Hidden Markov Models

Rafael E. Banchs (rbanchs@gps.tsc.upc.edu)

Session 1: Introduction to HMMs

Centre de Tecnologies i Aplicacions del Llenguatge i la Parla
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Definition of Markov Chains

- A sequence of random variables: $\boldsymbol{X} = X_1 , X_2 , X_3 \ldots X_k$
- Taking values in the same alphabet: $\boldsymbol{S} = \{s_1 , s_2 , s_3 \ldots s_m\}$

According to Bayes´ rule, the probability of the sequence is:

$$P(X_1 , X_2 , X_3 \ldots X_n) = \prod_{n=1}^{k} P(X_n \mid X_1 , X_2 \ldots X_{n-1})$$

Definition of Markov Chains

- A sequence of random variables: $\boldsymbol{X} = X_1 , X_2 , X_3 \ldots X_k$
- Taking values in the same alphabet: $\boldsymbol{S} = \{s_1 , s_2 , s_3 \ldots s_m\}$

According to Bayes´ rule, the probability of the sequence is:

$$P(X_1 , X_2 , X_3 \ldots X_n) = \prod_{n=1}^{k} P(X_n \mid X_1 , X_2 \ldots X_{n-1})$$

This random process is said to be a "Markov Chain" if it holds that $P(X_n \mid X_1 , X_2 \ldots X_{n-1}) = P(X_n \mid X_{n-1})$ for all values of *n*.

Basic definitions on Markov Chains

- A Markov Chain is *time invariant* or *homogeneous* if

$P(X_n = s_j \mid X_{n-1} = s_i) = P(X_1 = s_j \mid X_0 = s_i)$ for all n, j, i.

Basic definitions on Markov Chains

- A Markov Chain is *time invariant* or *homogeneous* if $P(X_n = s_j \mid X_{n-1} = s_i) = P(X_1 = s_j \mid X_0 = s_i)$ for all *n, j, i*.
- X_0 is referred to as the initial state of the Markov Chain, and its probability distribution π is a vector of m elements.

Basic definitions on Markov Chains

- A Markov Chain is *time invariant* or *homogeneous* if
  $P(X_n = s_j \mid X_{n-1} = s_i) = P(X_1 = s_j \mid X_0 = s_i)$ for all n, j, i.

- X_0 is referred to as the initial state of the Markov Chain, and its probability distribution π is a vector of m elements.

- The matrix **P** with elements $p_{ij} = P(X_n = s_j \mid X_{n-1} = s_i)$ is referred to as the transition matrix of the Markov Chain.

Basic definitions on Markov Chains

- A Markov Chain is *time invariant* or *homogeneous* if $P(X_n = s_j \mid X_{n-1} = s_i) = P(X_1 = s_j \mid X_0 = s_i)$ for all n, j, i.
- X_0 is referred to as the initial state of the Markov Chain, and its probability distribution π is a vector of m elements.
- The matrix **P** with elements $p_{ij} = P(X_n = s_j \mid X_{n-1} = s_i)$ is referred to as the transition matrix of the Markov Chain.
- It is an mxm stochastic matrix: has non-negative entries and its row sums are equal to one.

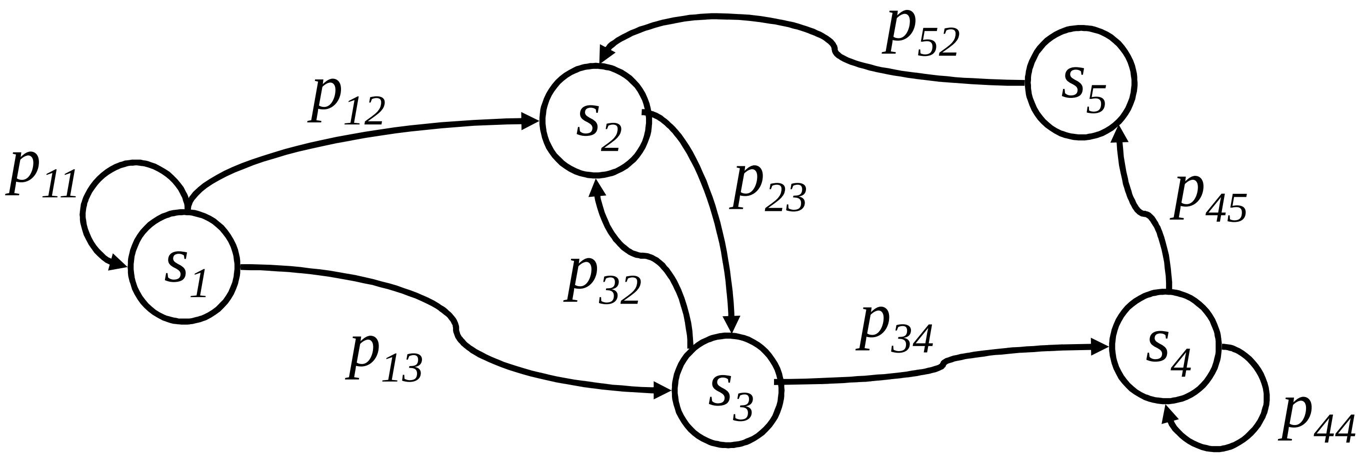
Markov Chains as finite state processes

- State alphabet: $\boldsymbol{S} = \{s_1 , s_2 , s_3 \ldots s_m\}$
- Initial probability distribution: $\boldsymbol{\pi}$
- Transition matrix: $\mathbf{P}$

Markov Chains as finite state processes

- State alphabet: $\boldsymbol{S} = \{s_1, s_2, s_3 \ldots s_m\}$
- Initial probability distribution: $\boldsymbol{\pi}$
- Transition matrix: **P**

Example for m = 5



More definitions on Markov Chains

- The n-step transition matrix is a matrix $\mathbf{P_n} = [p_{ij}(n)]$ with:
  $p_{ij}(n) = P(X_{h+n} = s_j \mid X_h = s_i)$

More definitions on Markov Chains

- The n-step transition matrix is a matrix $\mathbf{P_n} = [p_{ij}(n)]$ with:
  $p_{ij}(n) = P(X_{h+n} = s_j \mid X_h = s_i)$
- $\mathbf{P_n}$ can be computed by taking the *nth* power of **P**, so: $\mathbf{P_n} = \mathbf{P}^{\mathbf{n}}$

More definitions on Markov Chains

- The n-step transition matrix is a matrix $\mathbf{P_n} = [p_{ij}(n)]$ with:
  $p_{ij}(n) = P(X_{h+n} = s_j \mid X_h = s_i)$
- $\mathbf{P_n}$ can be computed by taking the *nth* power of $\mathbf{P}$, so: $\mathbf{P_n = P^n}$
- The probability distribution at the time step n, is given by:
  $\pi_n = \pi \, \mathbf{P_n}$

More definitions on Markov Chains

- The n-step transition matrix is a matrix $\mathbf{P_n} = [p_{ij}(n)]$ with:
  $p_{ij}(n) = P(X_{h+n} = s_j \mid X_h = s_i)$
- $\mathbf{P_n}$ can be computed by taking the *nth* power of $\mathbf{P}$, so: $\mathbf{P_n = P^n}$
- The probability distribution at the time step *n*, is given by:
  $\pi_n = \pi\ \mathbf{P_n}$
- And more general: $\mathbf{P_{k+n} = P_k\ P_n}$ and $\pi_{k+n} = \pi_k\ \mathbf{P_n}$

Exercise 1

Let us consider the following Markov Chain:

- State alphabet: $\boldsymbol{S} = \{s_1, s_2, s_3\}$
- Initial probability distribution: $\pi = [1, 0, 0]$
- Transition matrix: $\mathbf{P} = \begin{pmatrix} 0.2 & 0.8 & 0.0 \\ 0.0 & 0.6 & 0.4 \\ 0.5 & 0.2 & 0.3 \end{pmatrix}$

Exercise 1

1.- Compute the probability of being in state s_2 after three steps.

  1.a.- By considering all possible paths.

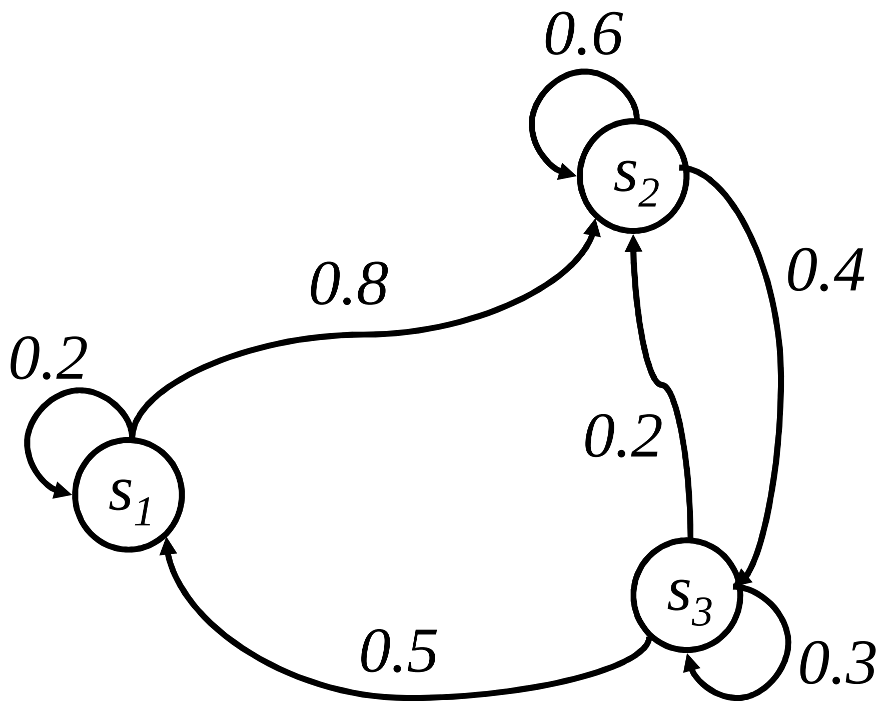
  1.b.- By applying Markov Chain properties.

2.- Compute the probability distribution of states after a large number of steps.

Exercise 1 (solution)

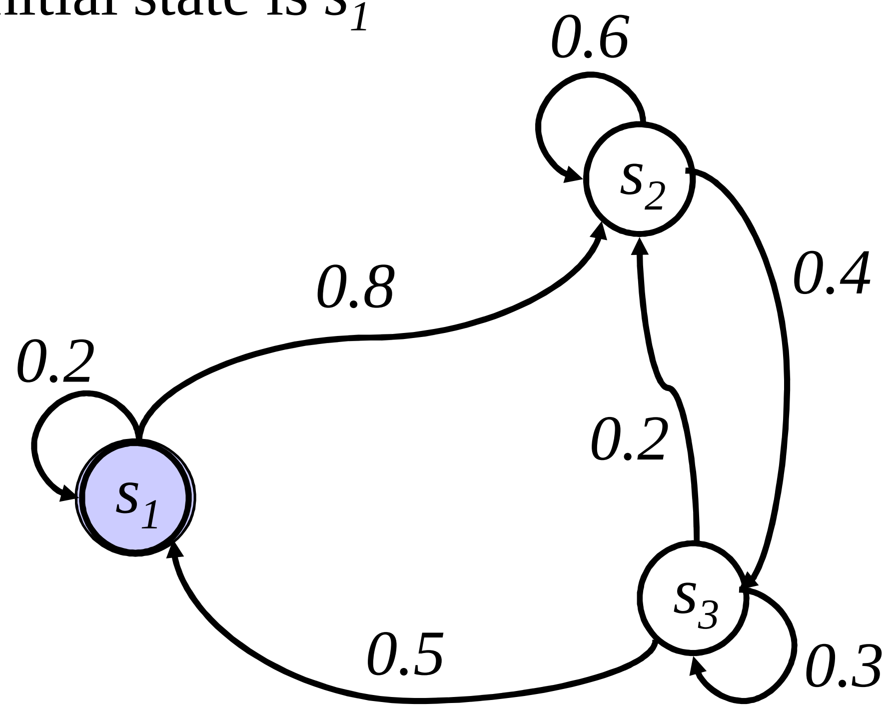
$$\pi = [1, 0, 0]$$

$$\mathbf{P} = \begin{pmatrix} 0.2 & 0.8 & 0.0 \\ 0.0 & 0.6 & 0.4 \\ 0.5 & 0.2 & 0.3 \end{pmatrix}$$



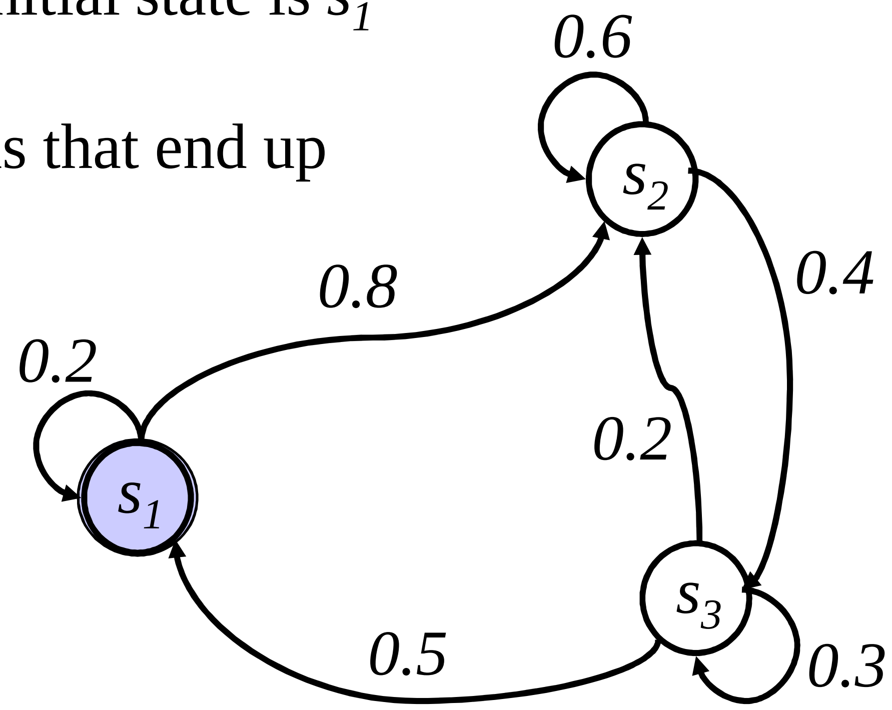
Exercise 1 (solution)

- Since $\pi = [1, 0, 0]$ the initial state is s_1



Exercise 1 (solution)

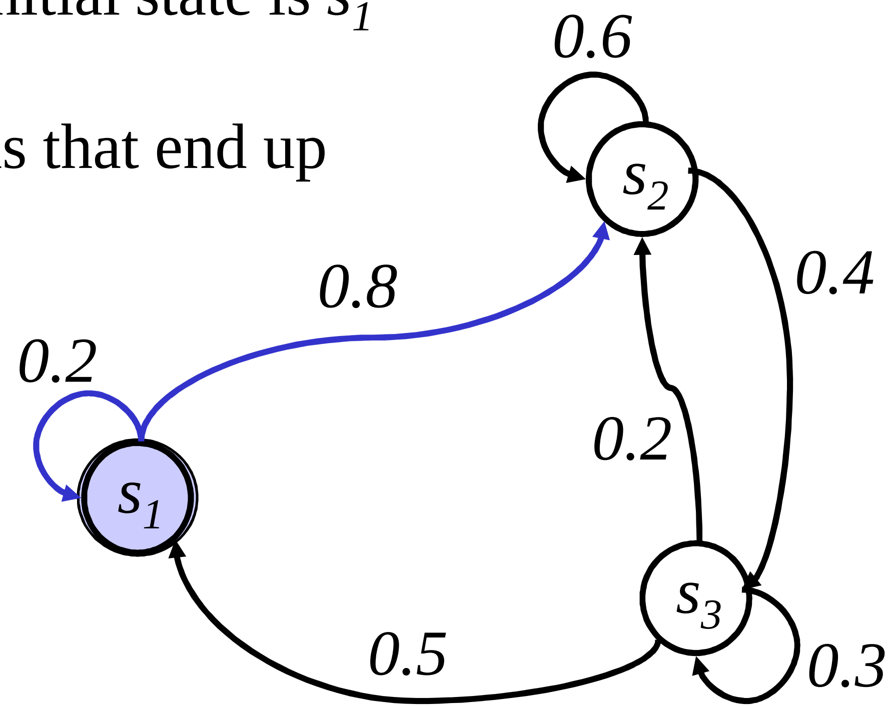
- Since $\pi = [1, 0, 0]$ the initial state is s_1
- There are only four paths that end up in s_2 after three steps.



Exercise 1 (solution)

- Since $\pi = [1, 0, 0]$ the initial state is s_1

- There are only four paths that end up in s_2 after three steps.



Path$_1$:

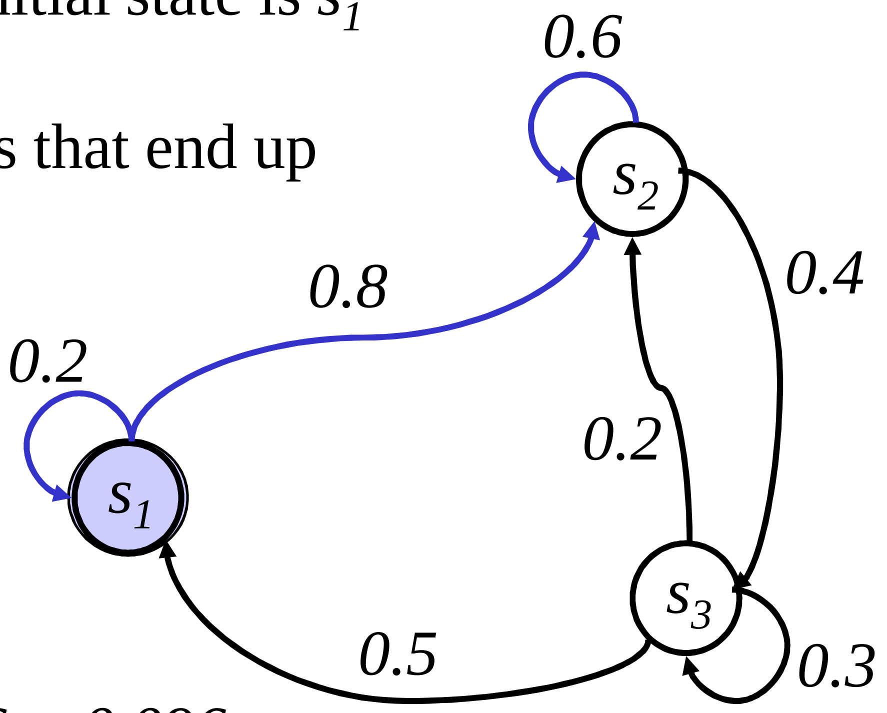
$s_1 \rightarrow s_1 \rightarrow s_1 \rightarrow s_2$

$0.2 \quad 0.2 \quad 0.8$

$p(path_1) = 0.2 \times 0.2 \times 0.8 = 0.032$

Exercise 1 (solution)

- Since $\pi = [1, 0, 0]$ the initial state is s_1
- There are only four paths that end up in s_2 after three steps.



Path$_2$:

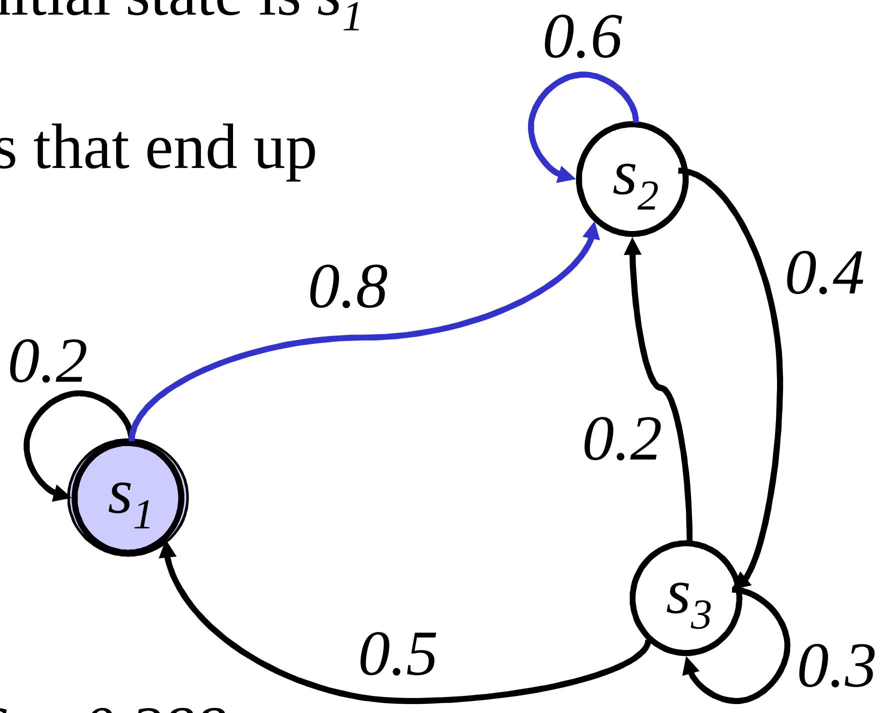
$s_1 \rightarrow s_1 \rightarrow s_2 \rightarrow s_2$

$0.2 \quad 0.8 \quad 0.6$

$p(path_2) = 0.2 \times 0.8 \times 0.6 = 0.096$

Exercise 1 (solution)

- Since $\pi = [1, 0, 0]$ the initial state is s_1
- There are only four paths that end up in s_2 after three steps.



Path$_3$:

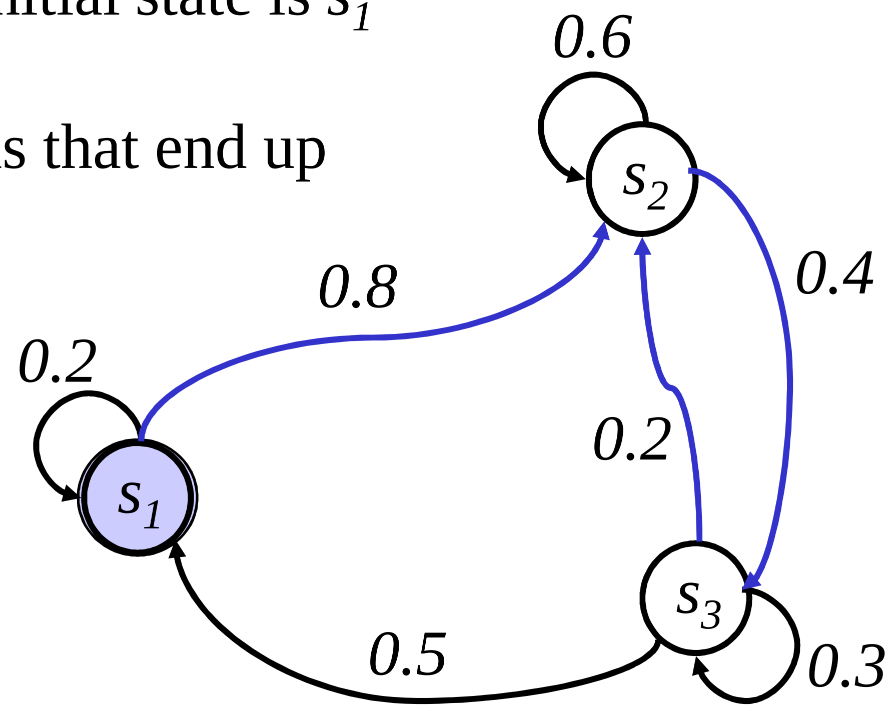
$s_1 \rightarrow s_2 \rightarrow s_2 \rightarrow s_2$

0.8 0.6 0.6

$p(path_3) = 0.8 \times 0.6 \times 0.6 = 0.288$

Exercise 1 (solution)

- Since $\pi = [1, 0, 0]$ the initial state is s_1
- There are only four paths that end up in s_2 after three steps.



Path$_4$:

$s_1 \rightarrow s_2 \rightarrow s_3 \rightarrow s_2$

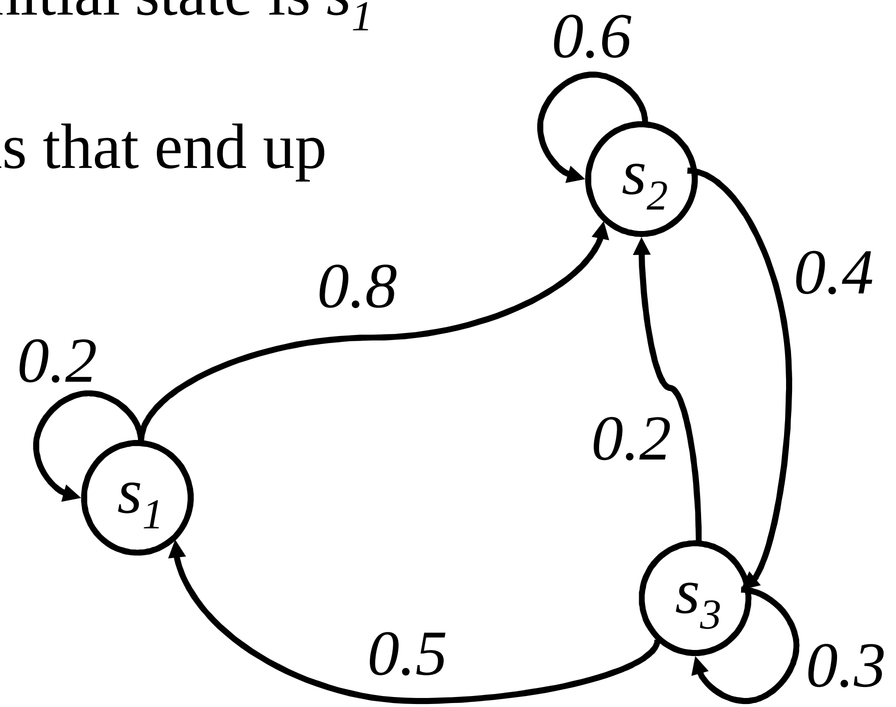
$0.8 \quad 0.4 \quad 0.2$

$p(path_4) = 0.8 \times 0.4 \times 0.2 = 0.064$

Exercise 1 (solution)

- Since $\pi = [1, 0, 0]$ the initial state is s_1
- There are only four paths that end up in s_2 after three steps.
- The probability of being in state s_2 after three steps is:

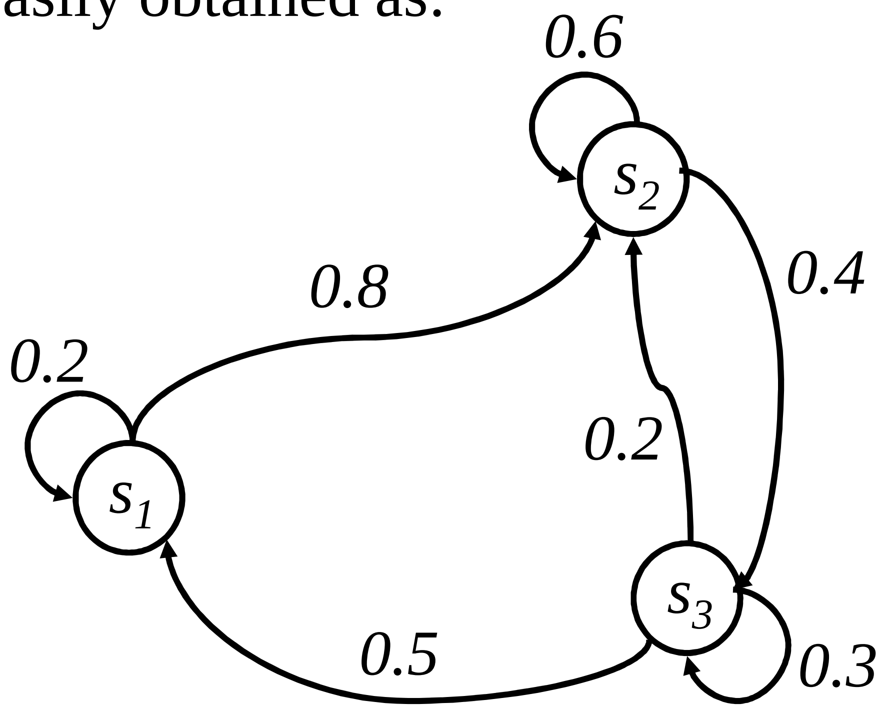
$$\sum_{i=1}^{4} p(path_i) = 0.480$$



Exercise 1 (solution)

- The same result can be easily obtained as:

$$\pi_3 = \pi \ \mathbf{P}_3 = \pi \ \mathbf{P} \, \mathbf{P} \, \mathbf{P}$$



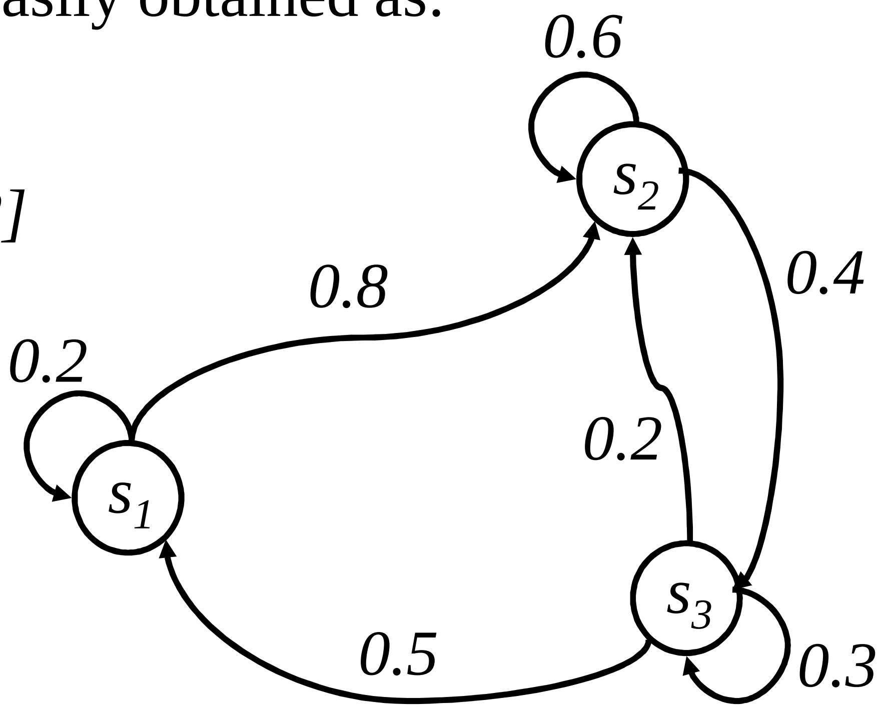
Exercise 1 (solution)

- The same result can be easily obtained as:

$$\pi_3 = \pi\ \mathbf{P}_3 = \pi\ \mathbf{P}\ \mathbf{P}\ \mathbf{P}$$

$$\pi_3 = [0.168\ \ 0.480\ \ 0.352]$$

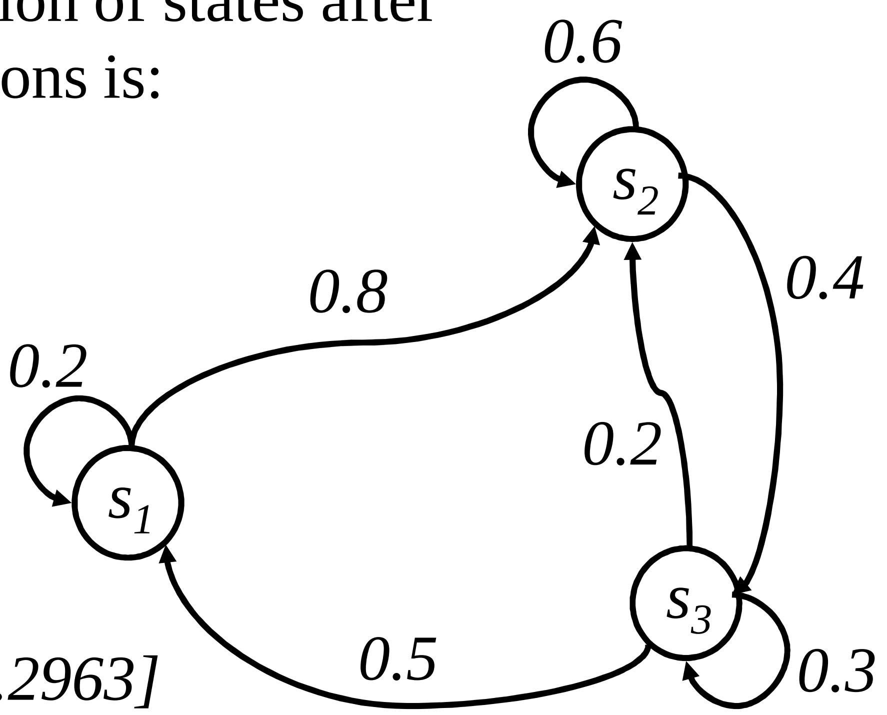
Probability of being in state s_2 after three steps



Exercise 1 (solution)

- The probability distribution of states after a large number of iterations is:

$\pi_n = \pi\ \mathbf{P_n}$ with n large



- So for $n = 50$;

$\pi_{50} = [0.1852\ \ 0.5185\ \ 0.2963]$

Hidden Markov Models (HMM)

- In some practical applications, the state sequence ***X*** of the Markov Chain is unknown.

Hidden Markov Models (HMM)

- In some practical applications, the state sequence $\boldsymbol{X}$ of the Markov Chain is unknown.
- And what is actually available to the observer is a sequence of observations: $\boldsymbol{Y} = Y_1 , Y_2 , Y_3 \ldots Y_k$

Hidden Markov Models (HMM)

- In some practical applications, the state sequence ***X*** of the Markov Chain is unknown.
- And what is actually available to the observer is a sequence of observations: $\boldsymbol{Y} = Y_1, Y_2, Y_3 \ldots Y_k$
- If these observations constitute a deterministic function of the state sequence, it is still a **Markov Chain**.

Hidden Markov Models (HMM)

- In some practical applications, the state sequence $\boldsymbol{X}$ of the Markov Chain is unknown.
- And what is actually available to the observer is a sequence of observations: $\boldsymbol{Y} = Y_1 , Y_2 , Y_3 \ldots Y_k$
- If these observations constitute a deterministic function of the state sequence, it is still a Markov Chain.
- If these observations constitute a probabilistic function of the state sequence, we talk about **Hidden Markov Models**.

Elements of HMMs

- State sequence: $\boldsymbol{X} = X_1 , X_2 , X_3 \ldots X_k$
- State alphabet (state set): $\boldsymbol{S} = \{s_1 , s_2 , s_3 \ldots s_m\}$

Elements of HMMs

- State sequence: $\boldsymbol{X} = X_1 , X_2 , X_3 \ldots X_k$
- State alphabet (state set): $\boldsymbol{S} = \{s_1 , s_2 , s_3 \ldots s_m\}$

- Observation sequence: $\boldsymbol{Y} = Y_1 , Y_2 , Y_3 \ldots Y_k$
- Observation alphabet (symbols): $\boldsymbol{O} = \{o_1 , o_2 , o_3 \ldots o_h\}$

Elements of HMMs

- State sequence: $\boldsymbol{X} = X_1 , X_2 , X_3 \ldots X_k$
- State alphabet (state set): $\boldsymbol{S} = \{s_1 , s_2 , s_3 \ldots s_m\}$

- Observation sequence: $\boldsymbol{Y} = Y_1 , Y_2 , Y_3 \ldots Y_k$
- Observation alphabet (symbols): $\boldsymbol{O} = \{o_1 , o_2 , o_3 \ldots o_h\}$

- Initial state probabilities: $\boldsymbol{\pi} = \{\pi_i\}$ with $i \in \boldsymbol{S}$
- State transition probabilities: $\mathbf{P} = \{p_{ij}\}$ with $i, j \in \boldsymbol{S}$

Elements of HMMs

- State sequence: $\boldsymbol{X} = X_1 , X_2 , X_3 \ldots X_k$
- State alphabet (state set): $\boldsymbol{S} = \{s_1 , s_2 , s_3 \ldots s_m\}$

- Observation sequence: $\boldsymbol{Y} = Y_1 , Y_2 , Y_3 \ldots Y_k$
- Observation alphabet (symbols): $\boldsymbol{O} = \{o_1 , o_2 , o_3 \ldots o_h\}$

- Initial state probabilities: $\boldsymbol{\pi} = \{\pi_i\}$ with $i \in \boldsymbol{S}$
- State transition probabilities: $\mathbf{P} = \{p_{ij}\}$ with $i, j \in \boldsymbol{S}$

- Symbol emission probabilities: $\mathbf{Q} = \{q_{ijk}\}$ with $i, j \in \boldsymbol{S}$, $k \in \boldsymbol{O}$

About HMM parameters

- As in the case of Markov Chains, $\boldsymbol{\pi}$ is a vector of m elements which represent the initial probability distribution of states.

About HMM parameters

- As in the case of Markov Chains, $\boldsymbol{\pi}$ is a vector of m elements which represent the initial probability distribution of states.
- Similarly, **P** is an $m_x m$ matrix with elements $p_{ij} = P(X_n = s_j | X_{n-1} = s_i)$ which represent the probability distribution of state transitions.

About HMM parameters

- As in the case of Markov Chains, π is a vector of m elements which represent the initial probability distribution of states.
- Similarly, **P** is an mxm matrix with elements $p_{ij} = P(X_n = s_j | X_{n-1} = s_i)$ which represent the probability distribution of state transitions.
- However, **Q** constitutes the observation probability distribution that is associated with the state transitions:

$$q_{ijk} = q(Y_n \mid X_n, X_{n-1}) = P(Y_n = o_k \mid X_n = s_j\,,\, X_{n-1} = s_i)$$

Important problems related to HMMs

1.- Given a HMM defined as $\lambda = \{\pi, \mathbf{P}, \mathbf{Q}\}$, what is the probability of an observation sequence $\boldsymbol{Y} = Y_1 , Y_2 , Y_3 \ldots Y_k$?

Important problems related to HMMs

1.- Given a HMM defined as $\lambda = \{\pi, \mathbf{P}, \mathbf{Q}\}$, what is the probability of an observation sequence $\boldsymbol{Y} = Y_1, Y_2, Y_3 \ldots Y_k$?

↓

Dynamic Programming

Important problems related to HMMs

2.- Given an observation sequence $\boldsymbol{Y} = Y_1 , Y_2 , Y_3 \ldots Y_k$ and a HMM $\lambda = \{\pi, \mathbf{P}, \mathbf{Q}\}$, which is the state sequence $\boldsymbol{X} = X_1 , X_2 , X_3 \ldots X_k$ that better explains the observations ?

Important problems related to HMMs

2.- Given an observation sequence $\boldsymbol{Y} = Y_1, Y_2, Y_3 \ldots Y_k$ and a HMM $\lambda = \{\pi, \mathbf{P}, \mathbf{Q}\}$, which is the state sequence $\boldsymbol{X} = X_1, X_2, X_3 \ldots X_k$ that better explains the observations ?

Viterbi Algorithm

Important problems related to HMMs

3.- Given an observation sequence $\mathbf{Y} = Y_1 , Y_2 , Y_3 \ldots Y_k$ and a space of models $\lambda = \{\pi, \mathbf{P}, \mathbf{Q}\}$, which is the model that better explains the observations ?

Important problems related to HMMs

3.- Given an observation sequence $\boldsymbol{Y} = Y_1 , Y_2 , Y_3 \ldots Y_k$ and a space of models $\lambda = \{\pi, \mathbf{P}, \mathbf{Q}\}$, which is the model that better explains the observations ?

⬇

Expectation Maximization

Exercise 2

Consider the same Markov Chain from exercise 1:

- State alphabet: $\boldsymbol{S} = \{s_1 , s_2 , s_3\}$
- Initial probability distribution: $\pi = [1, 0, 0]$
- Transition matrix: $\mathbf{P} = \begin{pmatrix} 0.2 & 0.8 & 0.0 \\ 0.0 & 0.6 & 0.4 \\ 0.5 & 0.2 & 0.3 \end{pmatrix}$

Exercise 2

But now, consider the binary observation alphabet: $\boldsymbol{O} = \{A , B\}$

And the following symbol emission probabilities:

$$q_{ijA} = q(\, A \mid X_n, X_{n-1}\,) \qquad q_{ijB} = q(\, B \mid X_n, X_{n-1}\,)$$

$$\begin{pmatrix} 1.0 & 0.5 & 0.5 \\ 0.5 & 1.0 & 0.5 \\ 0.5 & 0.5 & 1.0 \end{pmatrix} \qquad \begin{pmatrix} 0.0 & 0.5 & 0.5 \\ 0.5 & 0.0 & 0.5 \\ 0.5 & 0.5 & 0.0 \end{pmatrix}$$

Exercise 2

1.- What is the probability of observing the output sequence
$\boldsymbol{Y} = A, A, A$?

2.- What is the probability of observing the output sequence
$\boldsymbol{Y} = B, B, B$?

TO BE CONTINUED ...

Hidden Markov Models

Rafael E. Banchs (rbanchs@gps.tsc.upc.edu)

Session 1: Introduction to HMMs

TALP
UPC
Centre de Tecnologies i Aplicacions del Llenguatge i la Parla
UNIVERSITAT POLITÈCNICA DE CATALUNYA