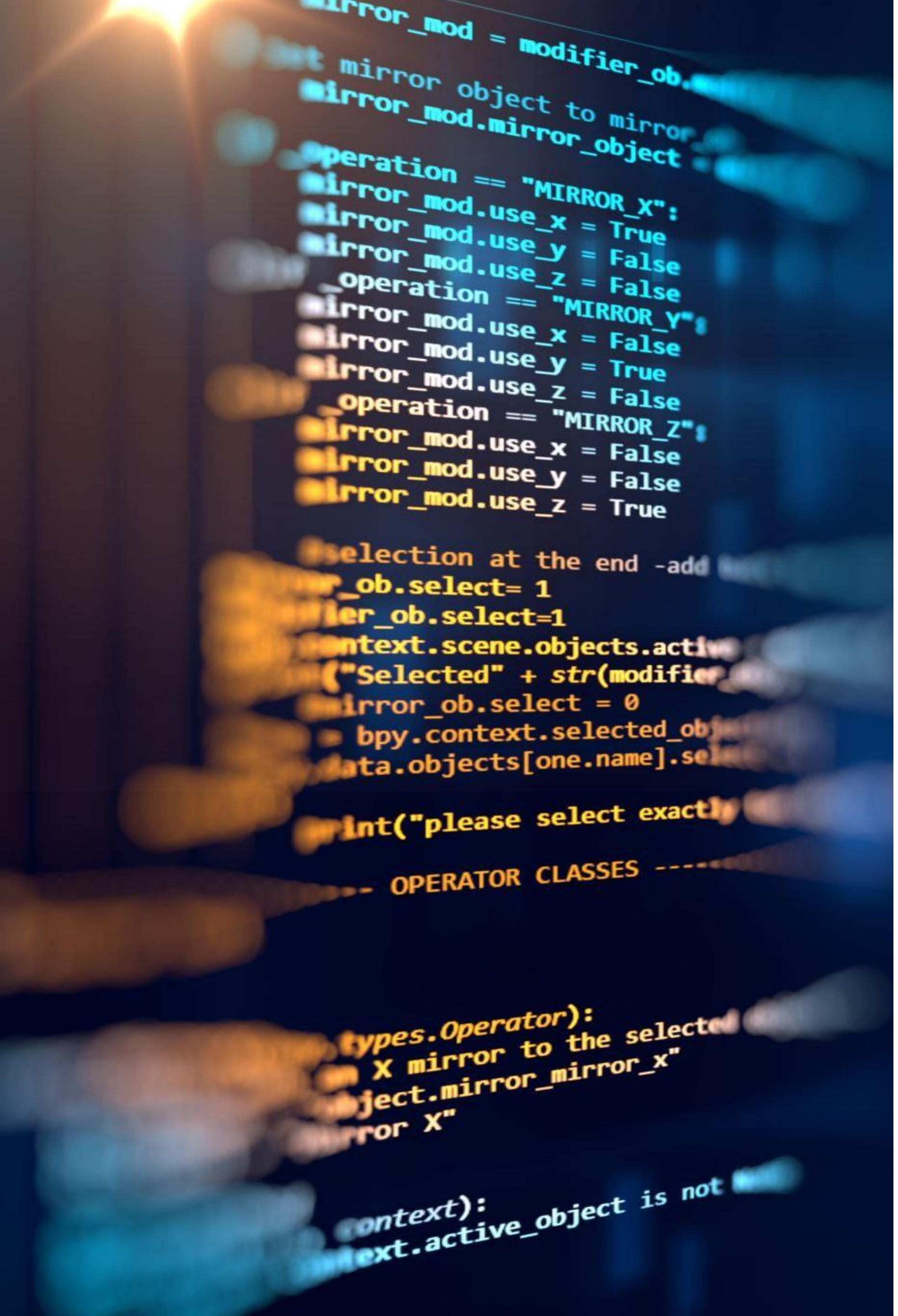
It is important to understand that the state of the model, and not the parameters of the model, are hidden.

A Markov model with fully known parameters is still called a HMM if states are hidden

While the model state may be hidden, the state-dependent output of the model is visible.

Information about the state of the model can be estimated from the probability distribution over possible output tokens





Application

The main usefulness of HMM is the recovery of a data sequence that is hidden by observing the output which is dependent on that hidden data sequence.

Computational finance

Speech recognition – Notably Apple’s Siri

Handwriting recognition

Time series analysis

Case Study

Two people, let’s call them Ram and Bhem, talk about food they like to eat.

Ram likes to eat pizza, pasta and pie. He tends to choose which to eat depending on his emotions.

Bhem has a rough understanding of the likelihood that Ram is happy or upset and his tendency to pick food based on those emotions.

Ram's food choice is the Markov process and Bhem knows the parameters but he does not know the state of Ram's emotions this is a hidden Markov model.

When they talk, Bhem can determine the probability of Ram being either happy or upset based on which of the three foods he chose to eat at a given moment.

Important Questions

Define	Define probability (slides2 & 3)
Define	Define joint probability (slide 4 & 5)
Mean by	What mean by mutual exclusive events (slide 6 & 7)
Explain	Explain all additive theorems (slide 8 & 11)
Explain	Explain independent events (slide 12) & solve problems (slide 13 & 15)

Important Questions

- Explain dependent events and conditional probability (slide 16 & 17)
- Explain Bayes Theorem (slide 18 & 21)
- Bayes Theorem – Problems (slide 22 & 23)
- What mean by local markov property (slide 24)
- Explain Bayesian Network (slide 25 & 29)
- Bayesian Network Problems (slide 30 & 31)
- What mean by temporal model explain with case study (slide 32 & 37)
- What mean by stationary process (Slide 38 and 39)
- Explain hidden markov process assumptions (slide 40 & 41)
- Expalin hidden markov model with applications (slide 42 & 44)

Thank you