Hidden Markov Models

Module No. 6	**Hidden Markov Models**	**8 Hours**
Introduction, discrete Markov processes, hidden Markov models, three basic problems of HMMs evaluation problem, finding the state sequence, learning model parameters, continuous observations, the HMM with input, model selection in HMM.		

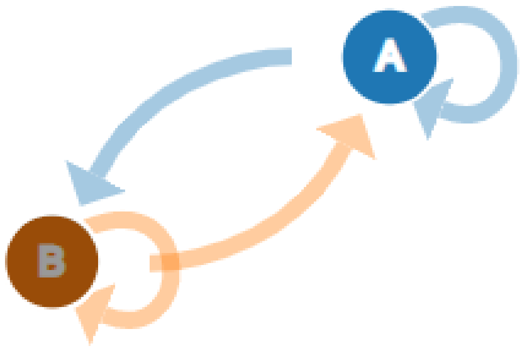
Probability Recap

- Conditional probability $$P(x|y) = \frac{P(x,y)}{P(y)}$$
- Product rule $$P(x,y) = P(x|y)P(y)$$
- Chain rule
$$\begin{aligned} P(X_1, X_2, \ldots X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots \\ &= \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1}) \end{aligned}$$
- X, Y independent if and only if: $$\forall x, y : P(x,y) = P(x)P(y)$$
- X and Y are conditionally independent given Z if and only if:

$$X \perp\!\!\!\perp Y|Z \qquad \forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

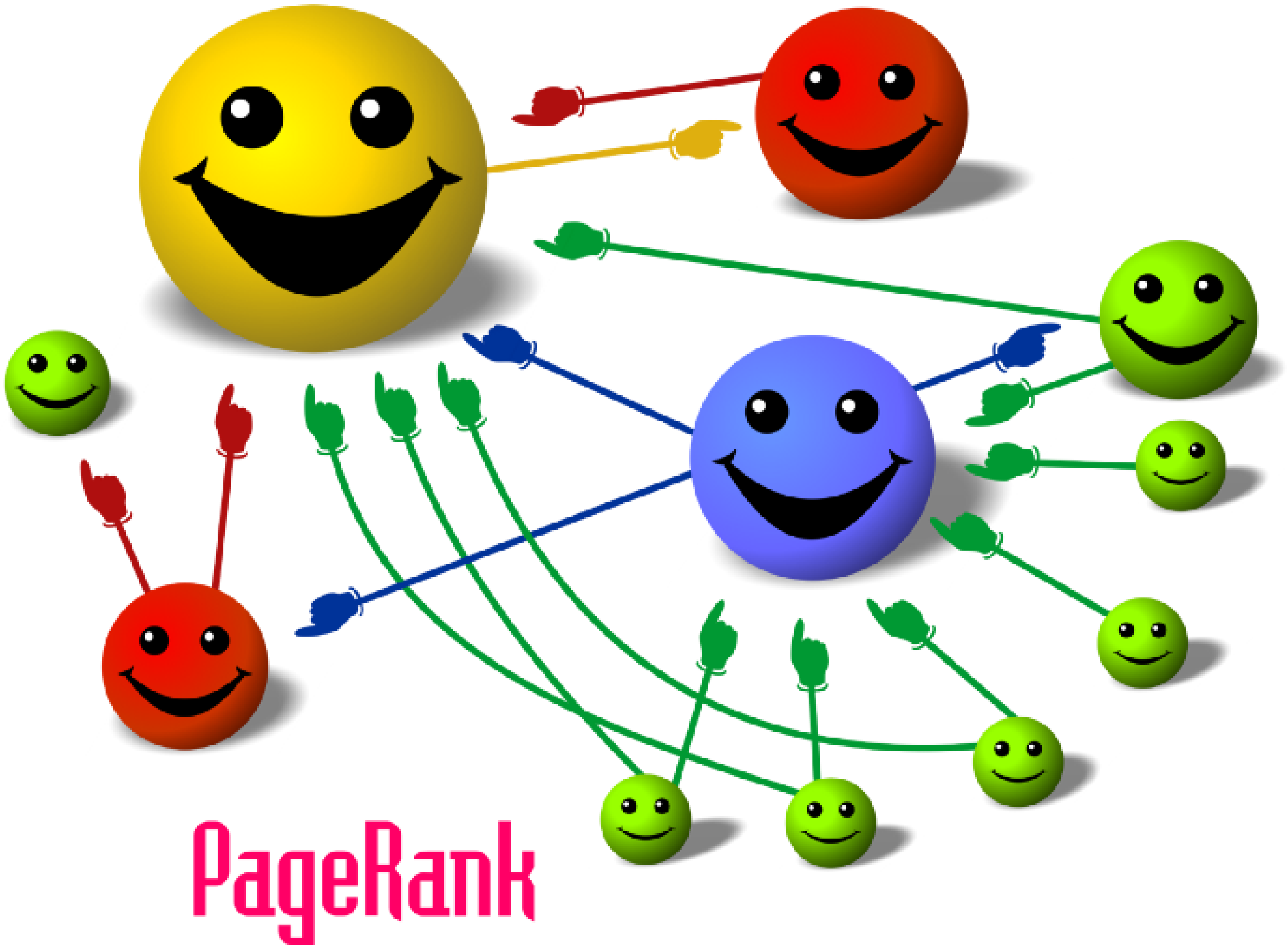
MARKOV Model

- Markov chains, named after Andrey Markov, are mathematical systems that hop from one "state" (a situation or set of values) to another.



	A	B
A	P(A|A): 0.76	P(B|A): 0.24
B	P(A|B): 0.50	P(B|B): 0.50

Markov Chain



What is a Markov Model?

- A Markov Model is a stochastic model which models temporal or sequential data, i.e., data that are ordered.

- It provides a way to model the dependencies of current information (e.g. weather) with previous information.

- It is composed of states, transition scheme between states, and emission of outputs (discrete or continuous).

- Several goals can be accomplished by using Markov models:
  - Learn statistics of sequential data.
  - Do prediction or estimation.
  - Recognize patterns.

Markov Chains

A Markov chain is simplest type of Markov model, where all states are observable and probabilities converge over time.

A Markov chain is **a mathematical process that transitions from one state to another within a finite number of possible states**. It is a collection of different states and probabilities of a variable, where its future condition or state is substantially dependent on its immediate previous state.

But there are other types of Markov Models. For instance, Hidden Markov Models are similar to Markov chains, but they have a few hidden states.

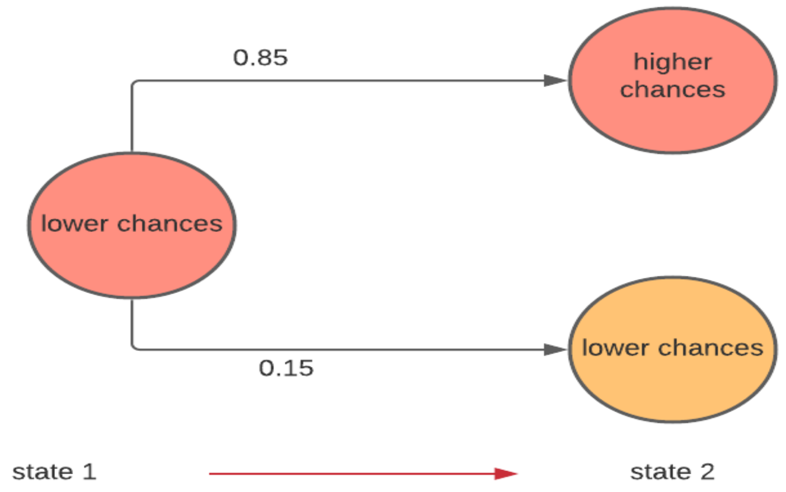
Markov Chains are one of the simple and very useful tools in order to model time-dependent, space-dependent stochastic processes.

A Markov chain is a **discrete-time** process for which the future behavior only depends on the present and not the past state. Whereas the Markov process is the **continuous-time** version of a Markov chain.

Many domains like finance (stock price movement), sales(sales quantity information), NLP algorithms (finite-state transducers, Hidden Markov Model for POS Tagging), weather forecasting, etc use the Markov chain to make their predictions easily and accurately.

Prediction Using Markov Chain

The Markov chain is a very powerful tool for making predictions for future value. Since it gives various useful insights, it becomes very necessary to know the transition probabilities, transition matrix, state-space, and trajectory to understand the insights.

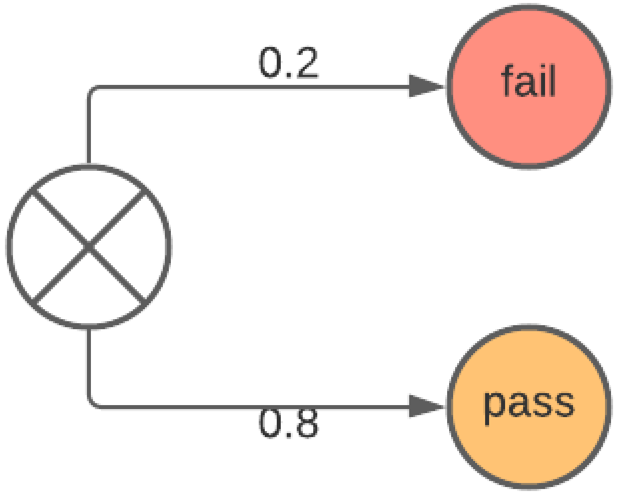


In what situation we use Markov chains?

They are stochastic processes for which the description of the present state fully captures all the information that could influence the future evolution of the process.

Predicting traffic flows, communications networks, genetic issues, and queues are examples where Markov chains can be used to model performance.

$$
\begin{array}{ccc}
 & \text{fail} & \text{pass} \\
\text{fail} & 0.75 & 0.25 \\
\text{pass} & 0.2 & 0.8
\end{array}
$$



$$\begin{Bmatrix} 0.2 & 0.8 \end{Bmatrix} \begin{Bmatrix} 0.75 & 0.25 \\ 0.2 & 0.8 \end{Bmatrix} = \begin{Bmatrix} 0.31 & 0.69 \end{Bmatrix}$$

Initial State X Transition Matrix = Prediction

V1 = V0 . T

to predict the second step the formula for prediction will be

V2 = V1 . T

V2 = (V0 . T) . T

V2 = V0 . T^2

Similarly, for the third step, the prediction will be

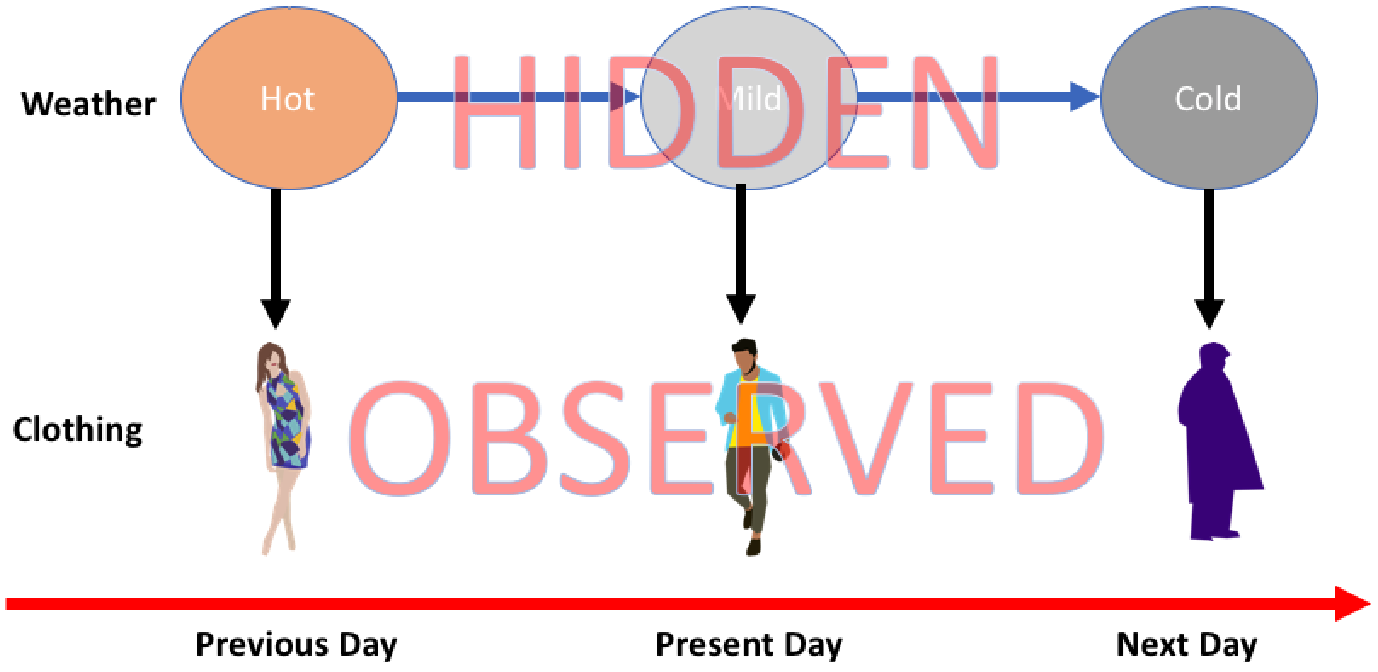
V3 = V2 . T = (V0 . T^2).T

V3 = V0 . T^3

Vn = Vn-1 . T = V0 .T^n

What is HMM ?

- Hidden Markov Models (HMMs) are a class of probabilistic graphical model that allow us to predict a sequence of unknown (hidden) variables from a set of observed variables.
- A simple example of an HMM is predicting the weather (hidden variable) based on the type of clothes that someone wears (observed).
-



- An HMM can be viewed as a Bayes Net unrolled through time with observations made at a sequence of time steps being used to predict the best sequence of hidden states.
- The hidden Markov model (HMM) is another type of Markov model where there are few states which are hidden.

Why Hidden, Markov Model?

- The reason it is called a Hidden Markov Model is because we are constructing an inference model based on the assumptions of a Markov process.
- The Markov process assumption is simply that the "*future is independent of the past given the present*". In other words, assuming we know our present state, we do not need any other historical information to predict the future state.

Figure 1: HMM hidden and observed states

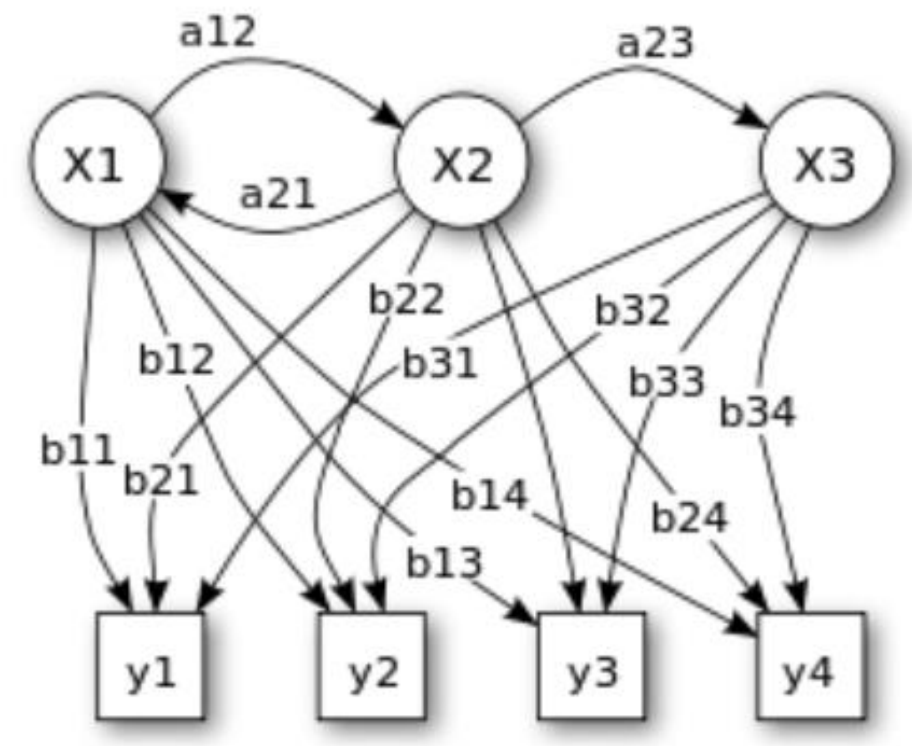


Figure 1. Probabilistic parameters of a hidden Markov model (example)

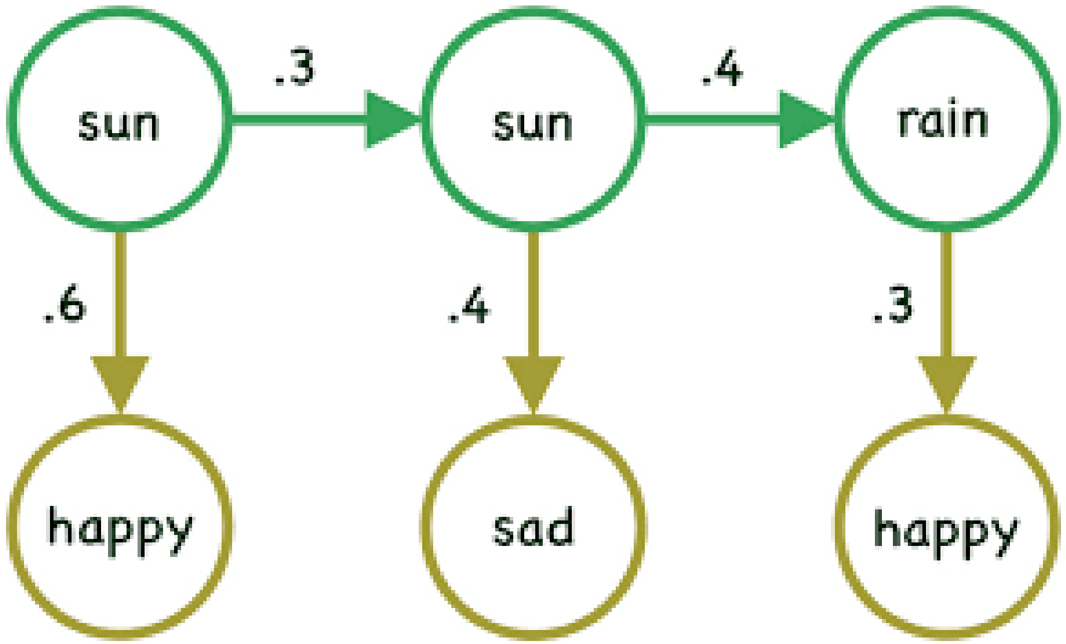
X — states

y — possible observations

a — state transition probabilities

b — output probabilities

- To make this point clear, let us consider the scenario below .
- The arrows represent transitions from a hidden state to another hidden state or from a hidden state to an observed variable.
- Notice that, true to the Markov assumption, <u>each state only depends on the previous state and not on any other prior states.</u>



HMM State Transitions

Intuition behind HMMs

- HMMs are probabilistic models. They allow us to compute the joint probability of a set of hidden states given a set of observed states. The hidden states are also referred to as latent states.
- Once we know the joint probability of a sequence of hidden states, we determine the best possible sequence i.e. the sequence with the highest probability and choose that sequence as the best sequence of hidden states.
- The ratio of hidden states to observed states is not necessarily 1 : 1 as is evidenced by **Figure 1** above(it may be many outputs)
- The key idea is that one or more observations allow us to make an inference about a sequence of hidden states.

HMM model consist of these basic parts:

- **hidden states**
- **observation symbols** (or states)
- transition from **initial state** to initial hidden state probability distribution
- transition to **terminal state** probability distribution (in most cases excluded from model because all probabilities equal to 1 in general use)
- **state transition** probability distribution
- **state emission** probability distribution

HMM State Transitions — Weather Example

- The example tables show a set of possible values that could be derived for the weather/clothing scenario.

Priors

Hot	0.6
Mild	0.3
Cold	0.1

Transitions

	Hot	**Mild**	**Cold**
Hot	0.6	0.3	0.1
Mild	0.4	0.3	0.2
Cold	0.1	0.4	0.5

Emissions

	Hot	**Mild**	**Cold**
Casual Wear	0.8	0.19	0.01
Semi Casual Wear	0.5	0.4	0.1
Winter apparel	0.01	0.2	0.79

Once this information is known, then the joint probability of the sequence, by the conditional probability chain rule and by Markov assumption, can be shown to be proportional to **P(Y)** below

The probability of observing a sequence

$$Y = y(0), y(1), \dots, y(L-1)$$

of length *L* is given by

$$P(Y) = \sum_{X} P(Y \mid X) P(X),$$

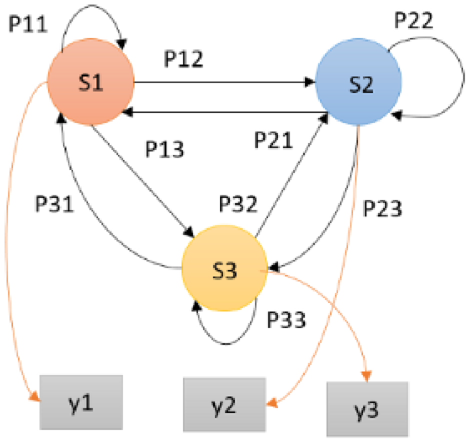
where the sum runs over all possible hidden-node sequences

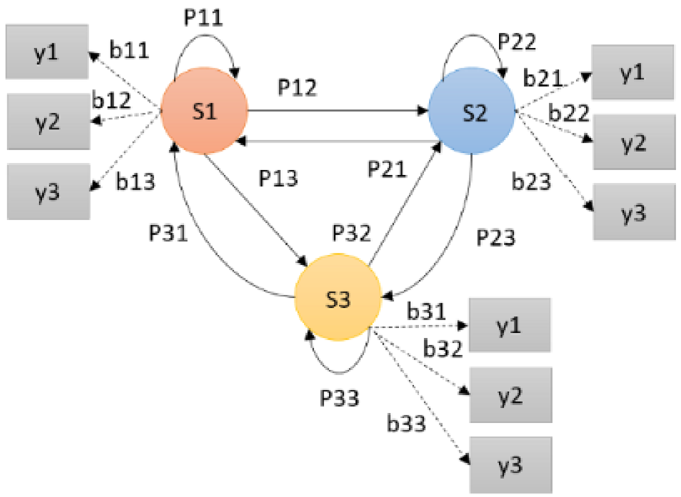
$$X = x(0), x(1), \dots, x(L-1).$$

Note that as the number of observed states and hidden states gets large, the computation gets more computationally intractable. If there are k possible values for each hidden sequence and we have a sequence length of n, then there are **n^k** total possible sequences that must be all scored and ranked in order to determine a winning candidate.

Hidden Markov Model (HMM)

- Hidden Markov Models (HMMs) are probabilistic models, it implies that the Markov Model underlying the data is hidden or unknown. More specifically, we only know observational data and not information about the states.

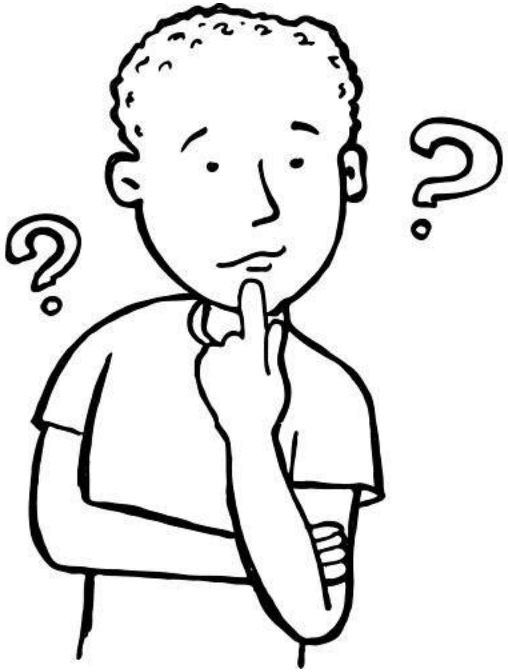




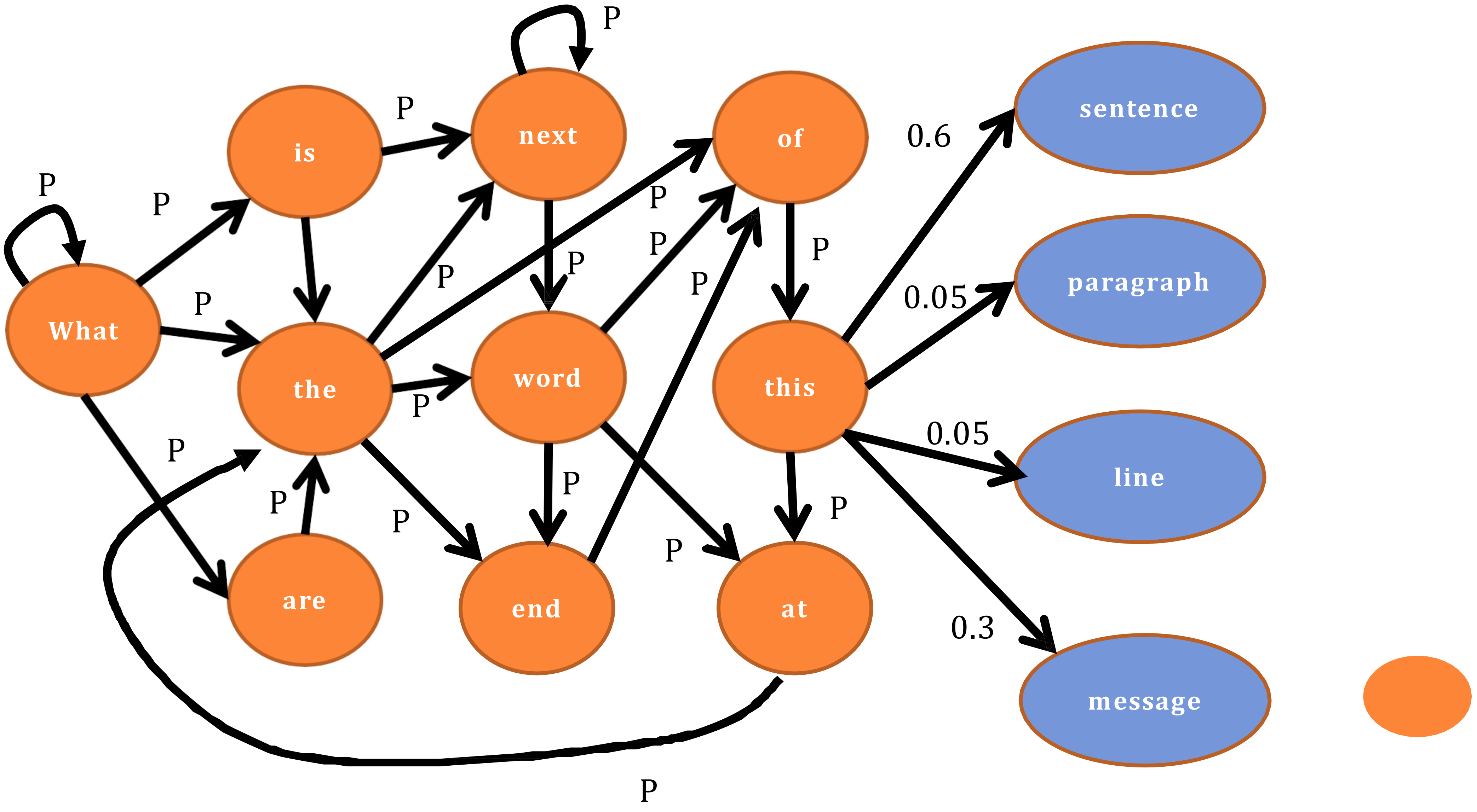
From Markov Chain (left) to Hidden Markov Model (right); where S=states, y=possible observations, P=state transition probabilities and b=observation probabilities

MOTIVATION

What is the word
at the end of this _____?



Markov Chain



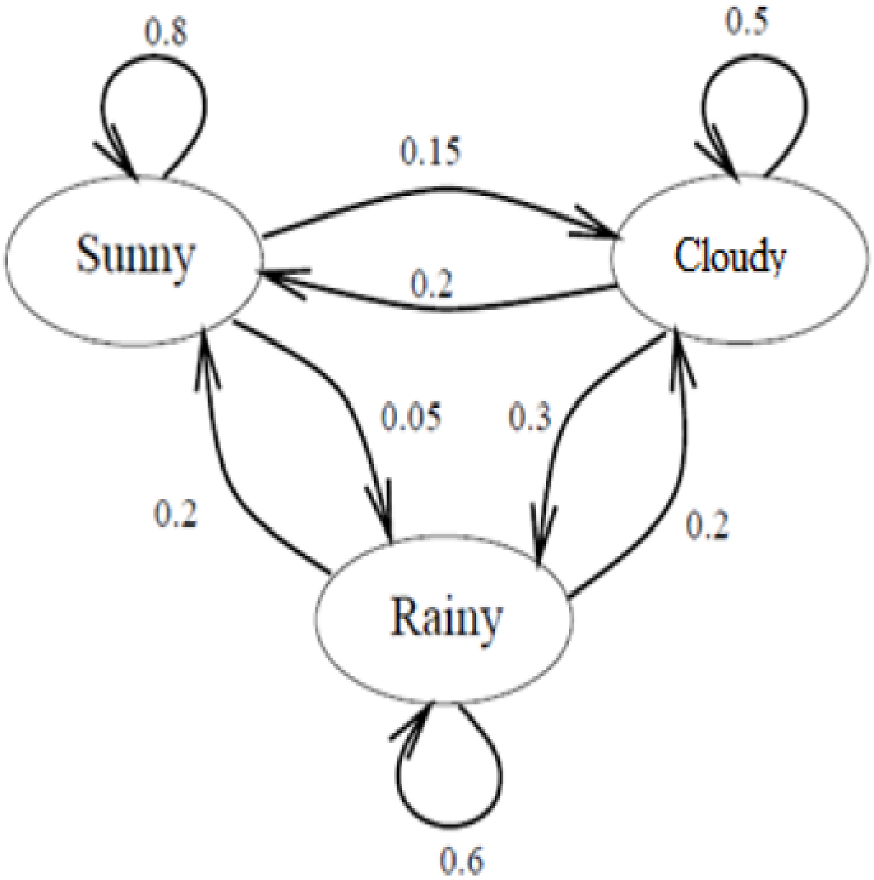
Markov Chain: Weather Example

- Design a Markov Chain to predict the weather of tomorrow using previous information of the past days.

- Our model has only 3 states: $S = \{S_1, S_2, S_3\}$, and the name of each state is $S_1 = Sunny$, $S_2 = Rainy, S_3 = Cloudy$.

- To establish the transition probabilities relationship between states we will need to collect data.

- Assume the data produces the following transition probabilities:



$$\left.\begin{aligned} P(Sunny|Sunny) &= 0.8 \\ P(Rainy|Sunny) &= 0.05 \\ P(Cloudy|Sunny) &= 0.15 \end{aligned}\right\} 1$$

$$\left.\begin{aligned} P(Sunny|Rainy) &= 0.2 \\ P(Rainy|Rainy) &= 0.6 \\ P(Cloudyy|Rainy) &= 0.2 \end{aligned}\right\} 1$$

$$\left.\begin{aligned} P(Sunny|Cloudy) &= 0.2 \\ P(Rainy|Cloudy) &= 0.3 \\ P(Cloudy|Cloudy) &= 0.5 \end{aligned}\right\} 1$$

- Let's say we have a sequence: Sunny, Rainy, Cloudy, Cloudy, Sunny, Sunny, Sunny, Rainy, ….; so, in a day we can be in any of the three states.

- We can use the following **state sequence** notation: q_1, q_2, q_3, q_4, q_5, … . ., where $q_i \epsilon \{Sunny, Rainy, Cloudy\}$.

- In order to compute the probability of tomorrow's weather we can use the Markov property:

$$P(q_1, \ldots, q_n) = \prod_{i=1}^{n} P(q_i | q_{i-1})$$

- Exercise 1: Given that today is Sunny, what's the probability that tomorrow is Sunny and the next day Rainy?

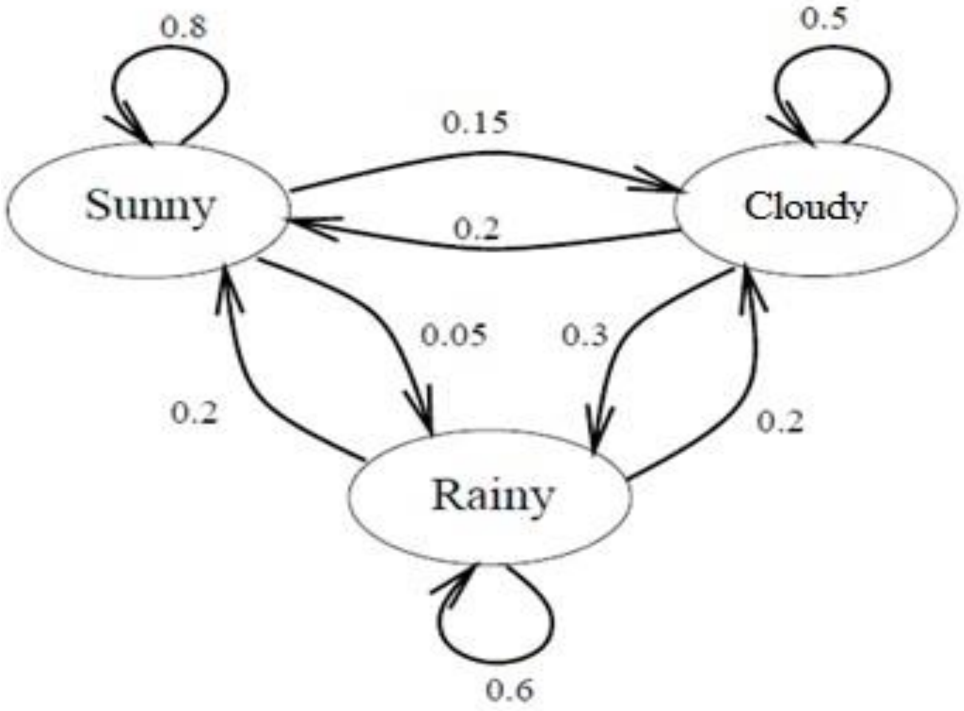
$P(q_2, q_3|q_1) = P(q_2|\ q_1)\ P(q_3|\ q_1, q_2)$

$= P(q_2|q_1)\ P(q_3|\ q_2)$

$= P(Sunny|Sunny)\ P(Rainy|Sunny)$

$= (0.8)(0.05)$

$= 0.04$

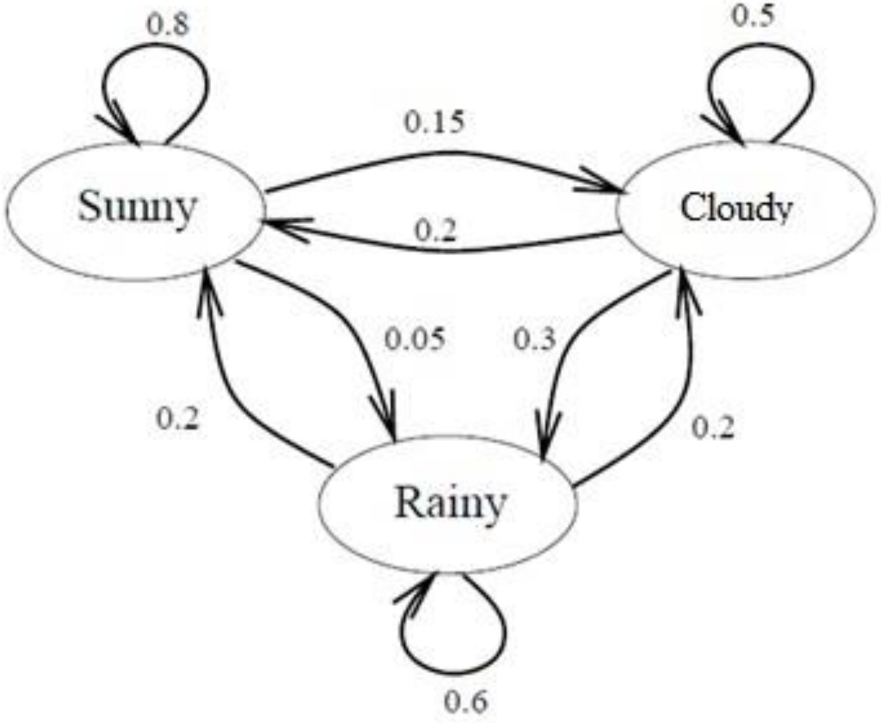


- Exercise 2: Assume that yesterday's weather was Rainy, and today is Cloudy, what is the probability that tomorrow will be Sunny?

$$P(q_3|q_1, q_2) = P(q_3|q_2)$$

$$= P(Sunny|Cloudy)$$

$$= 0.2$$



- Exercise 3 : Assume that today weather is Cloudy, and Yesterday was Rainy, what is the probability that tomorrow will be Sunny?
- P(Sunny| Rainy) = 0.2

- Exercise 4 : What is the probability of given sequence S,R,R,R,C,C

$$P(S) * P(R|S) * P(R|R) * P(R|R) * P(C|R) * P(C|C)$$

Problem 1

Consider the Markov chain with three states, $S = \{1, 2, 3\}$, that has the following transition matrix

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}.$$

a. Draw the state transition diagram for this chain.

b. If we know $P(X_1 = 1) = P(X_1 = 2) = \frac{1}{4}$, find $P(X_1 = 3, X_2 = 2, X_3 = 1)$.

Problem 1

Consider the Markov chain with three states, $S = \{1, 2, 3\}$, that has the following transition matrix

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}.$$

a. Draw the state transition diagram for this chain.
b. If we know $P(X_1 = 1) = P(X_1 = 2) = \frac{1}{4}$, find $P(X_1 = 3, X_2 = 2, X_3 = 1)$.

a. The state transition diagram is shown in Figure 11.6

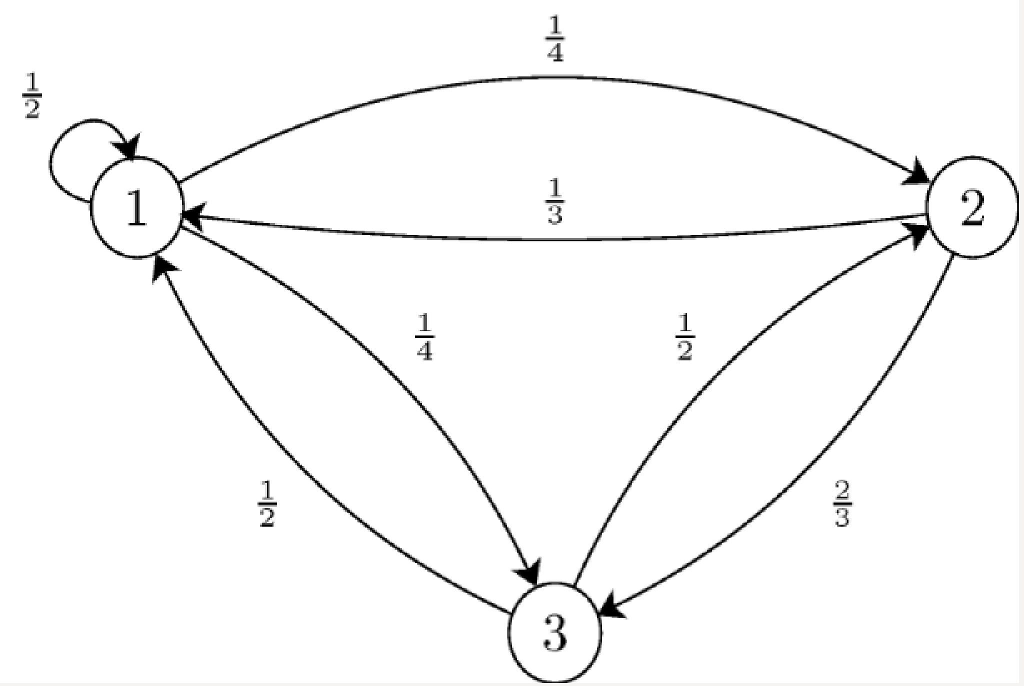


Figure 11.6 - A state transition diagram.

Use Product rule

b. First, we obtain

$$\begin{aligned} P(X_1 = 3) &= 1 - P(X_1 = 1) - P(X_1 = 2) \\ &= 1 - \frac{1}{4} - \frac{1}{4} \\ &= \frac{1}{2}. \end{aligned}$$

We can now write

$$\begin{aligned} P(X_1 = 3, X_2 = 2, X_3 = 1) &= P(X_1 = 3) \cdot p_{32} \cdot p_{21} \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} \\ &= \frac{1}{12}. \end{aligned}$$

Advantages of Markov Chain

- Markov chain is very easy to derive from a successional data
- We don’t need to dive deep into the mechanism of dynamic change.
- Markov chain is very insightful. It can tell the area of any process where we are lacking and further, we can make changes in accordance with improvement.
- Very low or modest computation requirements can easily be calculated by any size of the system.

Application of the Markov Chain

- Markov chains can be used for forecasting which can be any kind of forecasting like weather, temperature, sale, etc.
- This can be used for predicting customer behavior.
- As we know, it is good with sequential data so it can be merged with many NLP problem solutions like POS tagging.
- Brand loyalty and consumer behaviour can be analyzed.
- In the gaming field, various models can be developed in the game of chances.