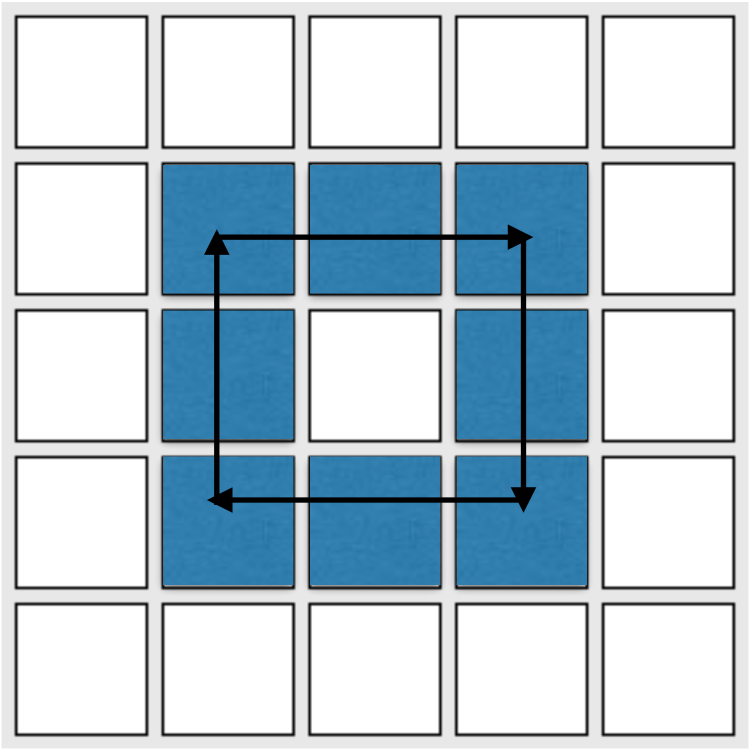
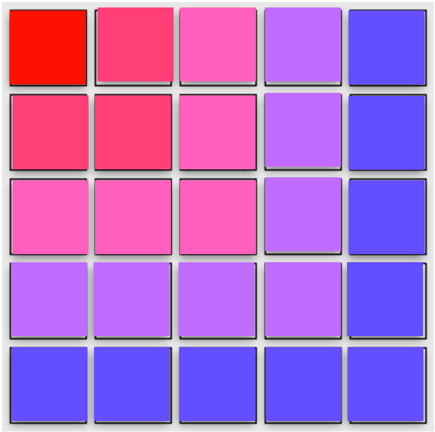
Machine Learning for Data Science (CS4786)
Lecture 19

Hidden Markov Models

Course Webpage :
http://www.cs.cornell.edu/Courses/cs4786/2016fa/

HIDDEN MARKOV MODEL (HMM)

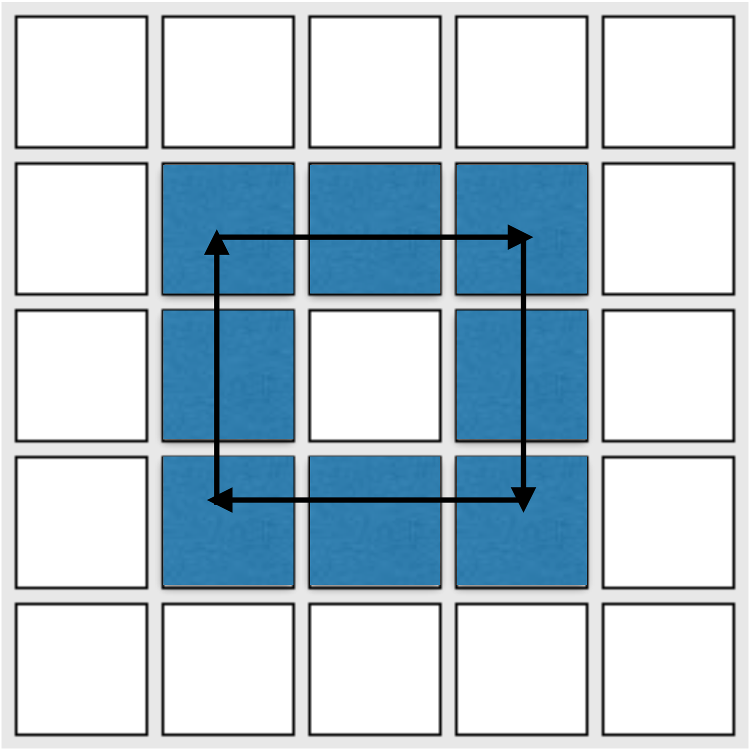


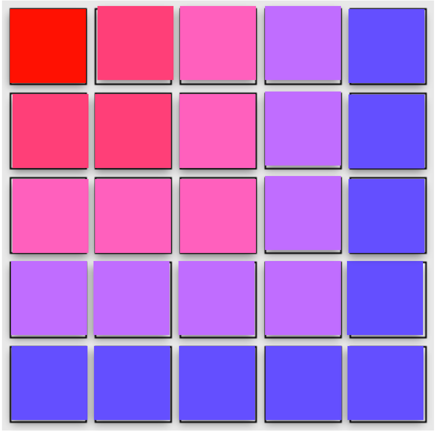


Can you catch the Bot?

Hidden Markov Model (HMM)

Same example:



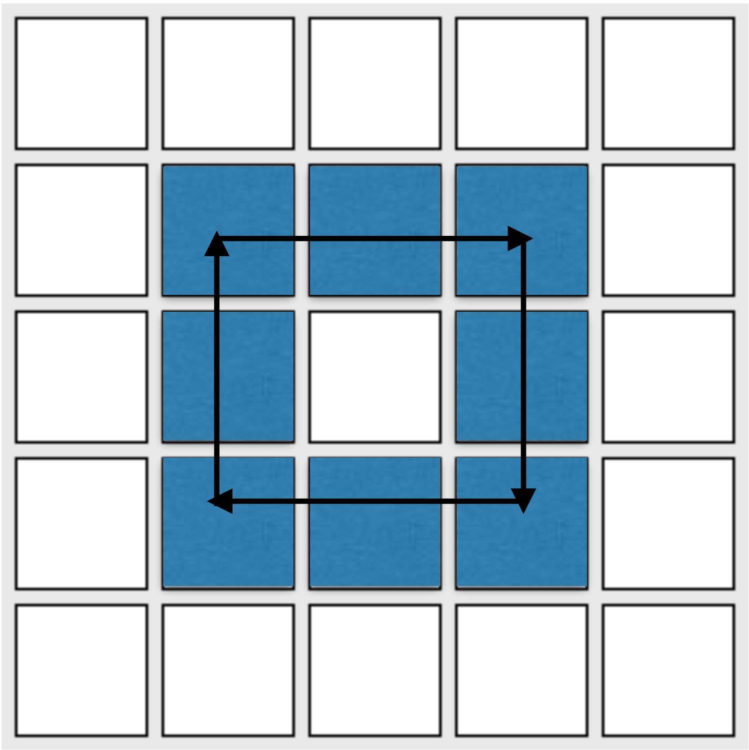


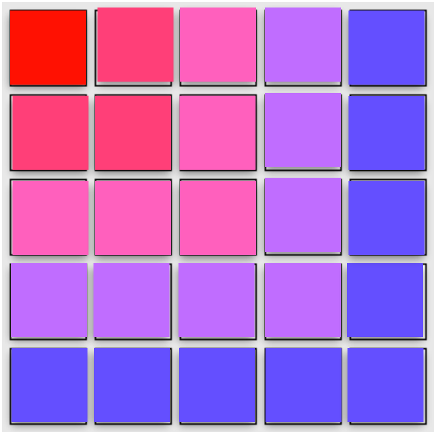
Can you catch the Bot?

Hidden Markov Model (HMM)

Same example:

But you don't observe location (dark room)

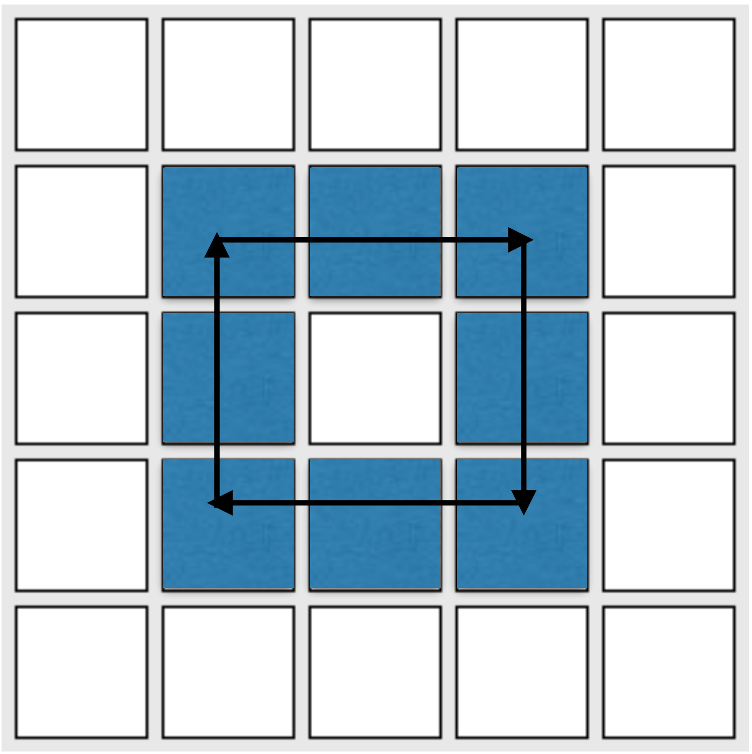




Can you catch the Bot?

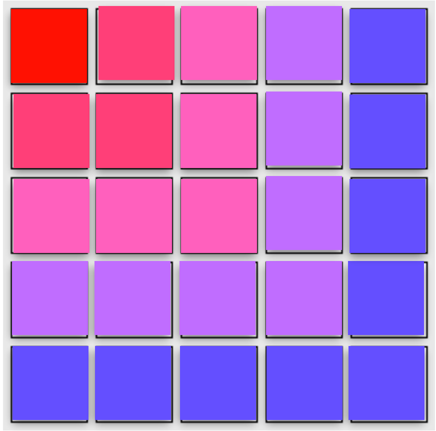
Hidden Markov Model (HMM)

Same example:



But you don't observe location (dark room)

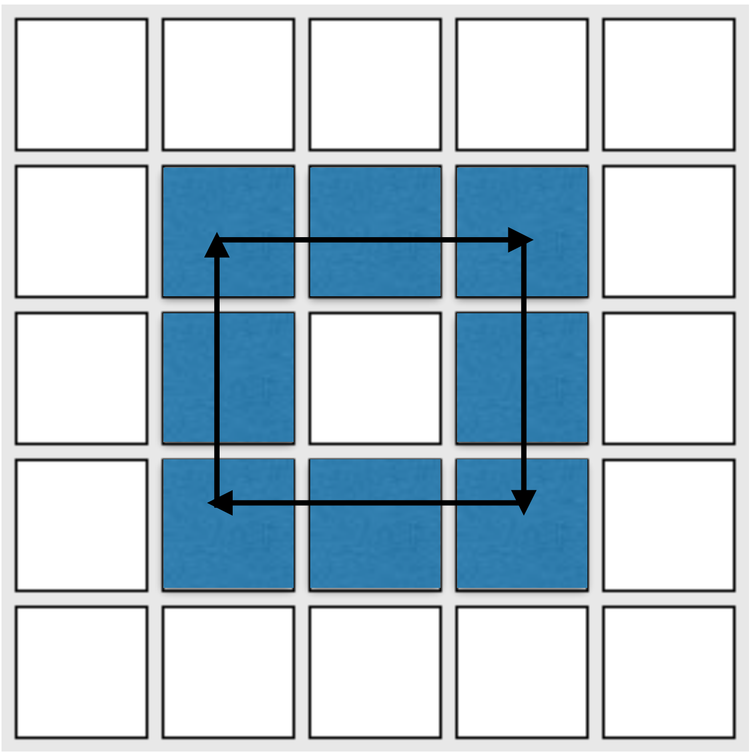
You hear how close the bot is!



Can you catch the Bot?

Hidden Markov Model (HMM)

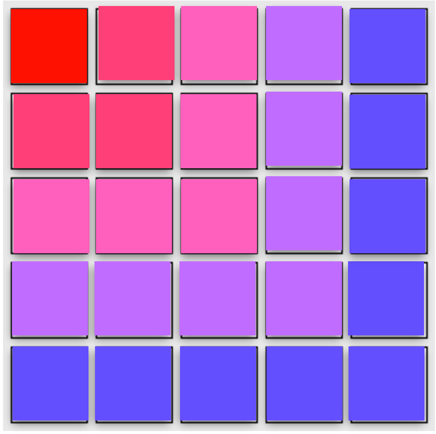
Same example:



But you don't observe location (dark room)

You hear how close the bot is!

What you hear:



Can you catch the Bot?

Hidden Markov Model (HMM)

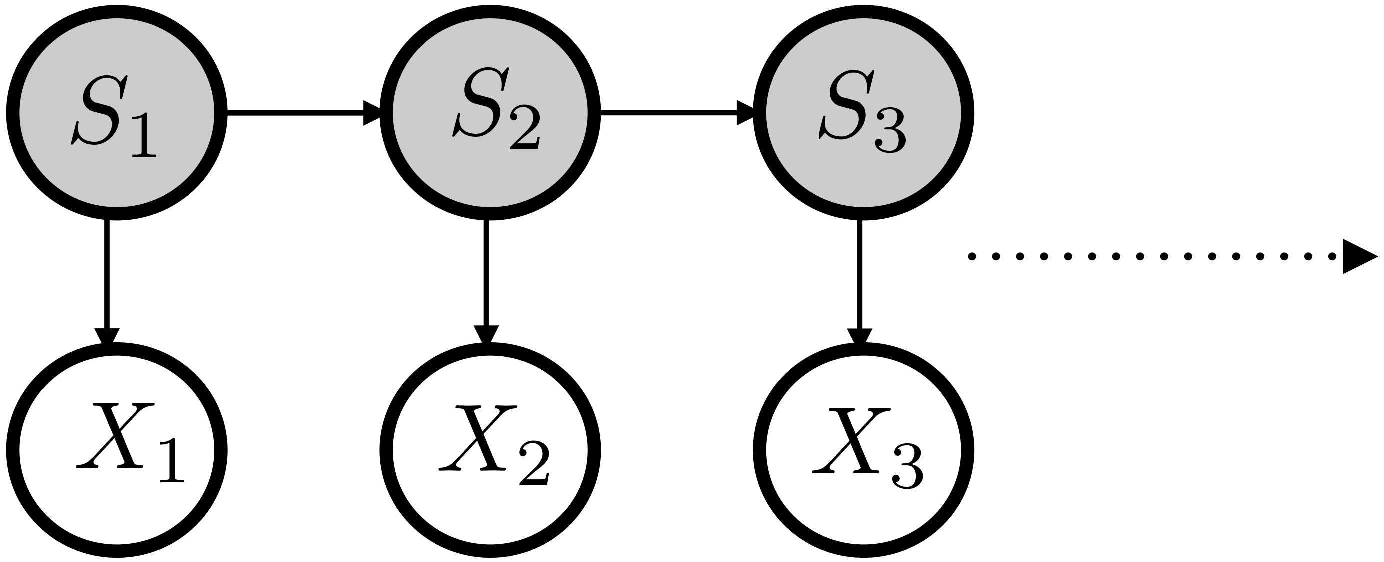
X_t's are what you hear (observation)

S_t's are the unseen locations (states)

Eg: for n x n grid we have, $K = n^2$ states

Number of alphabets = 5
(colors you can observe)

HIDDEN MARKOV MODEL (HMM)



X_t's are what you hear (observation)

S_t's are the unseen locations (states)

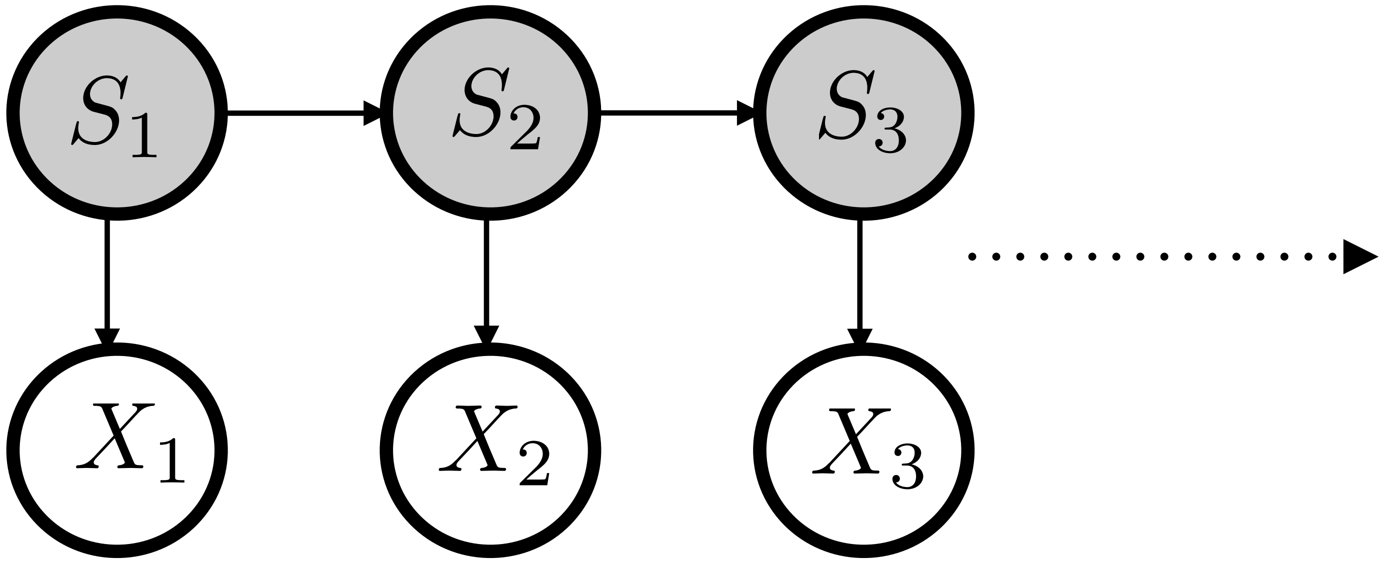
Eg: for n x n grid we have, $K = n^2$ states

Number of alphabets = 5
(colors you can observe)

Hidden Markov Model (HMM)

What are the parameters?

Hidden Markov Model (HMM)



What are the parameters?

Hidden Markov Model (HMM)

Hidden Markov Model (HMM)

- What is probability that bot will be in location k at time t given the entire sequence of observations?

Hidden Markov Model (HMM)

- What is probability that bot will be in location k at time t given the entire sequence of observations?

$$P(S_t = k | X_1, \ldots, X_N)?$$

Inference in HMM

$$
\begin{aligned}
P(S_t = k|X_1, \ldots, X_N) \\
\propto P(X_{t+1}, \ldots, X_N|S_t = k, X_1, \ldots, X_t)P(S_t = k|X_1, \ldots, X_t) \\
\propto P(X_{t+1}, \ldots, X_N|S_t = k, X_1, \ldots, X_t)P(S_t = k, X_1, \ldots, X_t) \\
\propto P(X_{t+1}, \ldots, X_N|S_t = k, X_1, \ldots, X_t)P(X_t|S_t = k, X_1, \ldots, X_{t-1})P(S_t = k, X_1, \ldots, X_{t-1}) \\
\propto P(X_{t+1}, \ldots, X_N|S_t = k)P(X_t|S_t = k)P(S_t = k, X_1, \ldots, X_{t-1})
\end{aligned}
$$

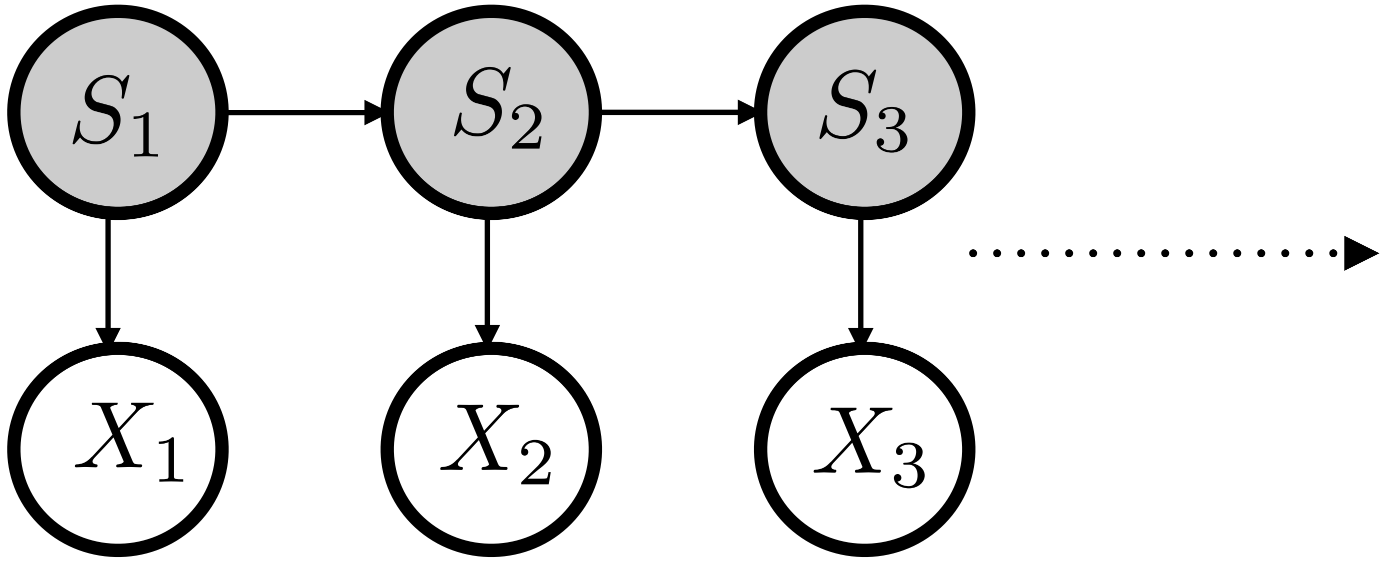
Inference in HMM

$$
\begin{aligned}
P(S_t = k | X_1, \dots, X_N) \\
&\propto P(X_{t+1}, \dots, X_N | S_t = k, X_1, \dots, X_t) P(S_t = k | X_1, \dots, X_t) \\
&\propto P(X_{t+1}, \dots, X_N | S_t = k, X_1, \dots, X_t) P(S_t = k, X_1, \dots, X_t) \\
&\propto P(X_{t+1}, \dots, X_N | S_t = k, X_1, \dots, X_t) P(X_t | S_t = k, X_1, \dots, X_{t-1}) P(S_t = k, X_1, \dots, X_{t-1}) \\
&\propto P(X_{t+1}, \dots, X_N | S_t = k) P(X_t | S_t = k) P(S_t = k, X_1, \dots, X_{t-1})
\end{aligned}
$$

We know $P(X_t | S_t = k)$'s and $P(S_t | S_{t-1})$

Compute $P(X_{t+1}, \dots, X_N)$ and $P(S_t = k, X_1, \dots, X_{t-1})$ recursively.

Inference in HMM

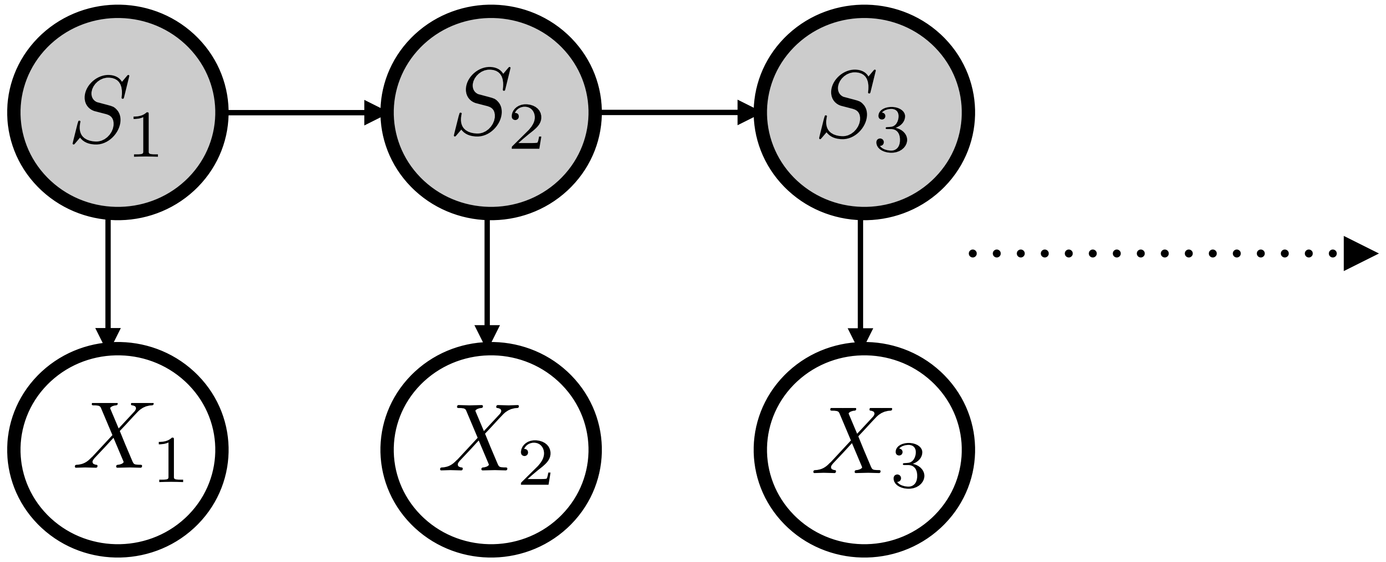


$$\text{message}_{S_{t-1} \mapsto S_t}(k) = P(S_t = k, X_1, \ldots, X_{t-1})$$

$$\text{message}_{S_{t+1} \mapsto S_t}(k) = P(X_n, \ldots, X_{t+1} | S_t = k)$$

$$P(S_t = k | X_1, \ldots, X_n) \propto \text{message}_{S_{t-1} \mapsto S_t}(k) \times \text{message}_{S_{t+1} \mapsto S_t}(k) \times P(X_t | S_t = k)$$

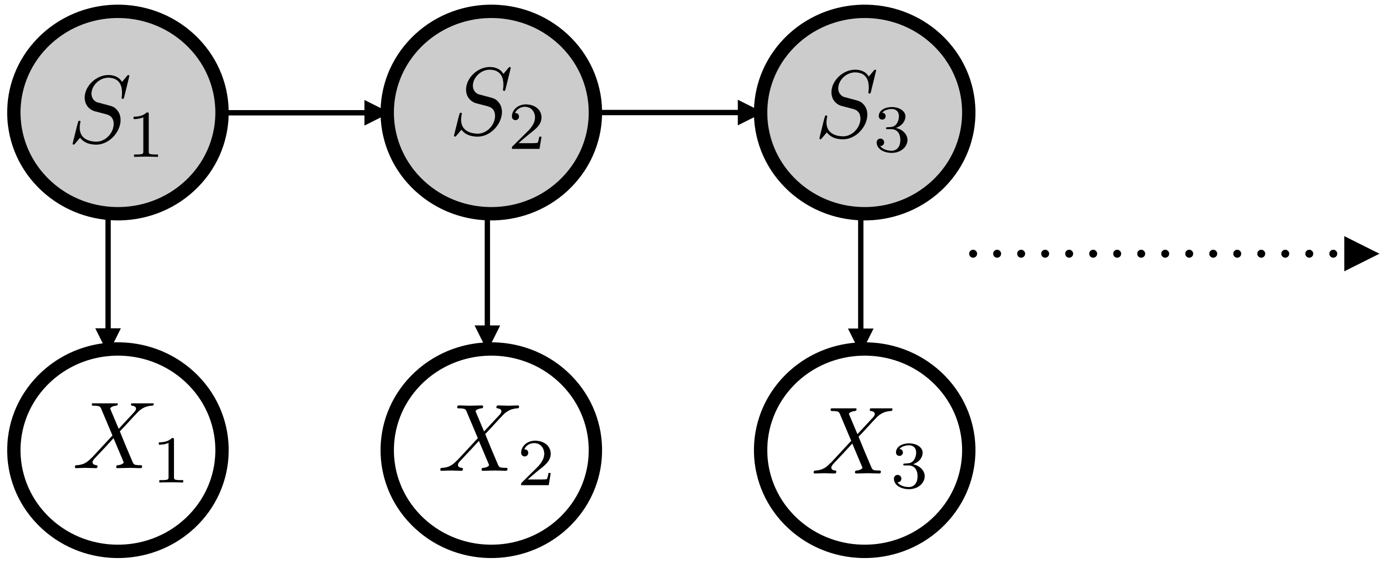
Inference in HMM



$$\text{message}_{S_{t-1} \mapsto S_t}(k) = P(S_t = k, X_1, \ldots, X_{t-1})$$

$$\text{message}_{S_{t+1} \mapsto S_t}(k) = P(X_n, \ldots, X_{t+1} | S_t = k)$$

Inference in HMM



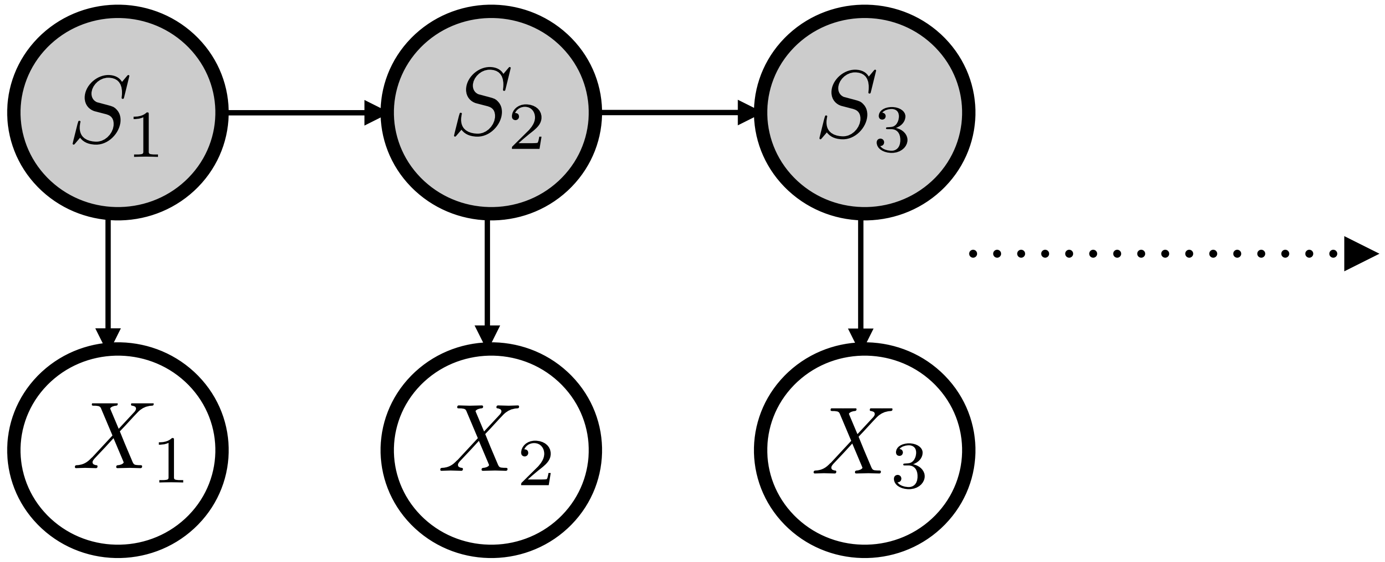
$$\text{message}_{S_{t-1} \mapsto S_t}(k) = P(S_t = k, X_1, \ldots, X_{t-1})$$

$$\text{message}_{S_{t+1} \mapsto S_t}(k) = P(X_n, \ldots, X_{t+1} | S_t = k)$$

Forward:

$$P(X_1, \ldots, X_{t-1}, S_t = k) = \sum_{j=1}^{K} P(S_t = k | S_{t-1} = j) P(X_{t-1} | S_{t-1} = j) P(X_1, \ldots, X_{t-2}, S_{t-1} = j)$$

Inference in HMM



$$\text{message}_{S_{t-1}\mapsto S_t}(k) = P(S_t = k, X_1, \ldots, X_{t-1})$$

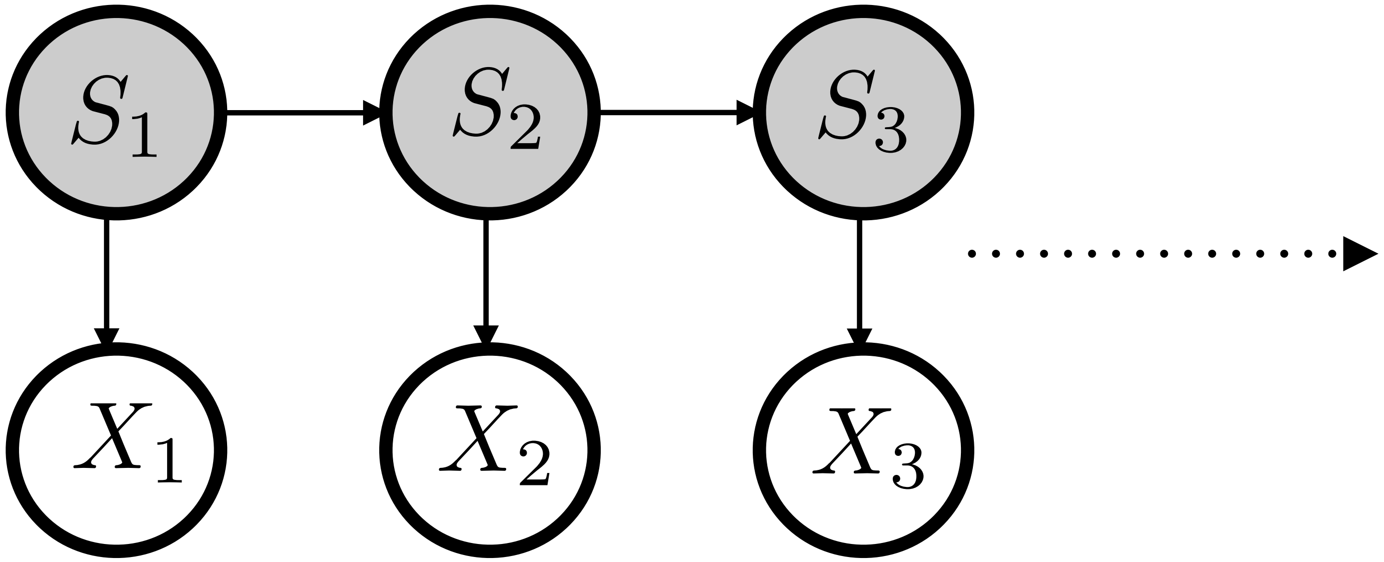
$$\text{message}_{S_{t+1}\mapsto S_t}(k) = P(X_n, \ldots, X_{t+1} | S_t = k)$$

Forward:

$$P(X_1, \ldots, X_{t-1}, S_t = k) = \sum_{j=1}^{K} P(S_t = k | S_{t-1} = j) P(X_{t-1} | S_{t-1} = j) P(X_1, \ldots, X_{t-2}, S_{t-1} = j)$$

$$\text{message}_{S_{t-1}\mapsto S_t}(k) = \sum_{j=1}^{K} P(S_t = k | S_{t-1} = j) P(X_{t-1} | S_{t-1} = j) \text{message}_{S_{t-2}\mapsto S_{t-1}}(j)$$

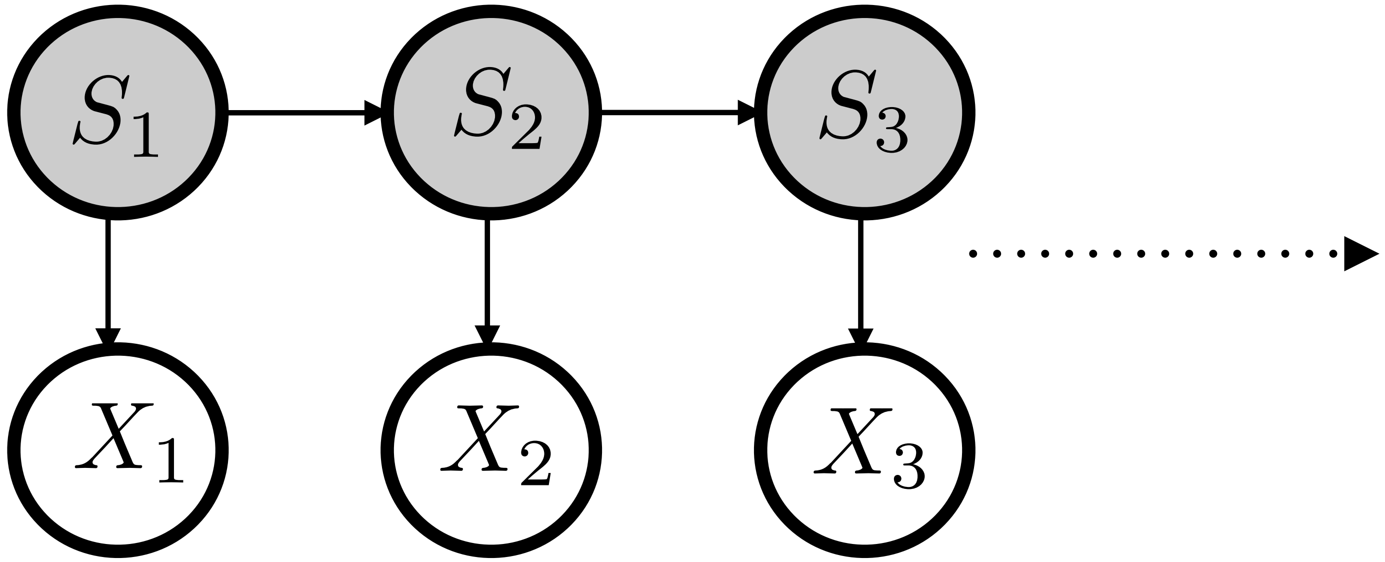
Inference in HMM



$$\text{message}_{S_{t-1} \mapsto S_t}(k) = P(S_t = k, X_1, \ldots, X_{t-1})$$

$$\text{message}_{S_{t+1} \mapsto S_t}(k) = P(X_n, \ldots, X_{t+1} | S_t = k)$$

Inference in HMM



$$\text{message}_{S_{t-1}\mapsto S_t}(k) = P(S_t = k, X_1, \ldots, X_{t-1})$$

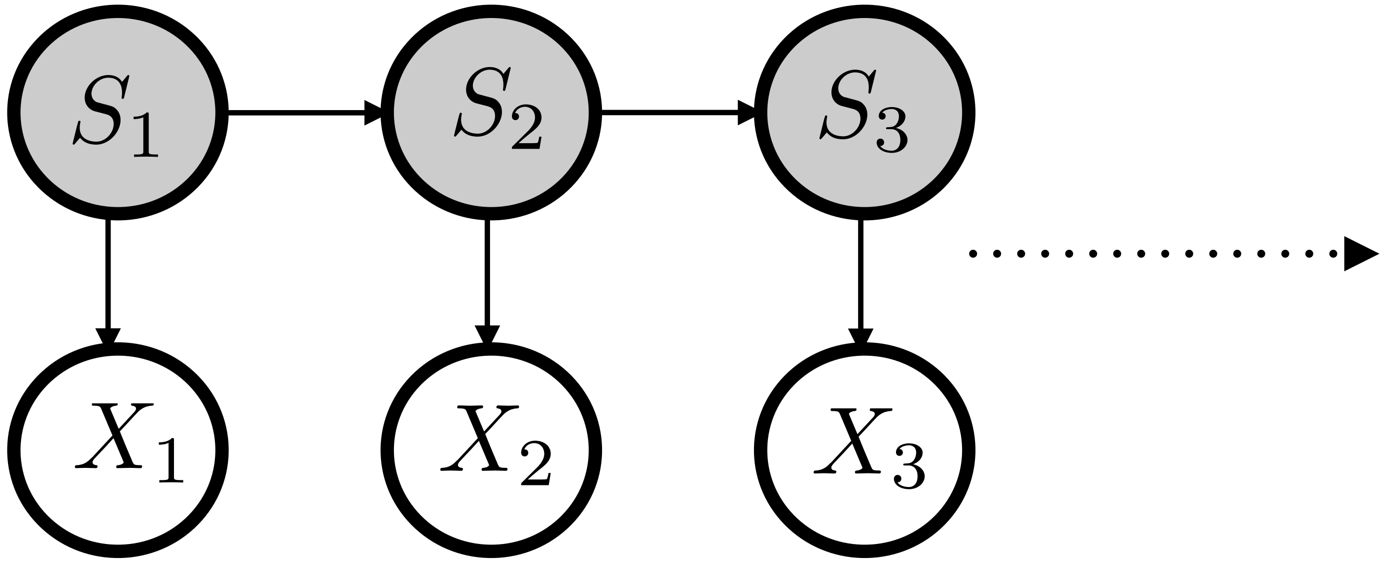
$$\text{message}_{S_{t+1}\mapsto S_t}(k) = P(X_n, \ldots, X_{t+1} | S_t = k)$$

Backward:

$$P(X_n, \ldots, X_{t+1}|S_t = k) = \sum_{j=1}^{K} P(X_n, \ldots, X_{t+2}|S_{t+1} = j)P(X_{t+1}|S_{t+1} = j)P(S_{t+1} = j|S_t = k)$$

K

Inference in HMM



$$\text{message}_{S_{t-1}\mapsto S_t}(k) = P(S_t = k, X_1, \ldots, X_{t-1})$$

$$\text{message}_{S_{t+1}\mapsto S_t}(k) = P(X_n, \ldots, X_{t+1}|S_t = k)$$

Backward:

$$P(X_n, \ldots, X_{t+1}|S_t = k) = \sum_{j=1}^{K} P(X_n, \ldots, X_{t+2}|S_{t+1} = j)P(X_{t+1}|S_{t+1} = j)P(S_{t+1} = j|S_t = k)$$

$$\text{message}_{S_{t+1}\mapsto S_t}(k) = \sum_{j=1}^{K} \text{message}_{S_{t+2}\mapsto S_{t+1}}(j)P(X_{t+1}|S_{t+1} = j)P(S_{t+1} = j|S_t = k)$$

Learning Parameters for HMM

- Now that we have algorithm for inference, what about learning
- Given observations, how do we estimate parameters for HMM? Three guesses …

EM FOR HMM (BAUM WELCH)

- EM algorithm of course, for HMM its referred to as Baum Welch algorithm
- Initialize Transition and Emission probability tables arbitrarily
- For $i = 1$ to convergence:

E-step For every state variable $t \in \{1, \ldots, n\}$,
Use forward-backward algorithm to compute probabilities of latent variables given obervation

M-step Optimize weighted log likelihood as usual:

$$\theta^{(i)} = \arg\max_{\theta \in \Theta} \sum_{s_1, \ldots, s_n} P(S_{1,\ldots,n} = s_{1,\ldots,n} | \theta^{(i-1)}) \log P(X_{1,\ldots,n}, S_{1,\ldots,n} = s_{1,\ldots,n} | \theta)$$