Unit 5 (Contd ….)

Bayes Network

Why Bayes Network

- To represent the probabilistic relationship between two different classes
- To avoid dependences between value of attributes by joint conditional probability distribution
- In Naïve Bayes Classifier, attributes are conditionally independent

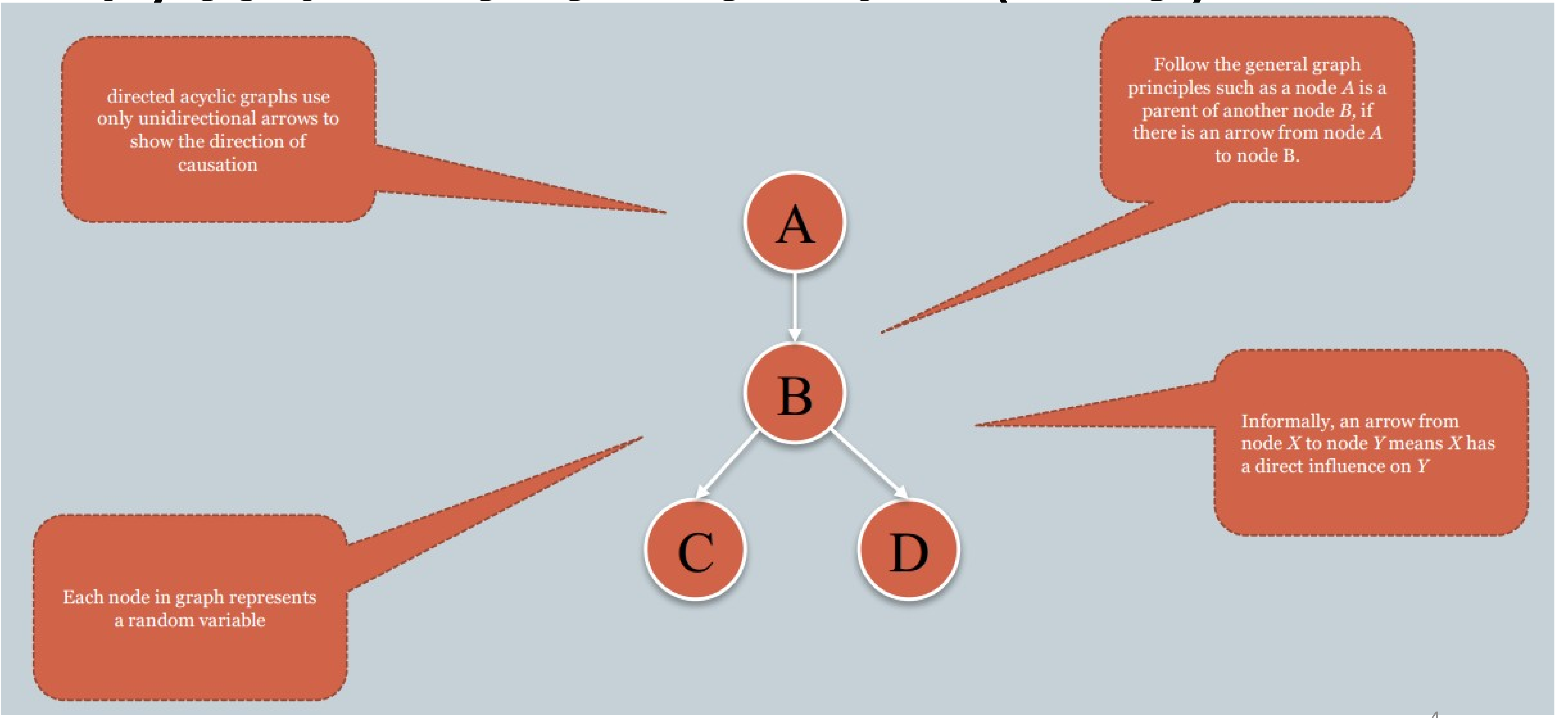
Bayesian Belief Network

- BN are also known as **Bayesian Network, Belief Networks** and **Probabilistic Networks**

- A BN is defined by two parts, **Directed Acyclic Graph (DAG)** and **Conditional Probability Tables (CPT)**

    Nodes Random Variables
    Arcs Indicates Probabilistic dependencies between nodes

Bayesian Belief Network (DAG)



Bayesian Belief Network

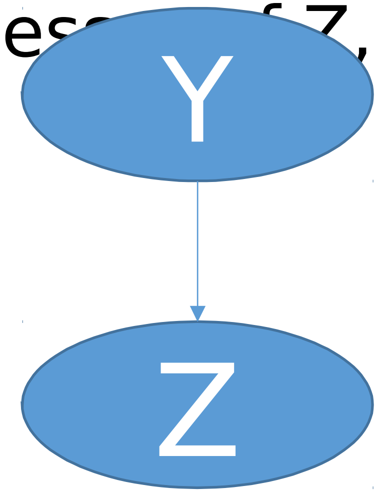
A BN is a directed graph with the following properties:

- **Nodes**: Set of Random Variables which may be discrete or continuous
- **Directed Links (Arcs)** : The real meaning od a link from node X to node Y is that X has a direct influence on Y
- Each node has a Conditional Probability Distribution $P(X_i|Parents(X_i))$ that quantifies the effects that the parent have on the node
- The graph has no directed cycles

Bayesian Belief Network

A BN is a directed graph with the following properties (contd…)

- If an arc is drawn from Y to Z, then Y is a parent or immediate predecessor of Z, and Z is a descendant of Y



- Each variable is conditionally independent of its non-descendants in the graph, given its parents

Bayes Belief Network

Incremental Network Construction:

1. ***Nodes:*** First determine the set of variables that are required to model the domain. Now order them, $\{X_1, X_2, \ldots, X_n\}$. Any order will work, but the resulting network will be more compact if the variables are ordered such that causes precede effects
2. ***Links*** : for $i = 1\ to\ n \quad do$:
   1. Choose, from $X_1, \ldots, X_{n-1}$, a minimal set of parents for X_i such that equation $\mathbf{P}(X_i|X_{i-1}, \ldots, X_1) = \mathbf{P}(X_i|Parents(X_i))$ is satisfied
   2. For each parent insert a link from the parent to X_i
   3. CPTs: Write down the Conditional Probability Table, $\mathbf{P}(X_i|Parents(X_i))$

Bayes Belief Network

- Conditional Independence:

Conditional Independence:

)$\mathbf{P}(X_1, X_2, \ldots, X_n) = \mathbf{P}(X_n | X_{n-1}, \ldots, X_1)\, \mathbf{P}(X_{n-1}, \ldots, X_1)$

=) $= \mathbf{P}(X_n | X_{n-1}, \ldots, X_1)\, \mathbf{P}(X_{n-1}, \ldots, X_1) \ldots \mathbf{P}(X_2 | X_1) \mathbf{P}(X_1)$

$= \sum_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i))$

A BN represents Conditional Independence

A BN represents Conditional Independence

$\mathbf{P}(X_i | X_{i-1}, \ldots, X_1) = \mathbf{P}(X_i | Parents(X_i))$

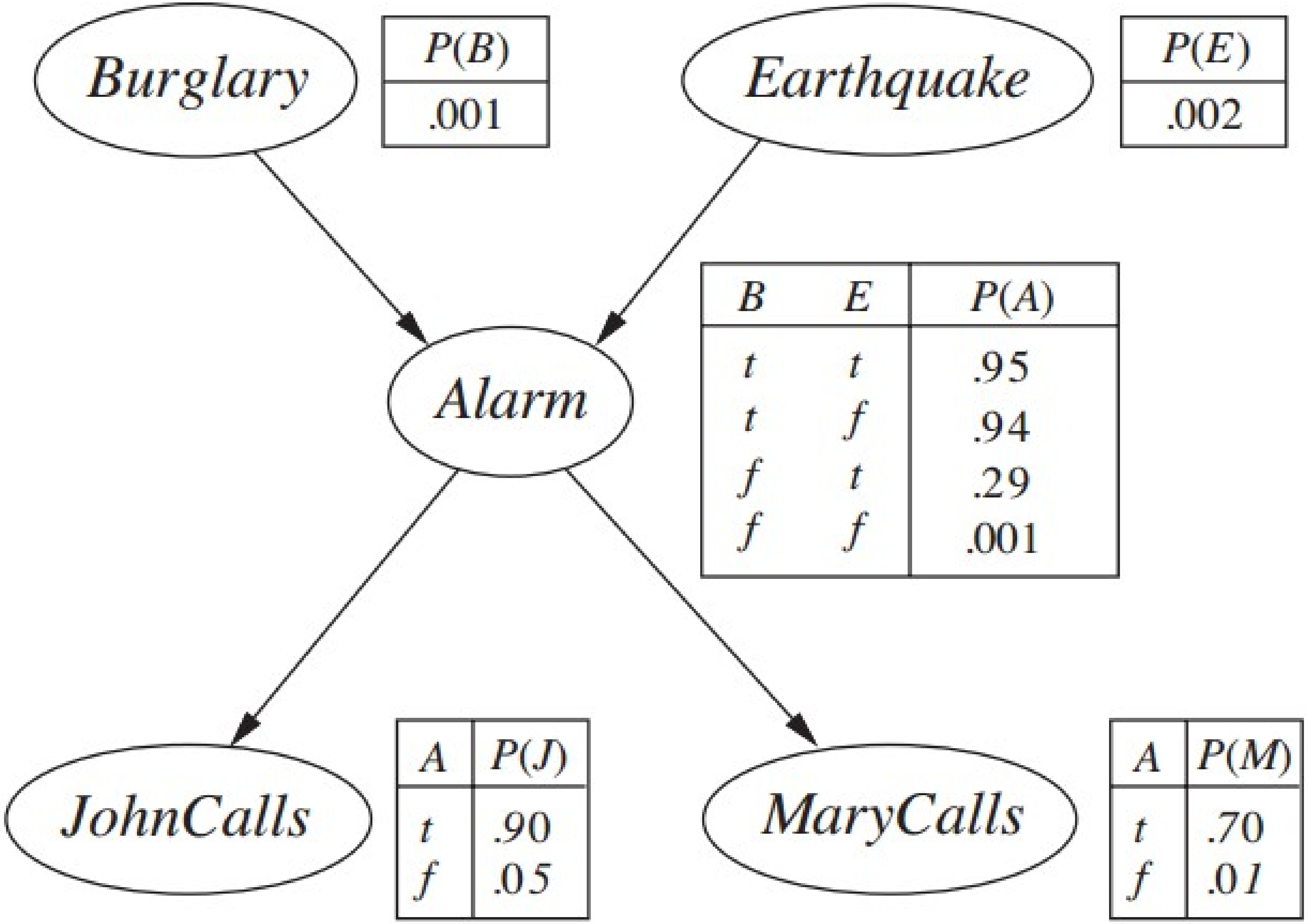
))

Bayes Belief Network

Example

- Burglar Alarm at Home
  - Fairly reliable at detecting a Burglary
  - Also Respond at times of Earthquake
- Two neighbours (John and Mary) on hearing Alarm calls you
  - John always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm and calls then too
  - Mary likes aloud music and sometimes misses the alarm altogether

Bayes Belief Network



$P(B)$
.001

$P(E)$
.002

B	E	$P(A)$
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	$P(J)$
t	.90
f	.05

A	$P(M)$
t	.70
f	.01

Bayes Belief Network

- Inference from Effect to cause; given Burglary, what is P(J|B)?

  $P(J \mid B) = ?$

  first calculate probability of Alarm ringing on burglary:

  $P(A \mid B) = P(B)P(E)P(BE) + P(B)P(E)P(BE)$

  $P(A \mid B) = 1*(0.998)*(0.94) + 1*(0.02)*(0.95)$

  $P(A \mid B) = 0.94$

  Now, Let us calculate P(J | B)

  $P(J \mid B) = P(A \mid B)*P(J) + P(\neg(A \mid B))*P(\neg J)$

  $P(J \mid B) = (0.94) * (0.9) + (0.06) * (0.05) = 0.85$

- Also calculate $P(M|B) = ?$

Bayes Belief Network

- It can readily handle incomplete data sets
- It allows one to learn about causal relationships
- It readily facilitate use of prior knowledge
- It Provide a natural representation for conditional independence
- It is more complex to construct the graph

Markov Network

- The canonical probabilistic model of temporal or sequential data is called Markov Model or Markov Network

- "**Future is independent of the past given the present**"

- They model a process that proceeds in steps (time, sequence, trials, etc.); like a series of probability trees

Markov Model

- Examples:

- What is the World at the end of this .... ?

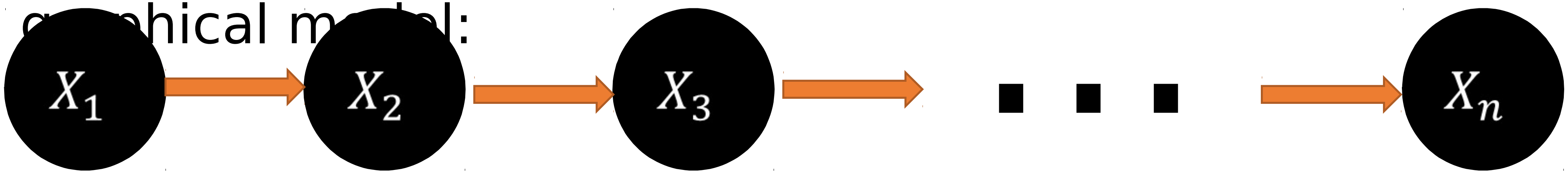
Markov Model

Definition:

Definition:

Discrete random variables $X_1, \ldots, X_n$ forms a Discrete time Markov Chain if their joint distribution respects the following graphical model:

Discrete random variables forms a Discrete time Markov Chain if their joint distribution respects the following graphical model:



Mathematically:

$$P(x_t|x_1, x_2, \ldots, x_{t-1}) = p(x_t|x_{t-1})$$

Mathematically:

What does it mean by resect the graph?

$$P(x_1, \ldots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_2)\ldots P(x_n|x_{n-1})$$

$$P(x_1, \ldots, x_n) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)\ldots P(x_n|x_1, \ldots, x_{n-1})$$

What does it mean by resect the graph?

Markov Model

Applications:

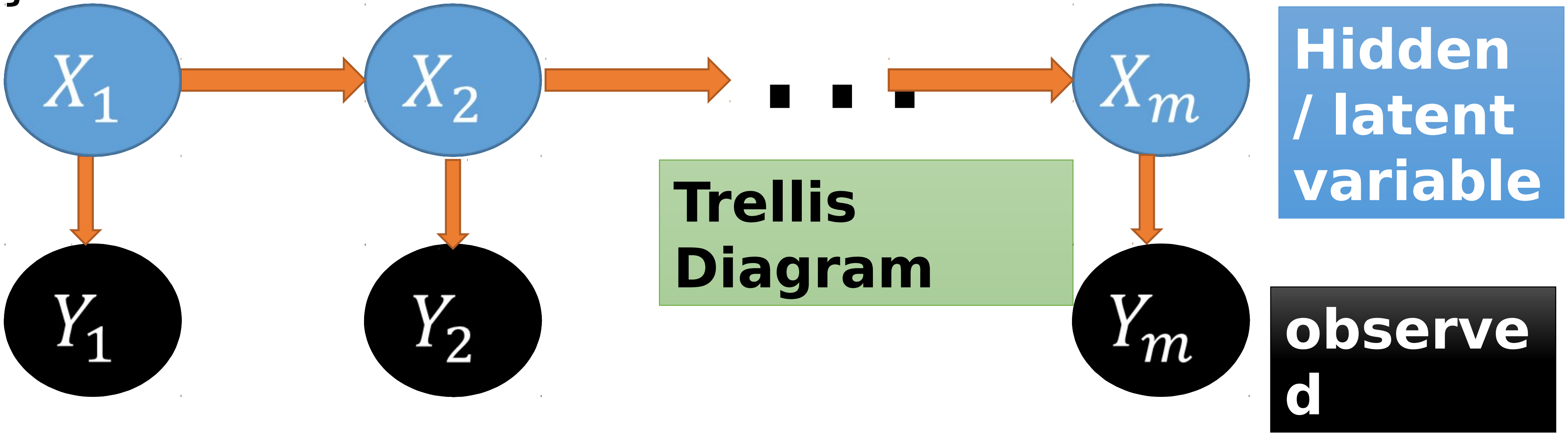
- Weather forecast
- Financial Analysis
- Language processing like Speech to Text Conversion or speech recognition
- Image tracking : Robot takes input as image and tries to find your next position by reading your coordinate
- Diagnosis : To self amend the internal state of machine. Ex. Mars Rovers use this to diagnosis if its wheels are broken and fix itself
- Bioinformatics : to determine how the patients are reacting toward the medication

Dynamic Bayes network

- A **Dynamic Bayesian Network** (DBN) is a Bayesian Network which relates variables to each other over adjacent time steps.
- This is often called a *Two-Timeslice* BN because it says that at any point in time T, the value of a variable can be calculated from the internal regressors and the immediate prior value (time T-1).
- DBNs are common in robotics, and have shown potential for a wide range of data mining applications. For example, they have been used in speech recognition, digital forensics, protein sequencing, and bioinformatics.

Hidden Markov Model(HMM)

- An HMM is a stochastic finite automation, where each state generates(emits) an observation.
- Let X_t = hidden states / variables and Y_t = observations and m = possible states then X_t ε {1, 2, … , m} and Y_t ε {1,2, … , m}



Hidden Markov Model

- The Joint Distribution of Trellis Graph respects following:

The Joint Distribution of Trellis Graph respects following:

$$p(X_1, \dots, X_m, Y_1, \dots, Y_m) = p(X_1)\mathrm{p}(Y_1|X_1)\prod_{k=2}^{m}\mathrm{p}(X_k|X_{k-1})\mathrm{p}(Y_k|X_k)$$

HMM

Instance

We can regard the weather as three states ：

state1 ： Rain

state2 ： Cloudy

state3 ： Sun



		Tomorrow		
		Rain	Cloudy	Sun
Today	Rain	0.4	0.3	0.3
	Cloudy	0.2	0.6	0.2
	Sun	0.1	0.1	0.8

We can obtain the transition matrix with long term observation

HMM

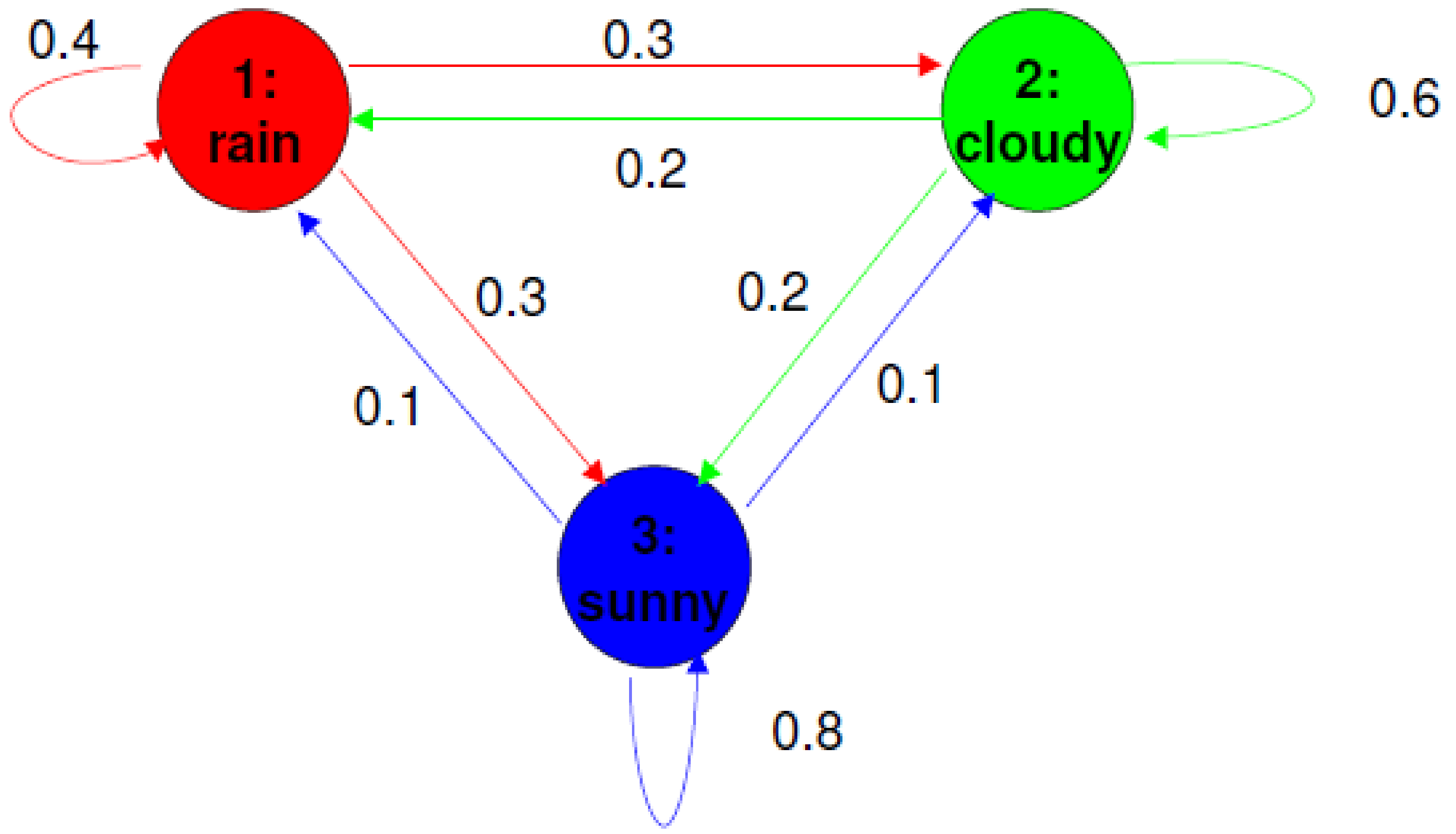


Fig: State Diagram

HMM

Parameters of HMM:

- State Space $\{1,2,\ldots,m\}$ and Observation Sequence
- **Transition probability:** $T(i,j) = P(X_{k+1} = j \mid X_k = i)\ \forall (i,j) \in \{1,2,\ldots,m\}$
- **State Transition Matrix:**

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ a_{N1} & a_{NN} & \cdots & a_{NN} \end{bmatrix}$$

$$a(i,j) = P(X_{k+1} = j \mid X_k = i)\ 1 \le i,\ j \le N$$

$$a(i,j) \ge 0$$

- **Initial State Transition Matrix / Distribution:** $\pi_i = P(X_1 = i),\ 1 \le i \le N$

HMM

- **Emission Probability:** $\varepsilon_i(y) = P(y|\, X_k = i)\ for\ i \in \{1, \dots, m\},\ y \in Y$

- **Emission Probability:**

Probability Distribution on Y

$\varepsilon_i(y) = P(\, Y_k = y \mid X_k = i)$

Joint Distributions in terms of above parameters:

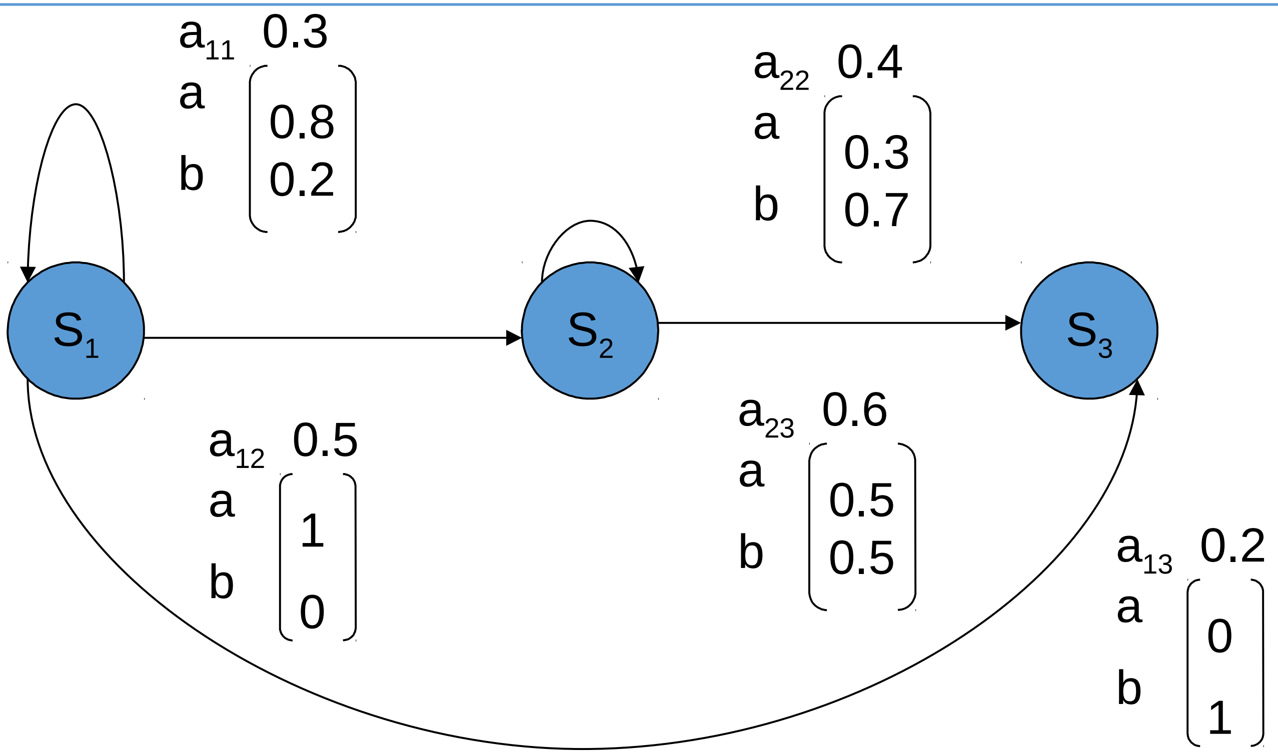
Joint Distributions in terms of above parameters:

$$p(X_1, \dots, X_m, Y_1, \dots, Y_m) = p(X_1)\mathrm{p}(Y_1|X_1) \prod_{k=2}^{m} \mathrm{p}(X_k|X_{k-1})\mathrm{p}(Y_k|X_k)$$

$$p(X_1, \dots, X_m, Y_1, \dots, Y_m) = \pi_i(\varepsilon_{X_1}(Y_1)) \prod_{k=2}^{m} \mathrm{T}(X_{k-1}, X_k)(\varepsilon_{X_k}(Y_k))$$

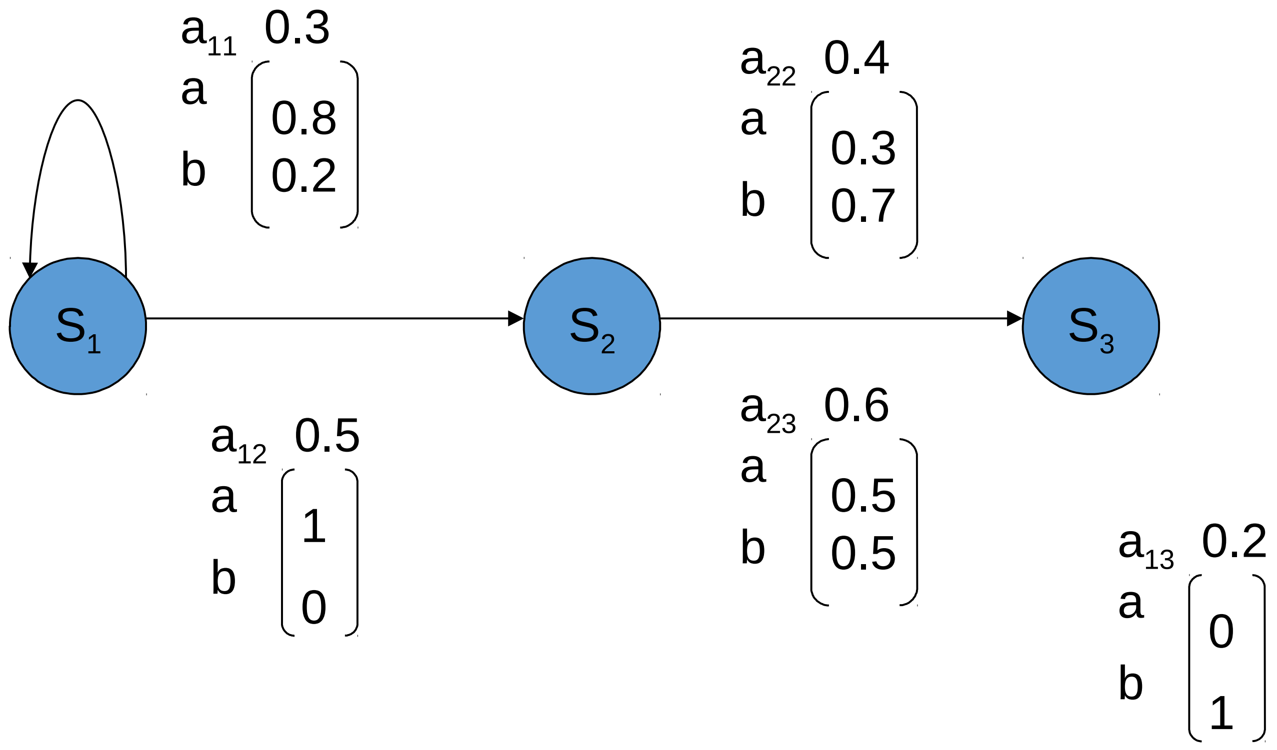
HMM

What's the probability of producing the sequence "aab" for this stochastic process?



HMM

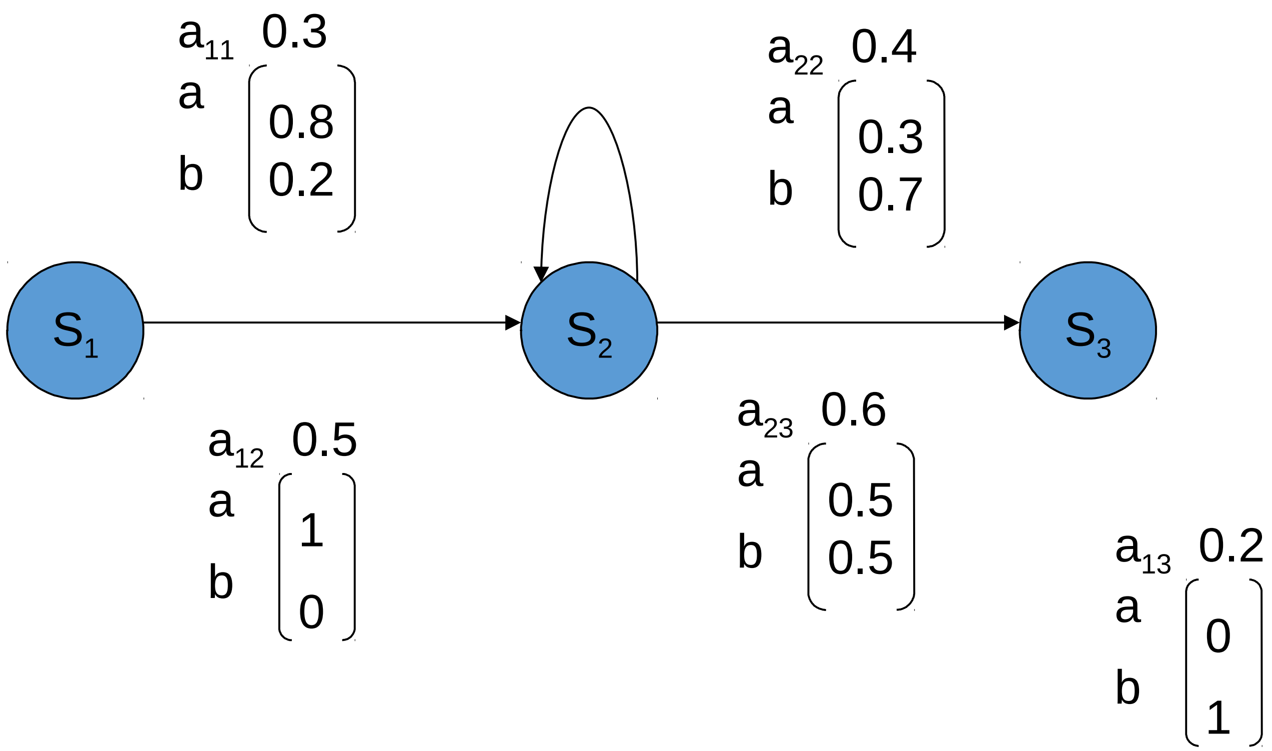
INSTANCE 1:



$S_1 \rightarrow S_1 \rightarrow S_2 \rightarrow S_3$ 0.3*0.8*0.5*1.0*0.6*0.5=0.036

HMM

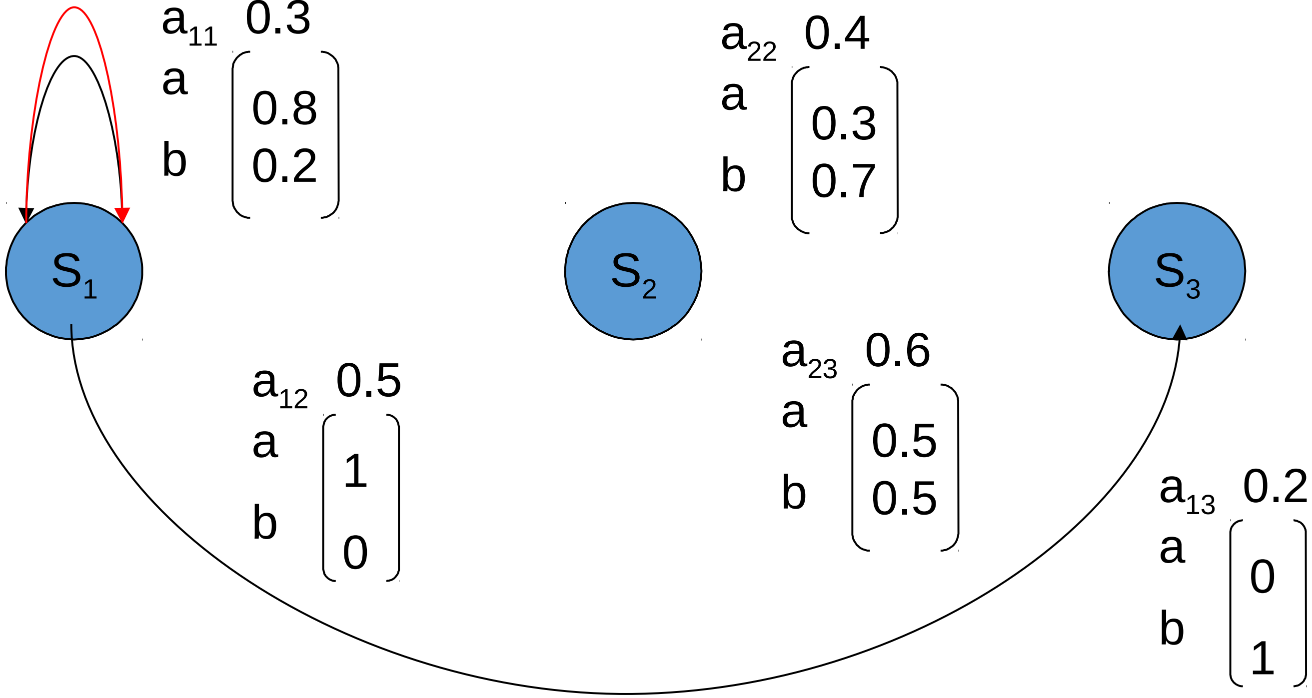
INSTANCE 2:



$S_1 \rightarrow S_2 \rightarrow S_2 \rightarrow S_3$ 0.5*1.0*0.4*0.3*0.6*0.5=0.018

HMM

INSTANCE 3



$S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_3$ 0.3*0.8*0.3*0.8*0.2*1.0=0.01152

Therefore, the total probability is:
0.036+0.018+0.01152=0.06552

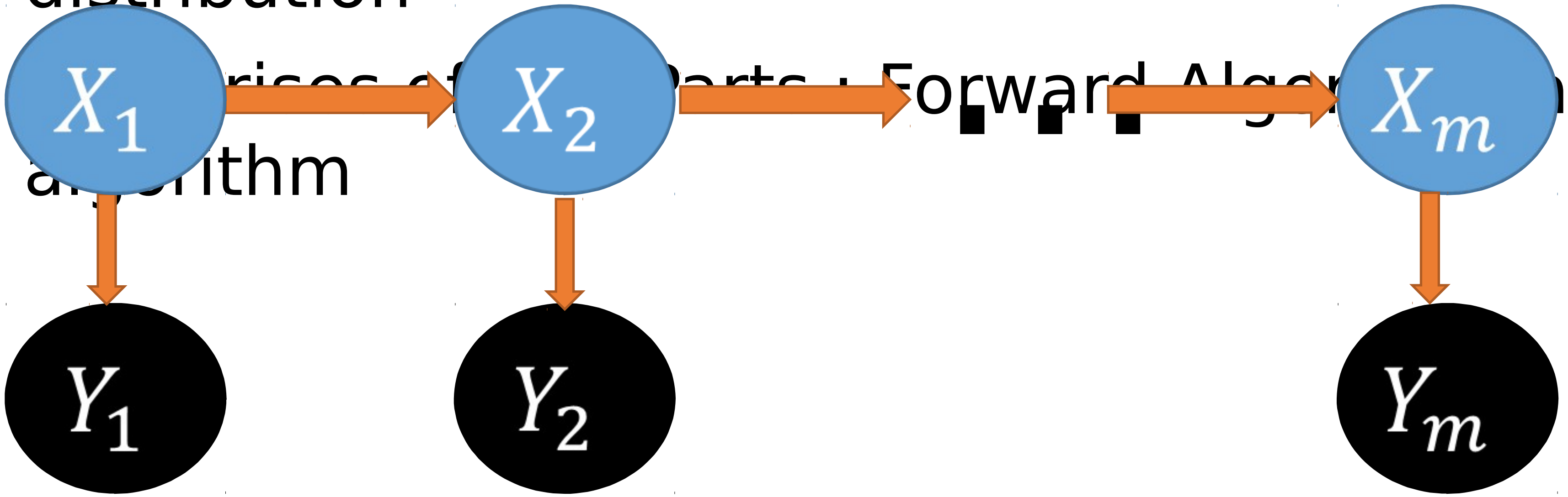
HMM

Forward Backward Algorithm:

used for Inference in HMM i.e Dynamic Programming first used by Richard Bellman

It assumes that we know emission probability and initial distribution

It comprises of two Parts : Forward Algorithm and Backward algorithm



HMM

- FB Algorithm’s goal is to compute:

FB Algorithm’s goal is to compute:

$$p(X_k|Y) \text{ where y} = (Y_1, Y_2, \dots, Y_m)$$

Forward Algorithm’s goal is to compute

Forward Algorithm’s goal is to compute

$$p(X_k|Y_{1:k}) \ \forall k = 1, \dots, m$$

Backward Algorithm’s goal is to compute

Backward Algorithm’s goal is to compute

$$p(Y_{k+1:m}|X_k) \ \forall k = 1, \dots, m$$