Machine Learning for Data Science (CS4786) Lecture 19

Hidden Markov Models

Course Webpage :
http://www.cs.cornell.edu/Courses/cs4786/2017fa/

Quiz

Quiz

- Two variables can be marginally independent but not conditionally independent given a third variable?
  - True False

Quiz

- Two variables can be marginally independent but not conditionally independent given a third variable?
  - True ✓ False

Quiz

- Two variables can be marginally independent but not conditionally independent given a third variable?
  - True ✓ False
- Two variables can be conditionally independent given a third variable but not marginally independent?
  - True False

Quiz

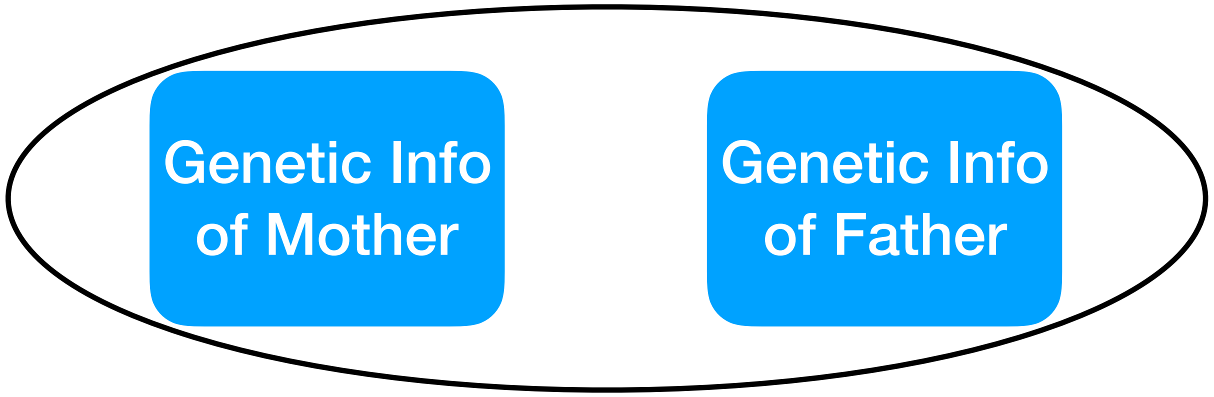
- Two variables can be marginally independent but not conditionally independent given a third variable?
    - True ✓ False

- Two variables can be conditionally independent given a third variable but not marginally independent?
    - True ✓ False

Example: CI and MI

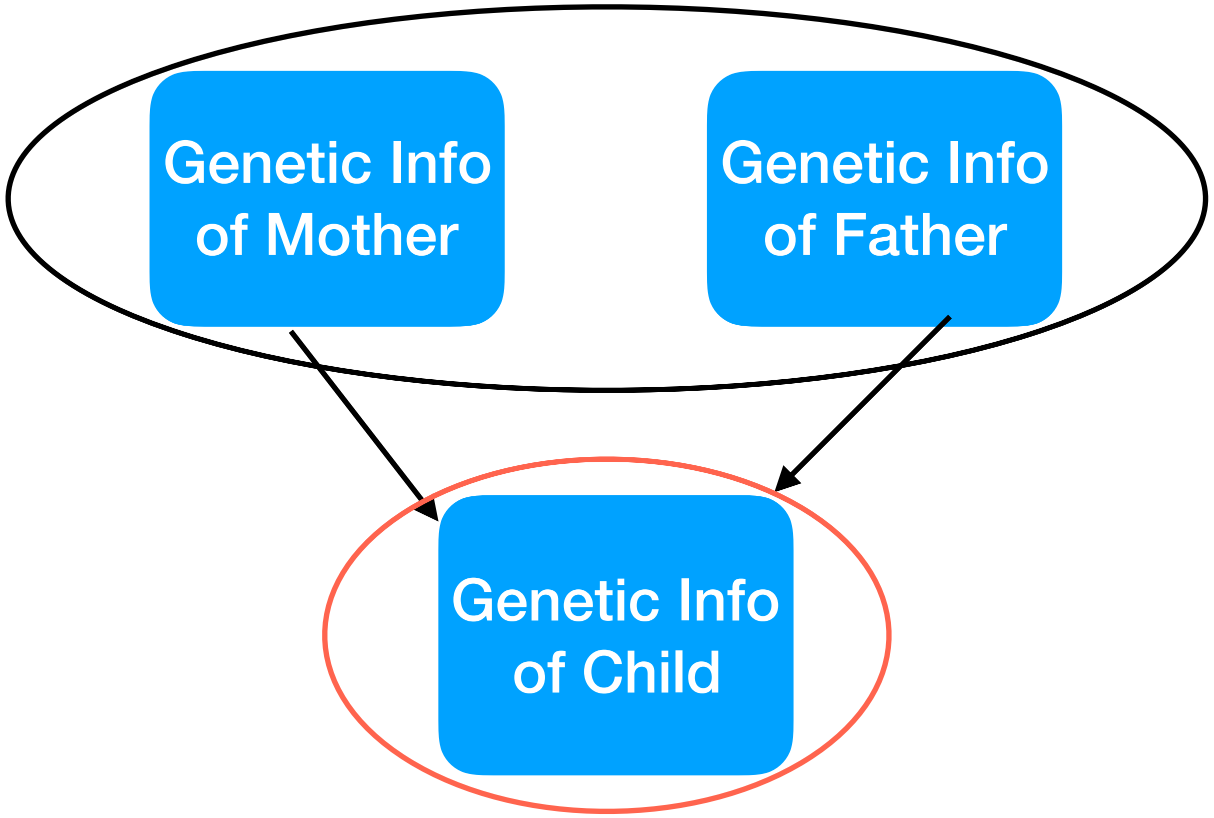
EXAMPLE: CI AND MI

Marginally independent



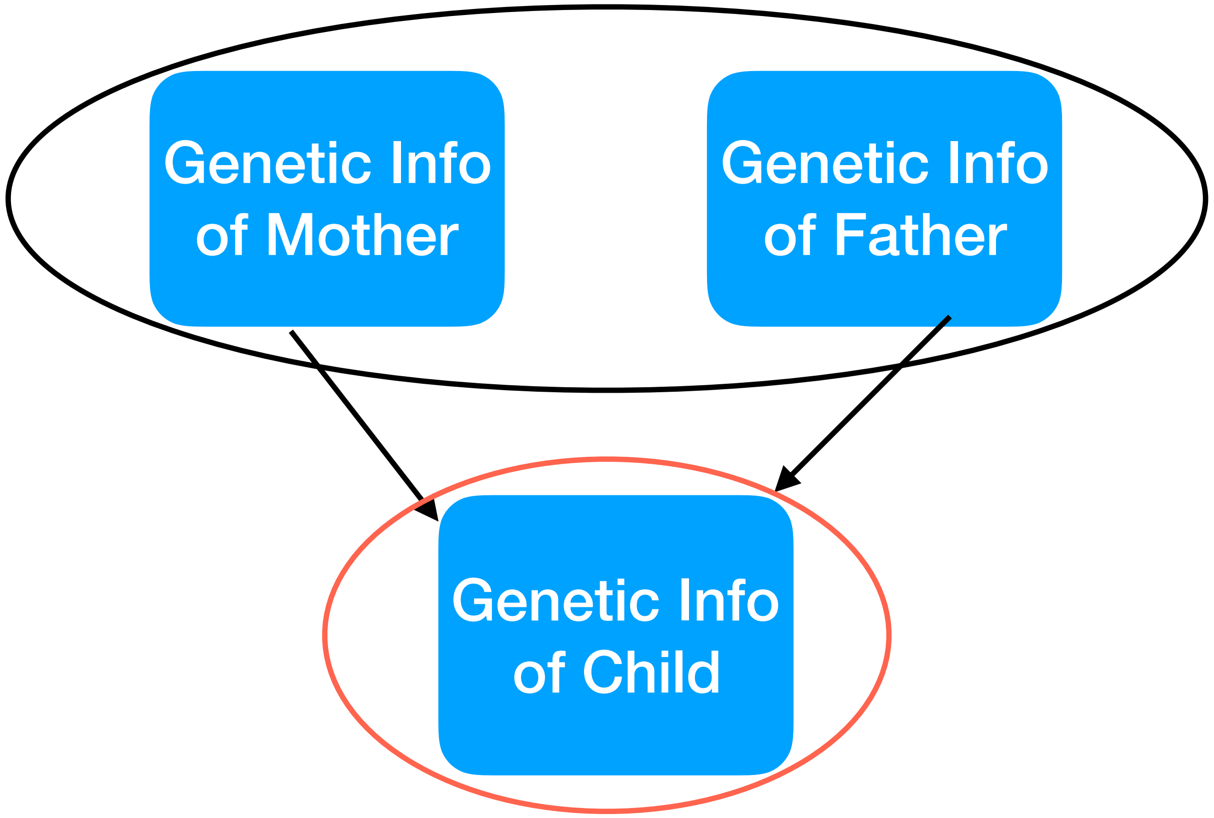
Example: CI and MI

Marginally independent
but Conditionally dependent
given child



EXAMPLE: CI AND MI

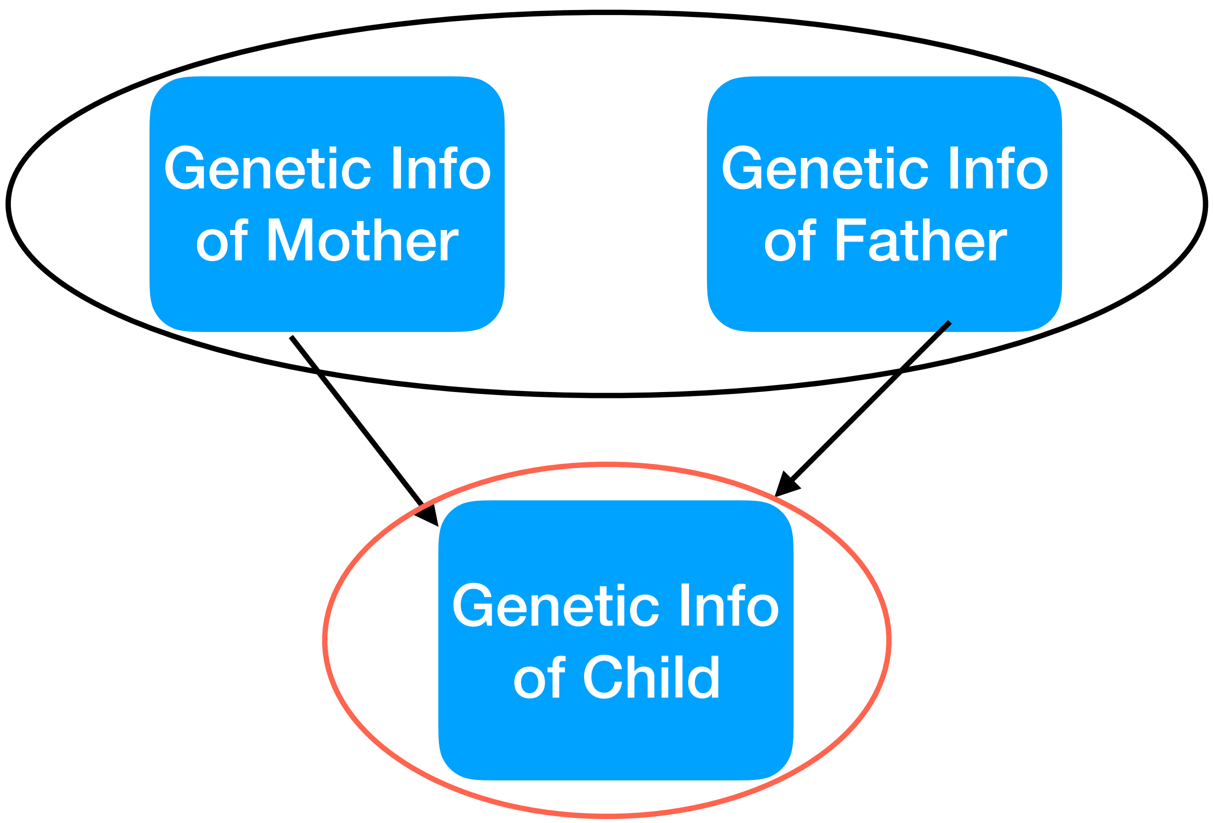
Marginally independent but Conditionally dependent given child



Given genetic info of child, with genetic info about mother, we can infer something about father

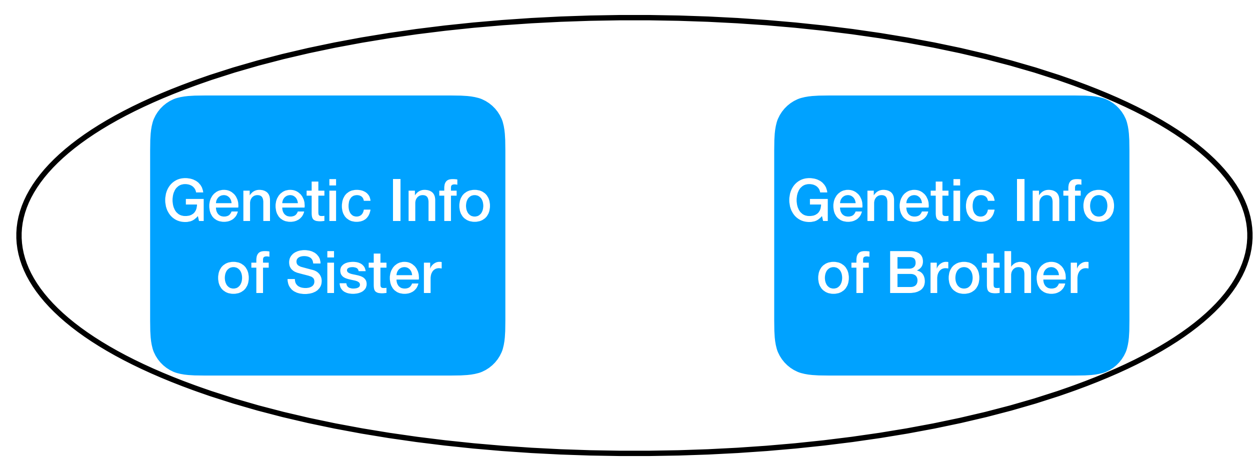
Example: CI and MI

Marginally independent
but Conditionally dependent
given child



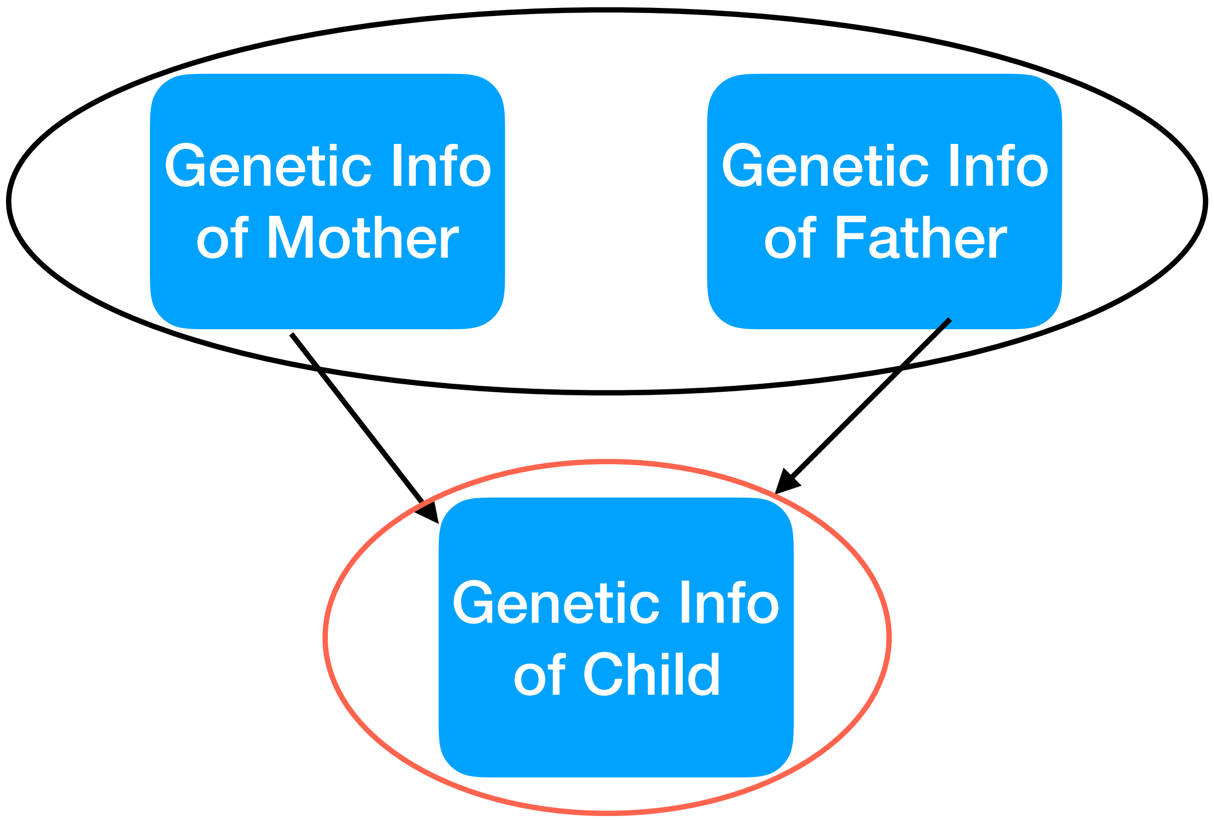
Given genetic info of child,
with genetic info about mother,
we can infer something about father

Marginally dependent



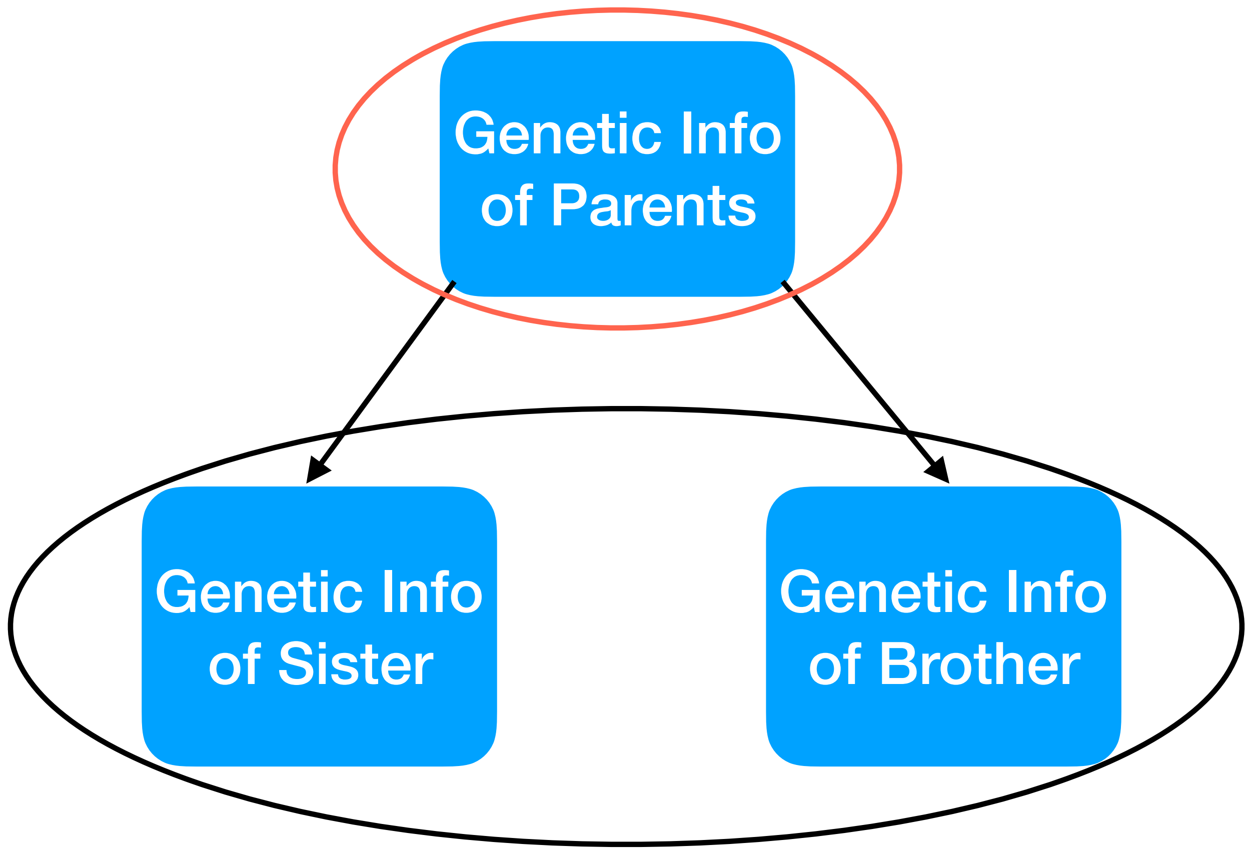
Example: CI and MI

Marginally independent
but Conditionally dependent
given child



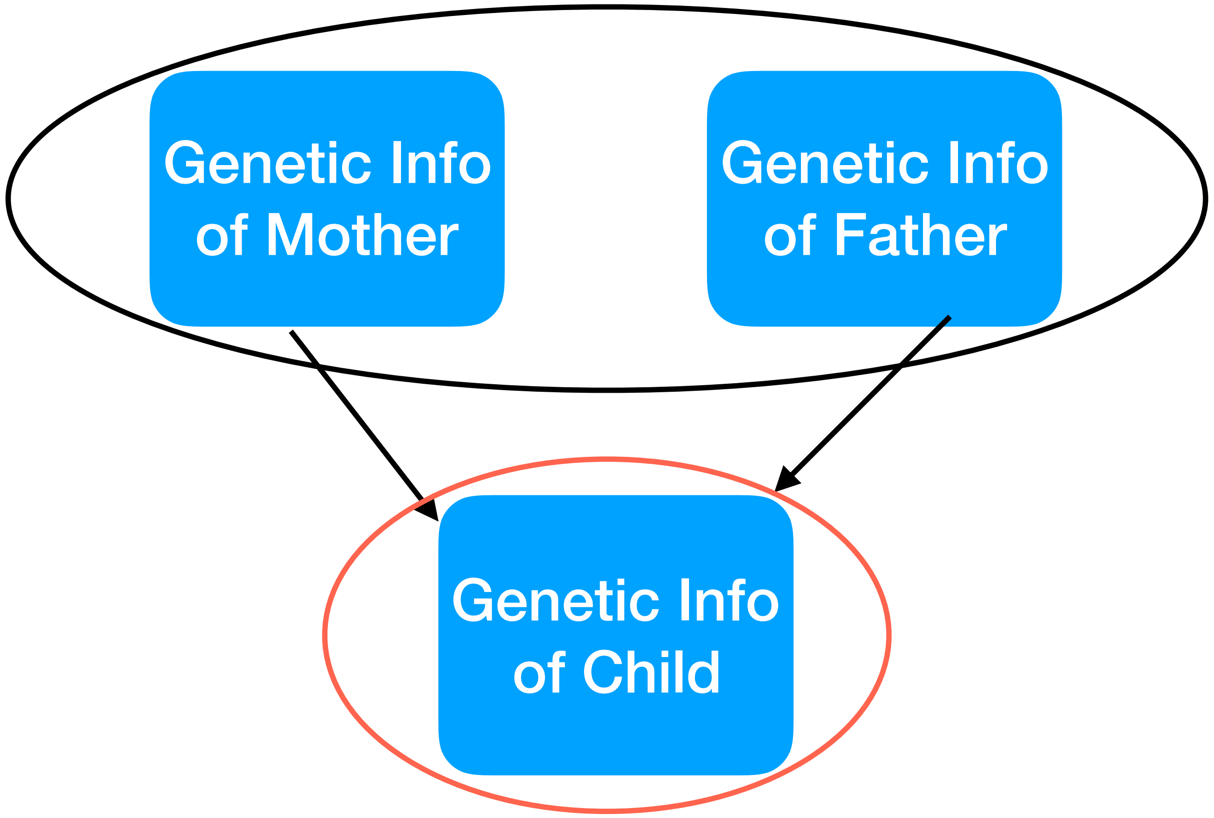
Given genetic info of child,
with genetic info about mother,
we can infer something about father

Marginally dependent
but Conditionally independent
given Parent



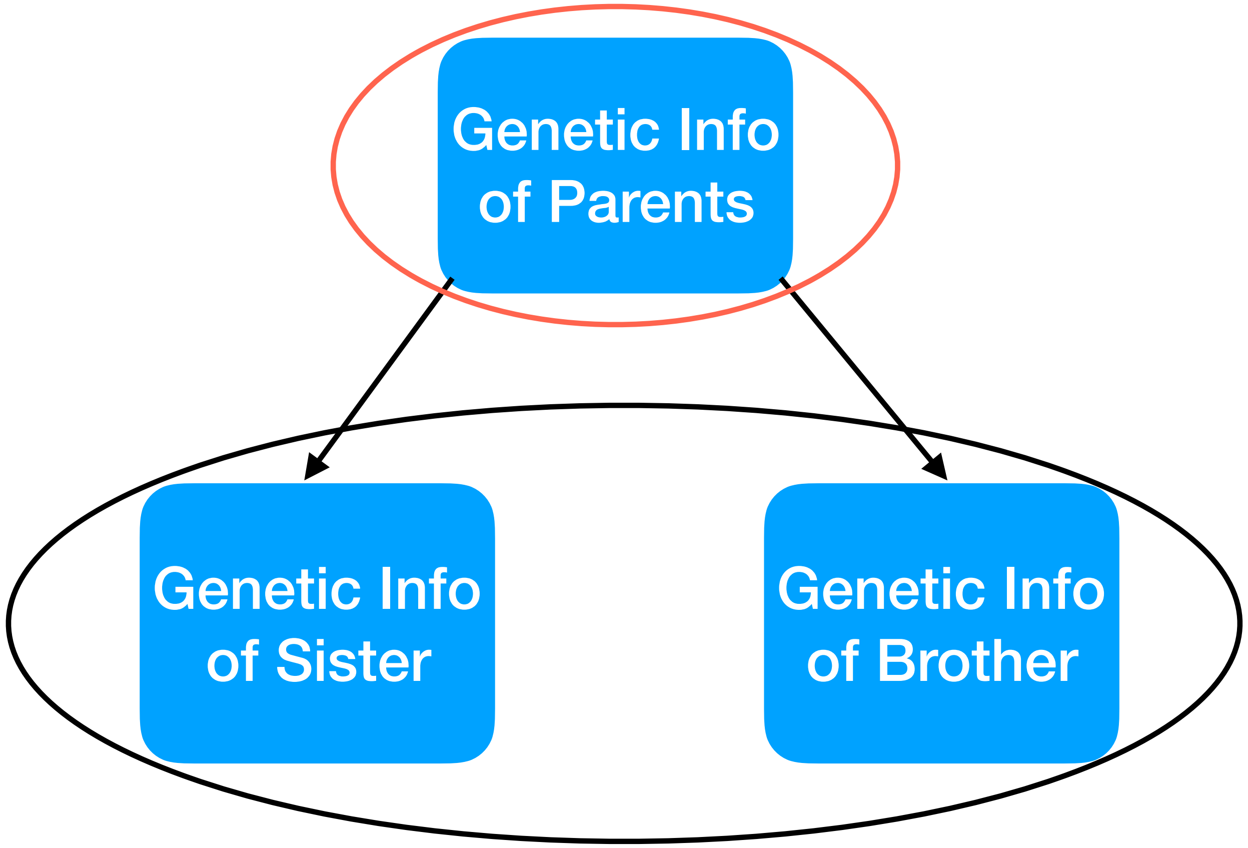
Example: CI and MI

Marginally independent
but Conditionally dependent
given child



Given genetic info of child,
with genetic info about mother,
we can infer something about father

Marginally dependent
but Conditionally independent
given Parent



Given genetic info about parents,
genetic info of sibling reveals nothing new
about myself.

Bayesian Networks

- Bayes net: directed acyclic graph + P(node|parents)

BAYESIAN NETWORKS

- Bayes net: directed acyclic graph + P(node|parents)
- Directed acyclic graph G = (V,E)
  - Edges going from parent nodes to child nodes
  - Direction indicates parent “generates” child

Bayesian Networks

- Bayes net: directed acyclic graph + P(node|parents)
- Directed acyclic graph $G = (V,E)$
  - Edges going from parent nodes to child nodes
  - Direction indicates parent "generates" child
- Provide conditional probability table/distribution P(node|parents)

BAYESIAN NETWORKS

- Bayes net: directed acyclic graph + P(node|parents)
- Directed acyclic graph G = (V,E)
  - Edges going from parent nodes to child nodes
  - Direction indicates parent "generates" child
- Provide conditional probability table/distribution P(node|parents)

$$P(X_1, \ldots, X_N) = \prod_{i=1}^{N} P(X_i | \text{Parents}(X_i))$$

Two main questions

Two main questions

- Learning/estimation: Given observations, can we learn the parameters for the graphical model ?

Two main questions

- Learning/estimation: Given observations, can we learn the parameters for the graphical model ?
- Inference: Given model parameters, can we answer queries about variables in the model

Two main questions

- Learning/estimation: Given observations, can we learn the parameters for the graphical model ?
- Inference: Given model parameters, can we answer queries about variables in the model
  - Eg. what is the most likely value of a latent variable given observations

Two main questions

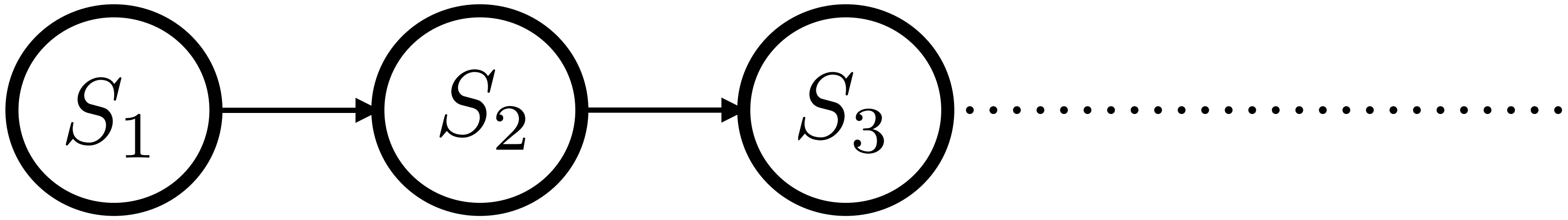
- Learning/estimation: Given observations, can we learn the parameters for the graphical model ?
- Inference: Given model parameters, can we answer queries about variables in the model
  - Eg. what is the most likely value of a latent variable given observations
  - Eg. What is the distribution of a particular variable conditioned on others

Hidden Markov Model (HMM)

- Speech recognition
- Natural language processing models
- Robot localization
- User attention modeling
- Medical monitoring

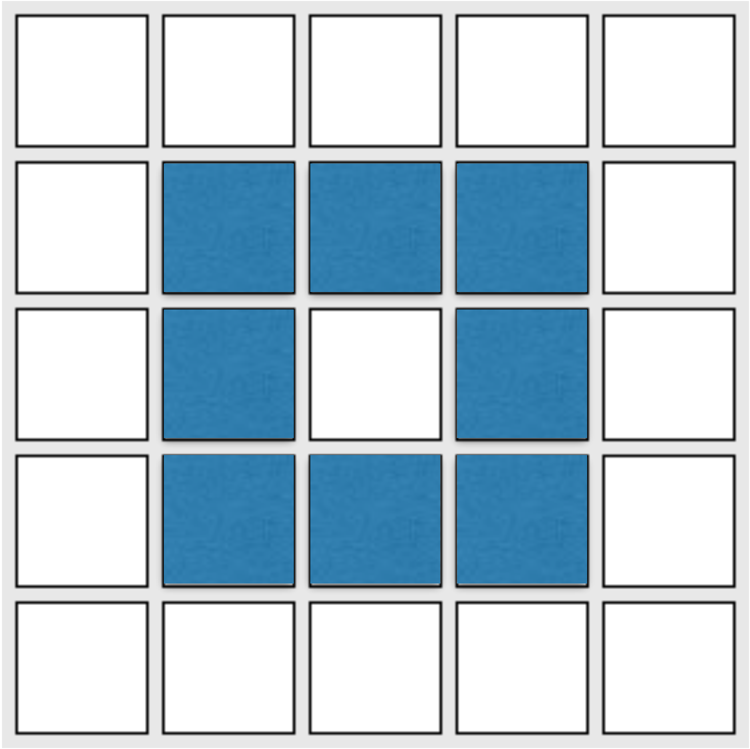
Time! … sequence of observations

Markov Model



- Each node is identically distributed given its predecessor (stationary)
- The values the nodes take are called states
- Parameters?
  - $P(S_1)$ the initial probability table
  - $P(S_t|S_{t-1})$ the transition probabilities

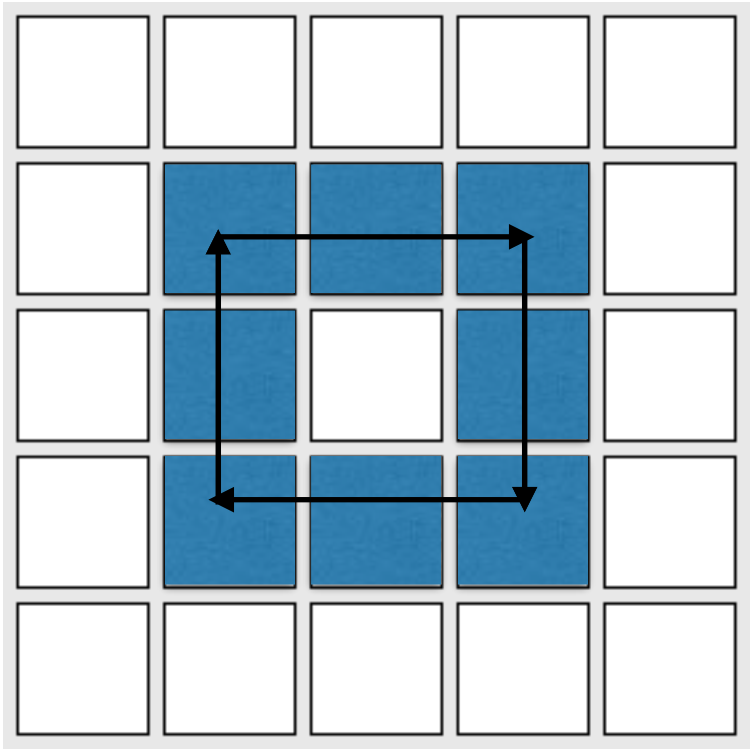
Markov Model



Bot tends to follow outlined path, but with some probability jumps to arbitrary neighbor

- Number of states: 25 (one for each location)
- For white boxes probability of jumping to any of the 4 neighbors is same 1/4
- For Blue boxes, probability of following path is 0.9 and jumping to some other neighbor is 0.0333333

Markov Model



Bot tends to follow outlined path, but with some probability jumps to arbitrary neighbor

- Number of states: 25 (one for each location)
- For white boxes probability of jumping to any of the 4 neighbors is same 1/4
- For Blue boxes, probability of following path is 0.9 and jumping to some other neighbor is 0.0333333

Markov Model

- If we observe the bot long enough, we get an estimate of its behavior (the transition table of jumping from state to state)
- If we observe enough number of times, we can also estimate initial distribution over states

MARKOV MODEL

- Inference question: what is probability that we will be in state k at time t? $P(S_t = k)$?

Answer:

MARKOV MODEL

- Inference question: what is probability that we will be in state k at time t? $P(S_t = k)$?

Answer:

$$P(S_t = k) = \sum_{s_1=1}^{K} \cdots \sum_{s_{t-1}=1}^{K} P(S_1 = s_1, \ldots, S_{t-1} = s_{t-1}, S_t = k)$$

$$= \sum_{s_1=1}^{K} \cdots \sum_{s_{t-1}=1}^{K} \prod_{i=1}^{t-1} (P(S_i = s_i | S_{i-1} = s_{i-1}) \times P(S_t = k | S_{t-1} = s_{t-1}))$$

Markov Model

- Inference question: what is probability that we will be in state k at time t? $P(S_t = k)$?

Answer:

$$P(S_t = k) = \sum_{s_1=1}^{K} \cdots \sum_{s_{t-1}=1}^{K} P(S_1 = s_1, \ldots, S_{t-1} = s_{t-1}, S_t = k)$$

$$= \sum_{s_1=1}^{K} \cdots \sum_{s_{t-1}=1}^{K} \prod_{i=1}^{t-1} \left(P(S_i = s_i | S_{i-1} = s_{i-1}) \times P(S_t = k | S_{t-1} = s_{t-1})\right)$$

For every t we can repeat the above or…

Markov Model

- Inference question: what is probability that we will be in state k at time t? $P(S_t = k)?$

Answer:

$$P(S_t = k) = \sum_{s_1=1}^{K} \cdots \sum_{s_{t-1}=1}^{K} P(S_1 = s_1, \ldots, S_{t-1} = s_{t-1}, S_t = k)$$

$$= \sum_{s_1=1}^{K} \cdots \sum_{s_{t-1}=1}^{K} \prod_{i=1}^{t-1} (P(S_i = s_i | S_{i-1} = s_{i-1}) \times P(S_t = k | S_{t-1} = s_{t-1}))$$

For every t we can repeat the above or…

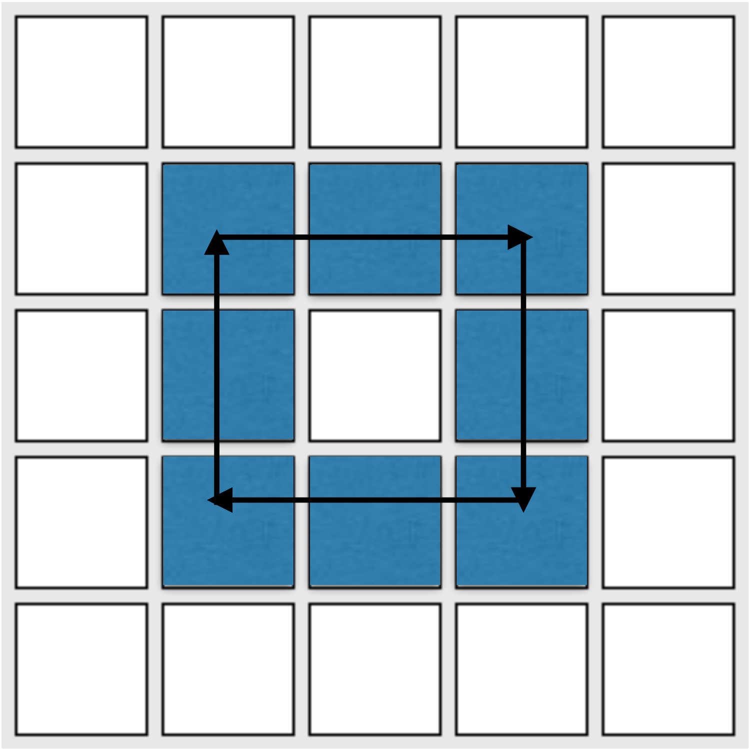
$$P(S_t = k) = \sum_{s_{t-1}=1}^{K} P(S_t = k | S_{t-1} = s_{t-1}) P(S_{t-1} = s_{t-1})$$

recursively compute probability of previous state

MARKOV MODEL

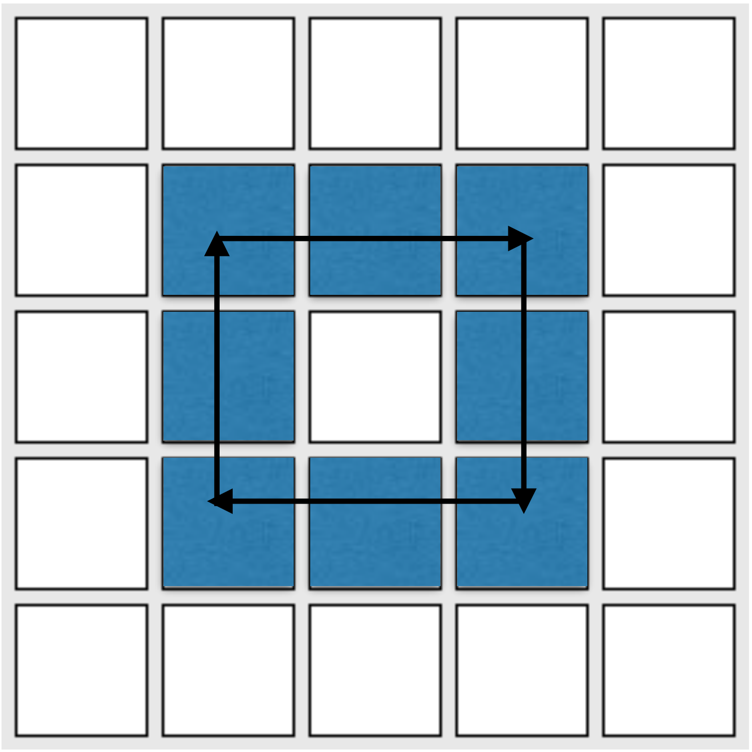
- As time goes by, $P(S_t = k)$ approaches a fixed distribution called stationary distribution
- Without any further observations, you are unlikely to find the bot on a new run (only by luck)

Hidden Markov Model (HMM)



Hidden Markov Model (HMM)

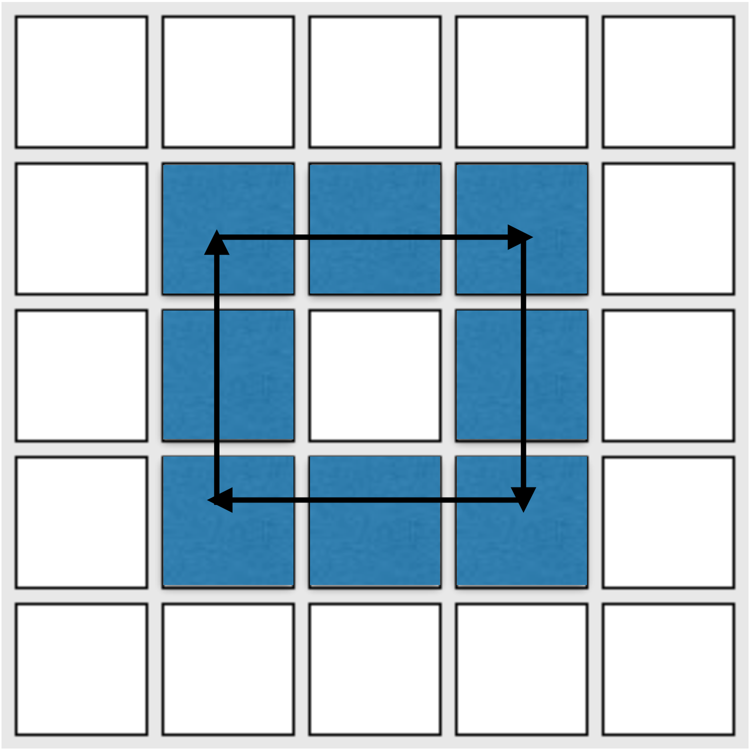
Same example:



Hidden Markov Model (HMM)

Same example:

But you don't observe location (dark room)

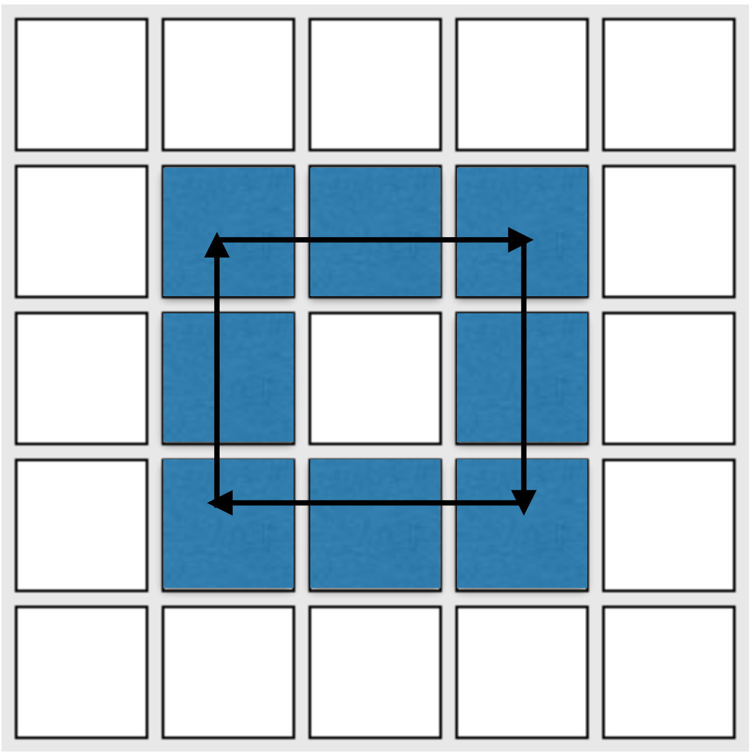


HIDDEN MARKOV MODEL (HMM)

Same example:

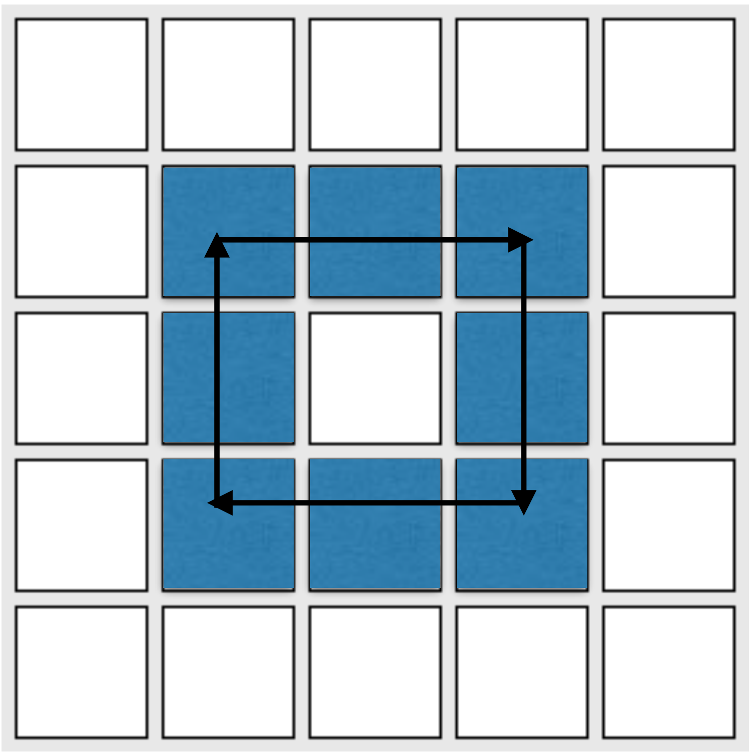
But you don't observe location (dark room)

You hear how close the bot is!



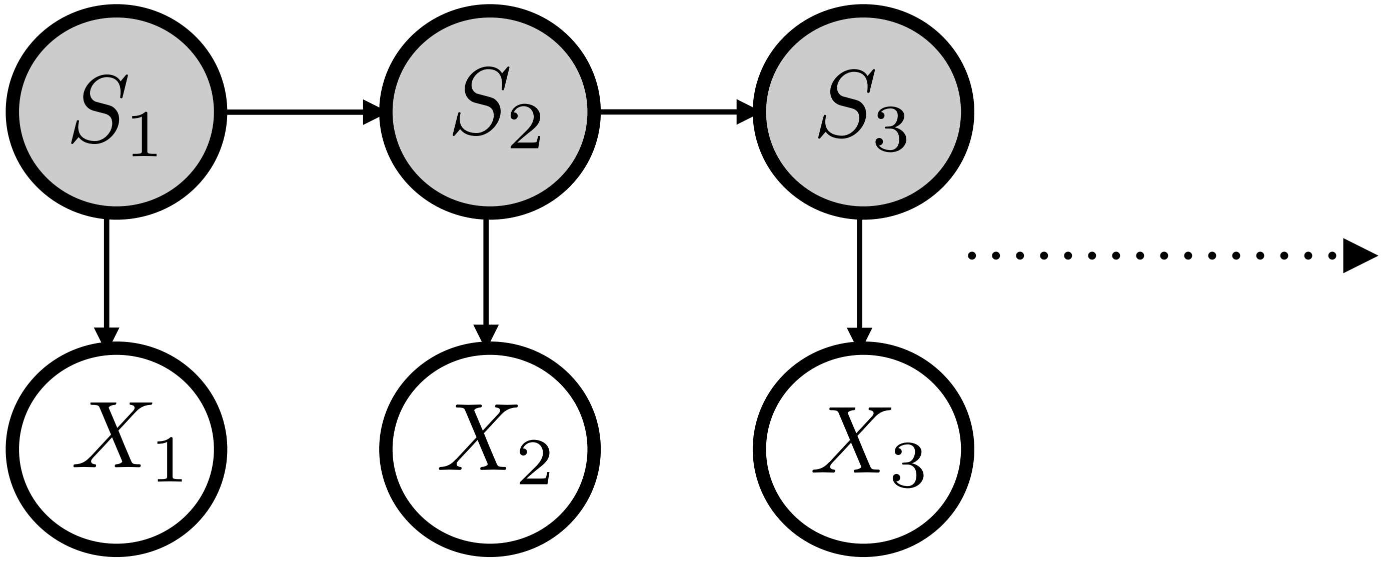
Hidden Markov Model (HMM)

Same example:



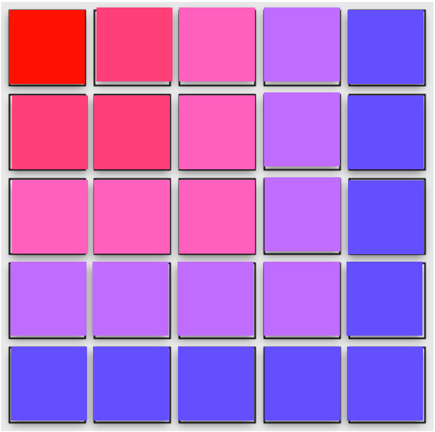
But you don't observe location (dark room)

You hear how close the bot is!

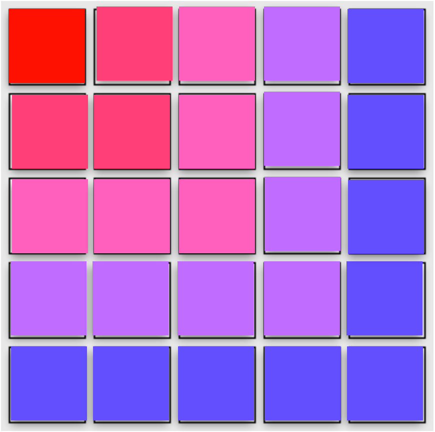


X_t's are loudness of what you hear

Hidden Markov Model (HMM)

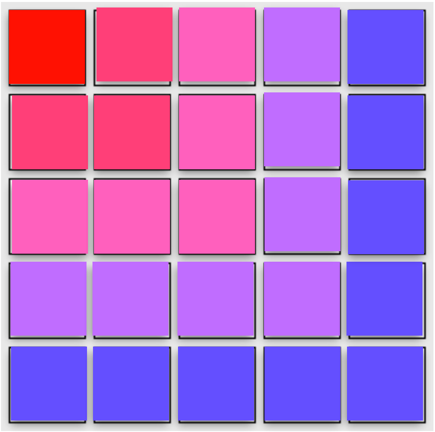


HIDDEN MARKOV MODEL (HMM)



- Both during the initial training/estimation phase, you never see the bot you only hear it

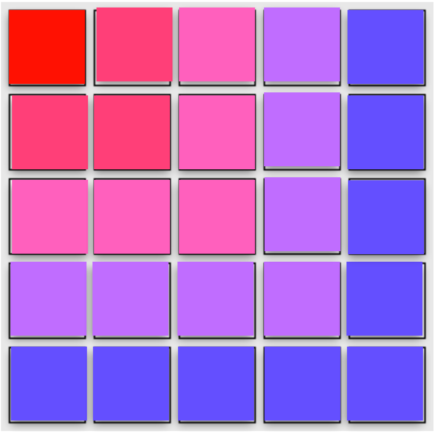
HIDDEN MARKOV MODEL (HMM)



- Both during the initial training/estimation phase, you never see the bot you only hear it

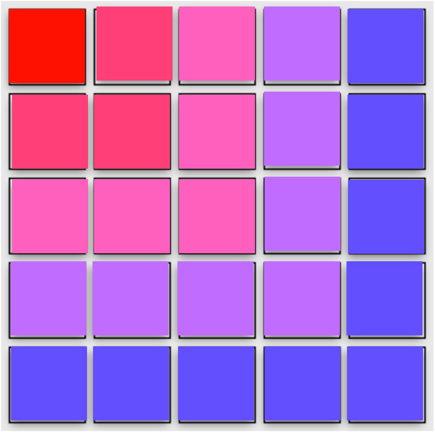
- But you hear it at any point in time

Hidden Markov Model (HMM)



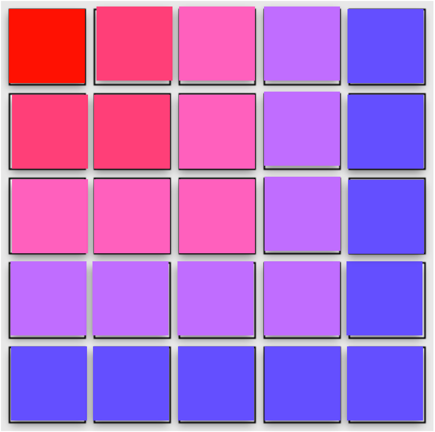
- Both during the initial training/estimation phase, you never see the bot you only hear it
- But you hear it at any point in time
- We will come back to learning next class.

HIDDEN MARKOV MODEL (HMM)



- Both during the initial training/estimation phase, you never see the bot you only hear it
- But you hear it at any point in time
- We will come back to learning next class.
- What is probability that bot will be in state k at time t given the entire sequence of observations?

Hidden Markov Model (HMM)



- Both during the initial training/estimation phase, you never see the bot you only hear it

- But you hear it at any point in time

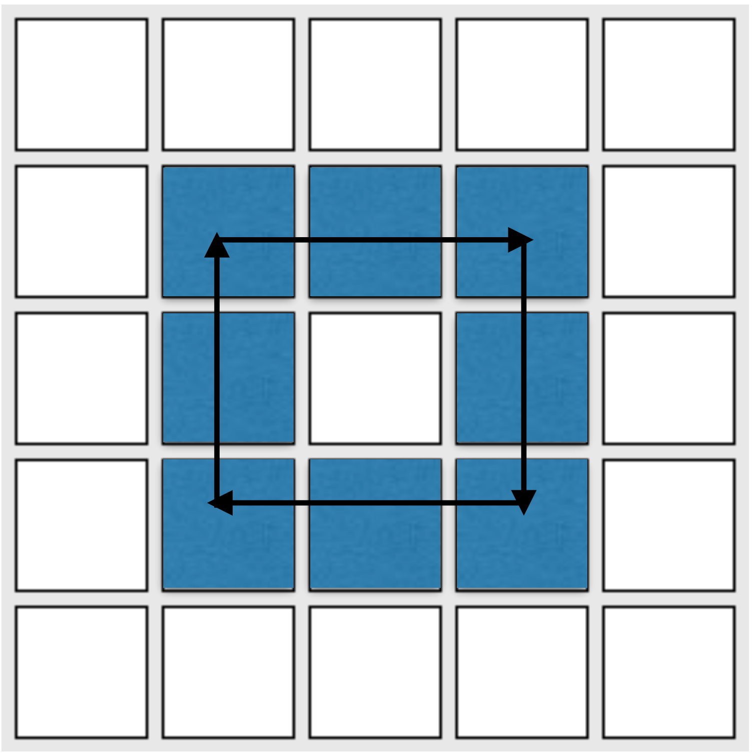
- We will come back to learning next class.

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

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

Hidden Markov Model (HMM)

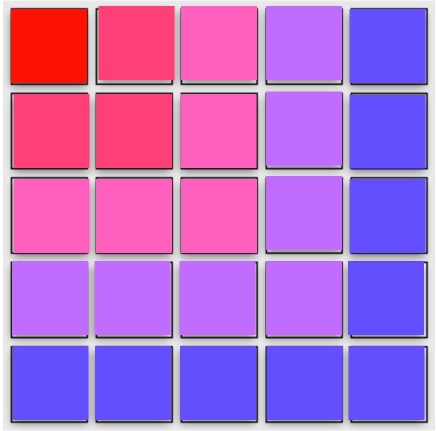
Same example:



But you don't observe location (dark room)

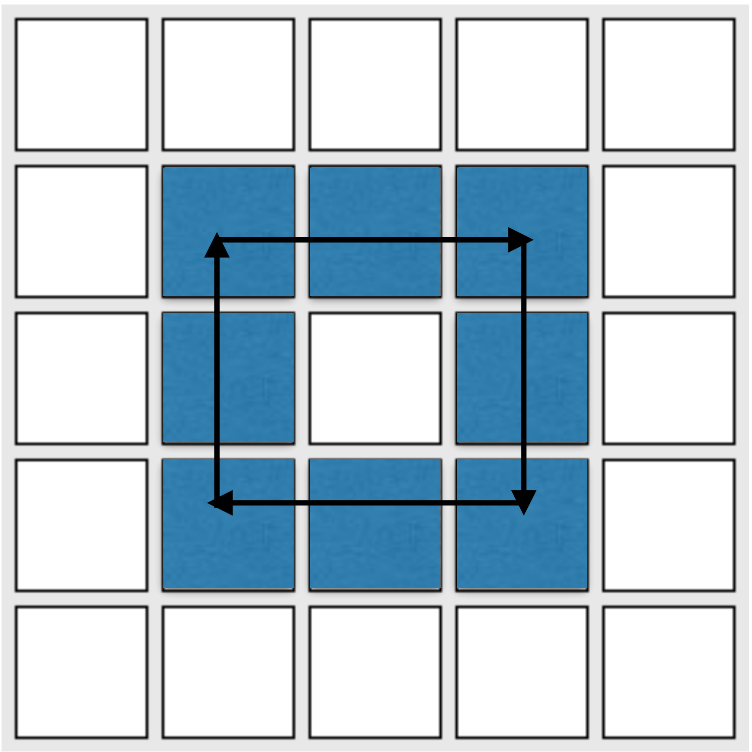
You hear how close the bot is!

What you hear:



Hidden Markov Model (HMM)

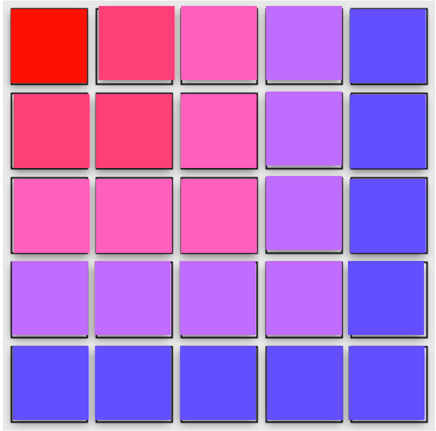
Same example:



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)

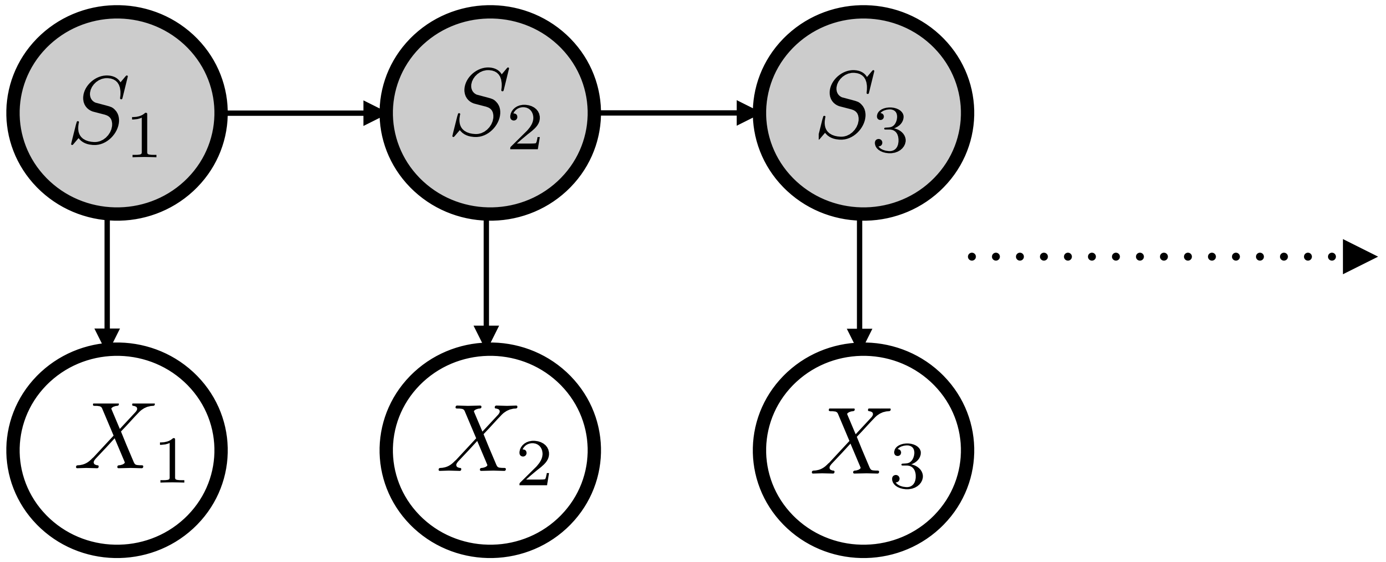
X_t's are what you hear (observation)

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)

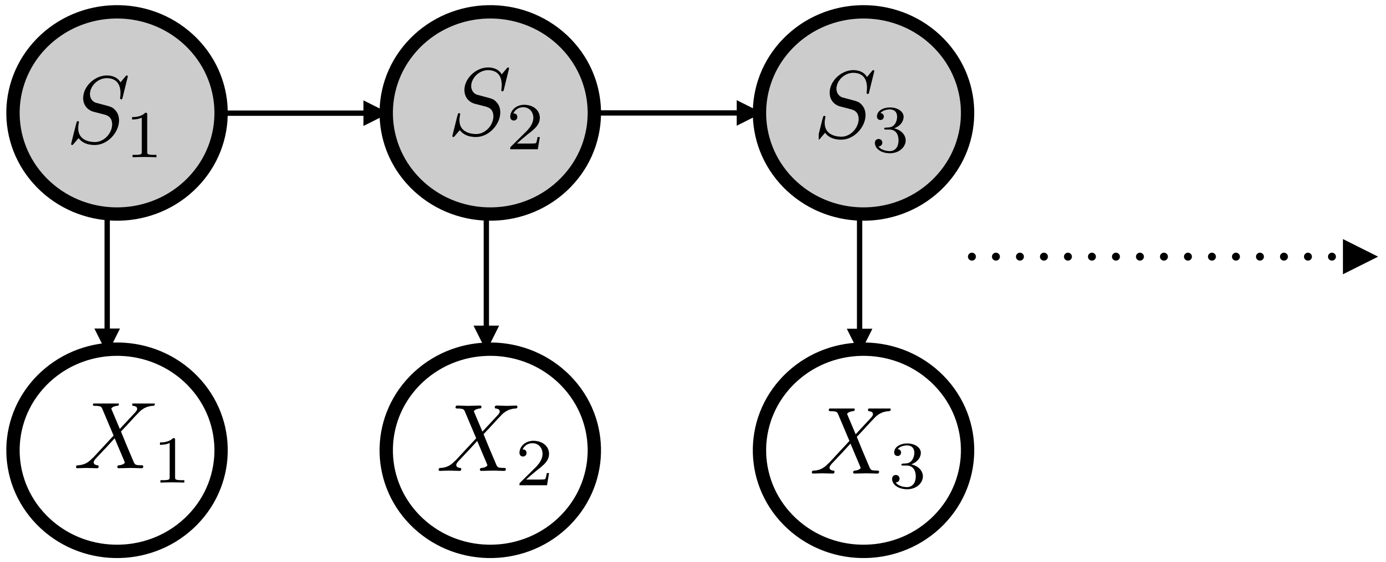
Eg: for n x n grid we have, $K = n^2$ states

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

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

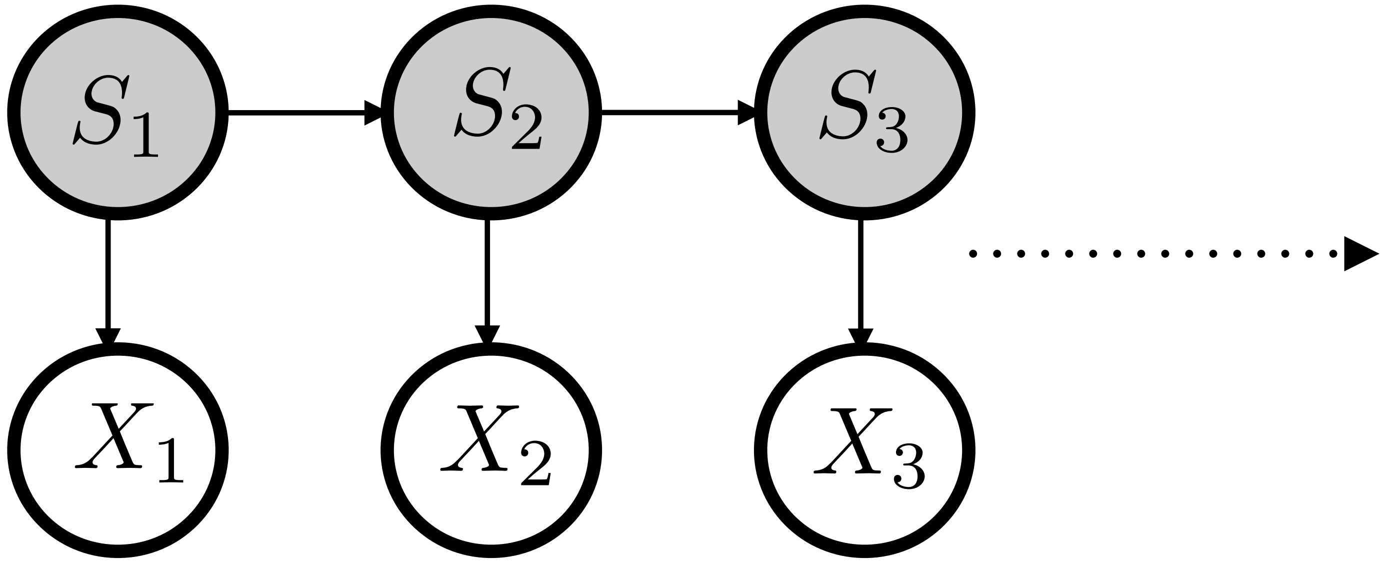
Inference in HMM

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

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

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

Inference in HMM

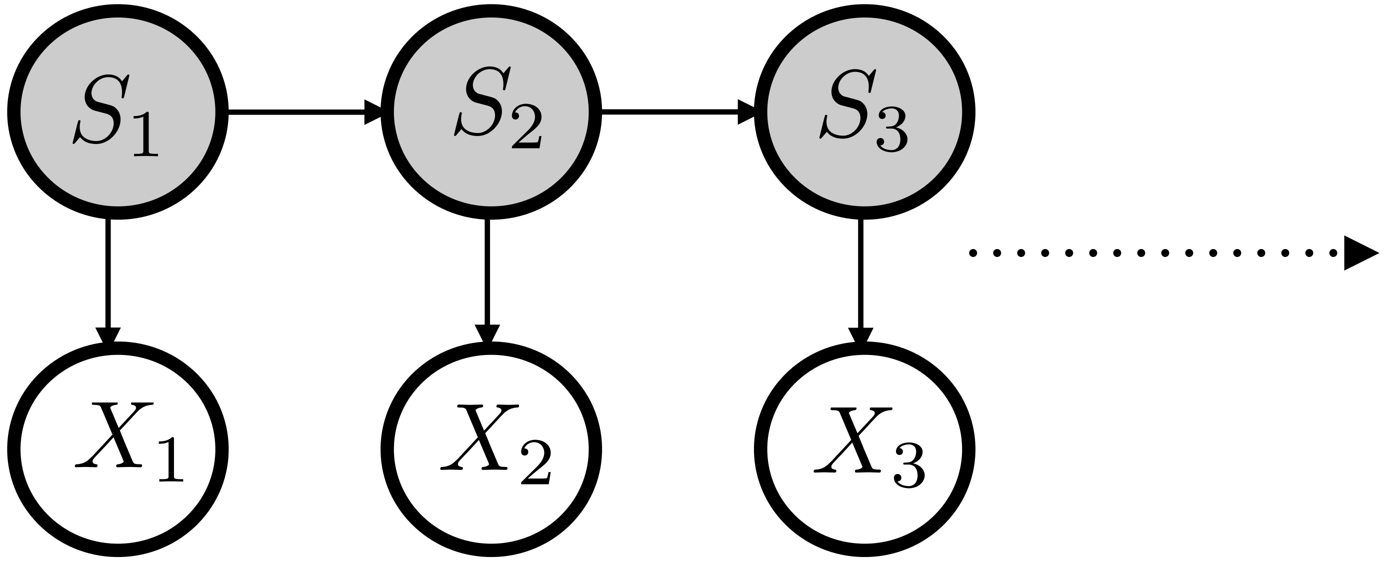


$$\text{message}_{S_{t-1}\mapsto S_t}(k) = P(S_t = k, X_1, \ldots, X_{t-1})$$

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

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

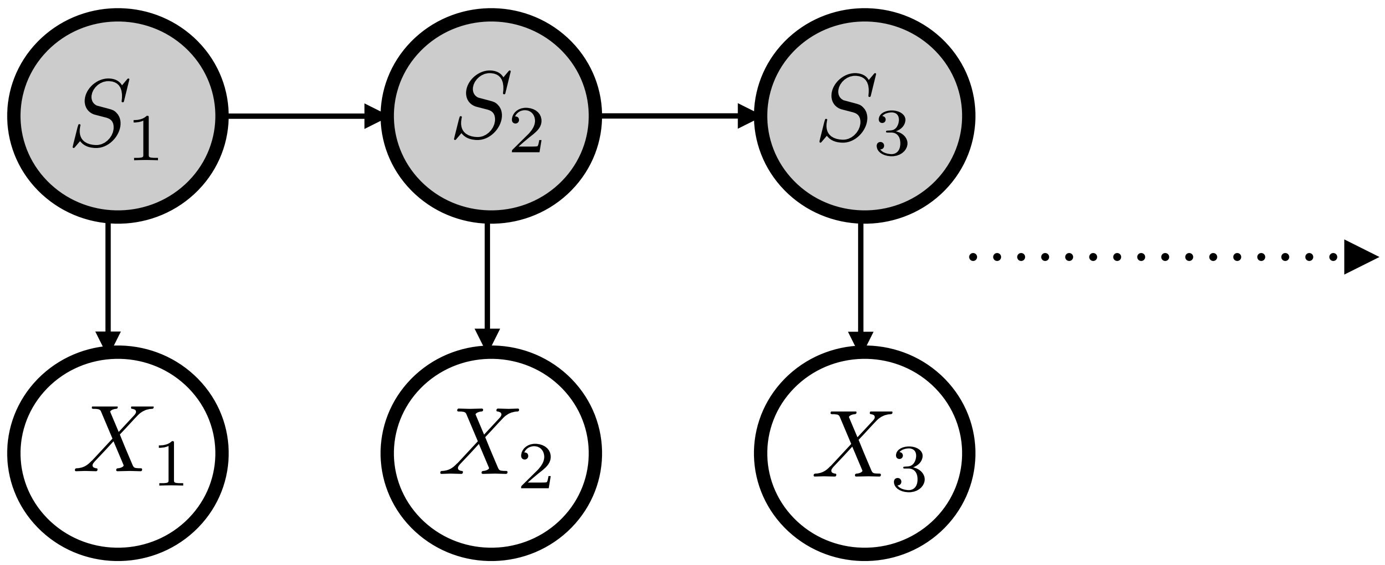
Inference in HMM



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

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

Inference in HMM



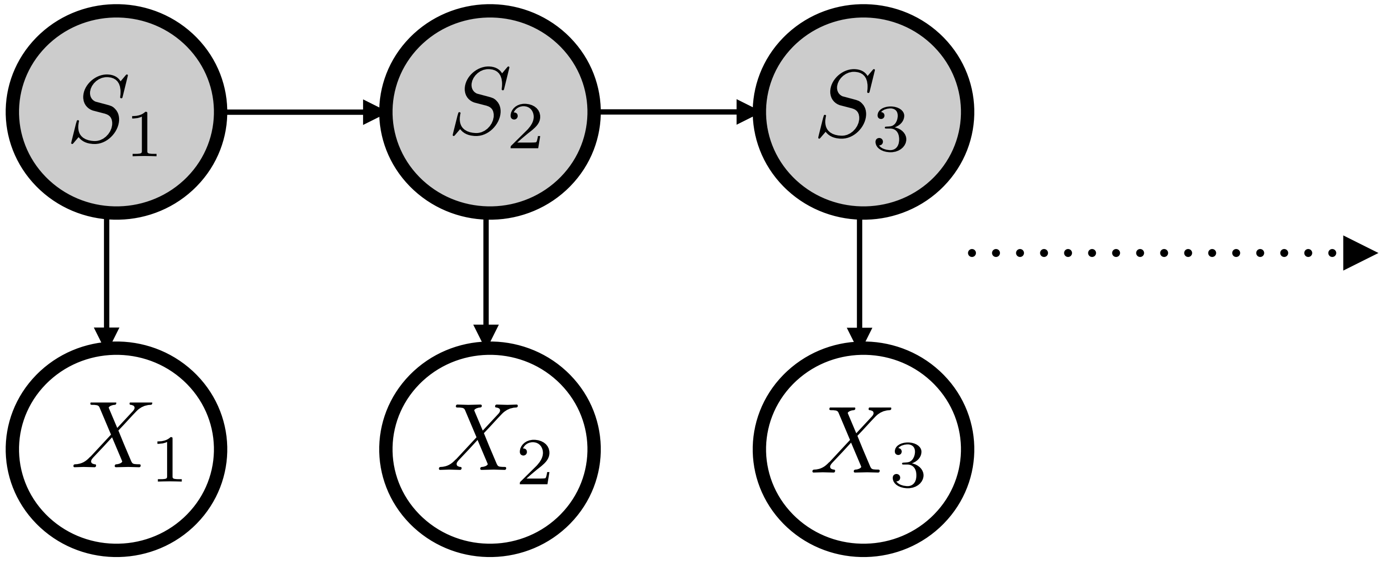
$$\text{message}_{S_{t-1} \mapsto S_t}(k) = P(S_t = k, X_1, \ldots, X_{t-1})$$

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

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



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

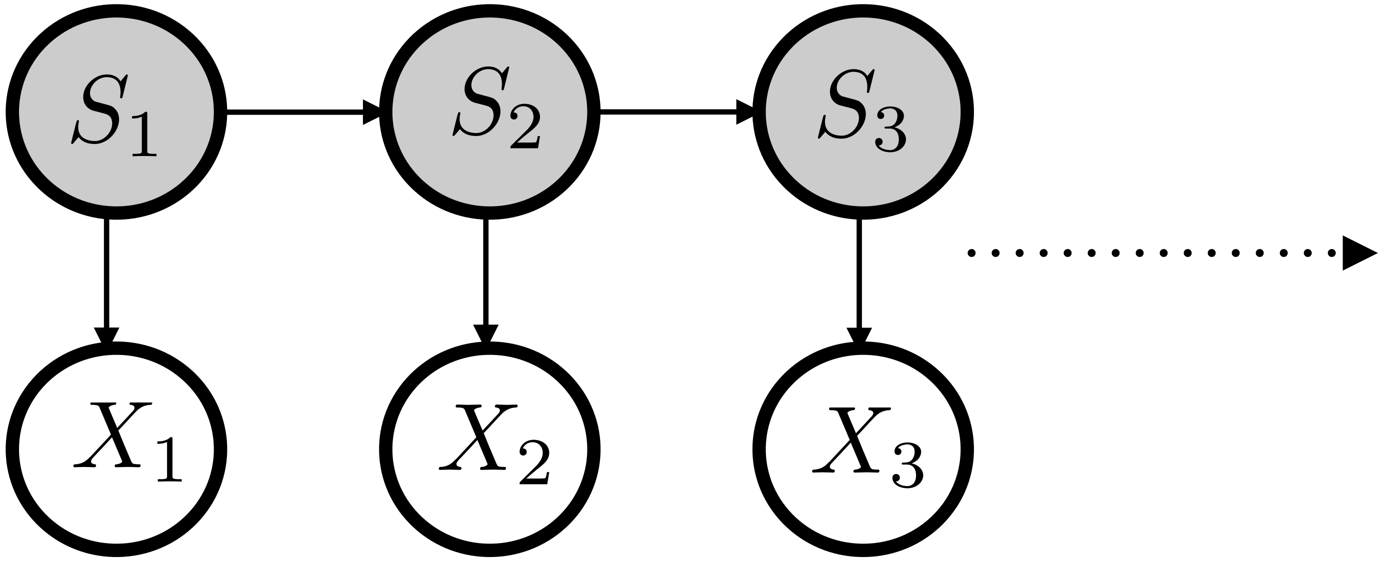
$$\text{message}_{S_{t+1}\mapsto S_t}(k) = P(X_n, \ldots, X_{t+1} | S_t = k)$$

Forward:

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

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

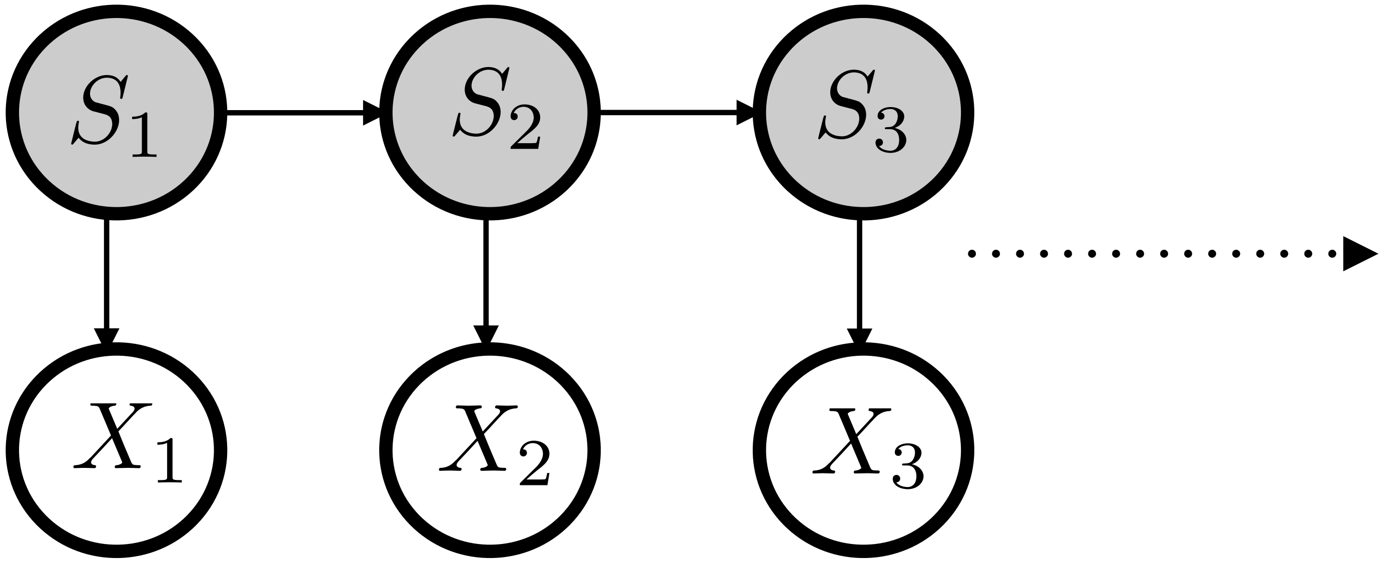
Inference in HMM



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

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

INFERENCE IN HMM



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

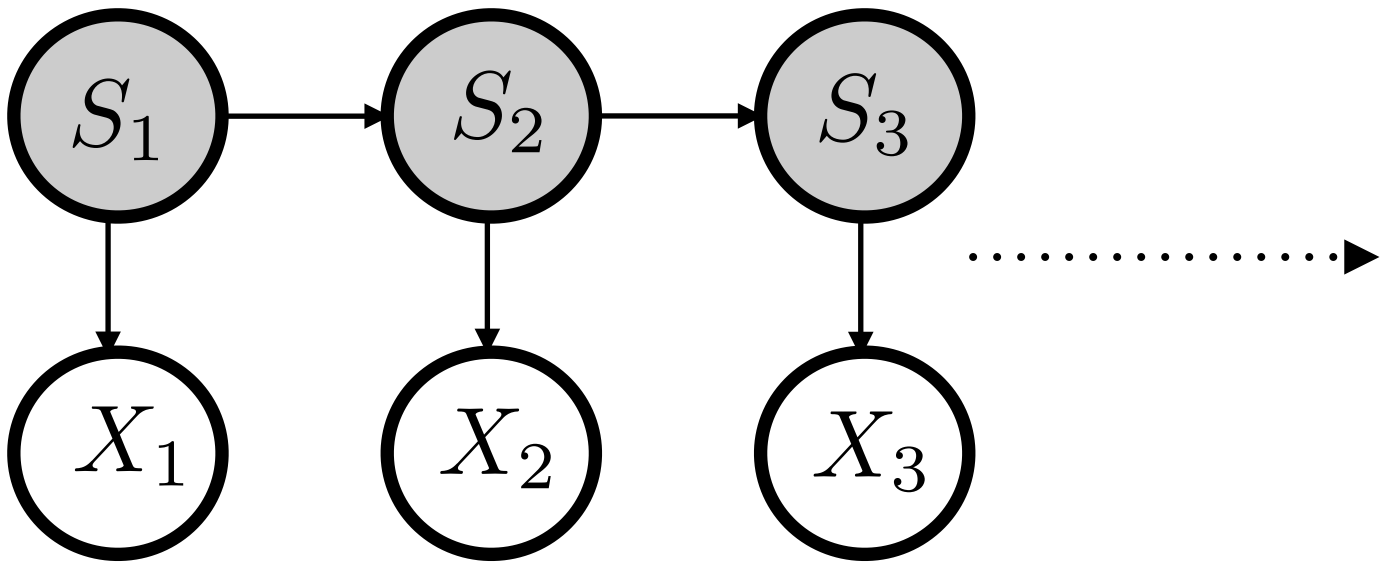
$$\text{message}_{S_{t+1} \mapsto S_t}(k) = P(X_n, \ldots, X_{t+1} | S_t = k)$$

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 . . .