
CIS 471/571 (Fall 2020): Introduction to Artificial Intelligence

Assignment Project Exam Help

Lecture 17: Hidden Markov Model

<https://tutorcs.com>

WeChat: cstutorcs

Thanh H. Nguyen

Source: <http://ai.berkeley.edu/home.html>



Reminder

- Homework 4: Bayes Nets
 - Deadline: Nov 24th, 2020

Assignment Project Exam Help

<https://tutorcs.com>

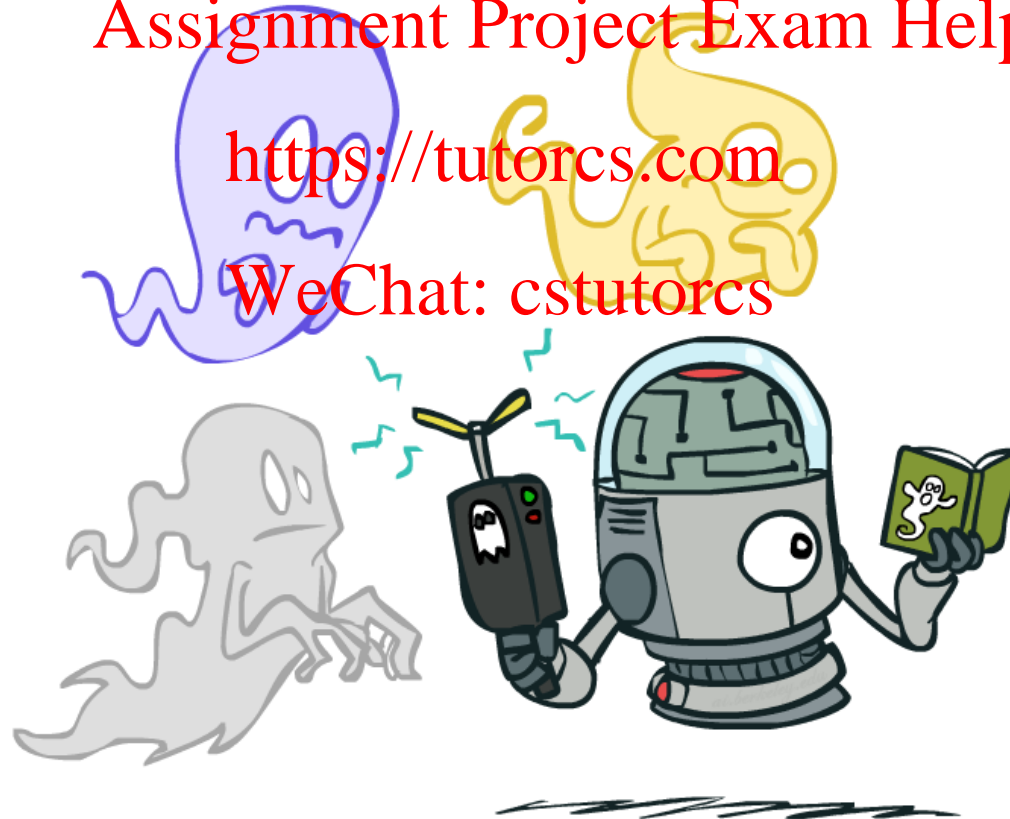
WeChat: cstutorcs

Hidden Markov Model

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



Reasoning over Time or Space

- Often, we want to reason about a sequence of observations
 - Speech recognition
 - Robot localization
 - User attention
 - Medical monitoring
- Need to introduce time (or space) into our models

Assignment Project Exam Help

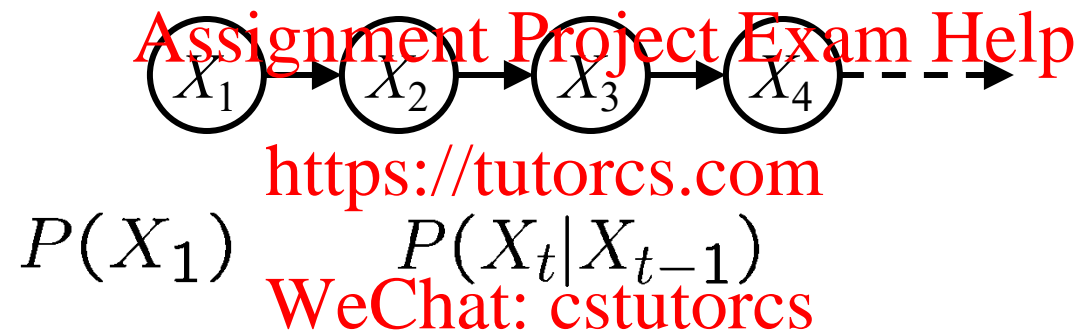
<https://tutorcs.com>

WeChat: cstutorcs



Markov Models

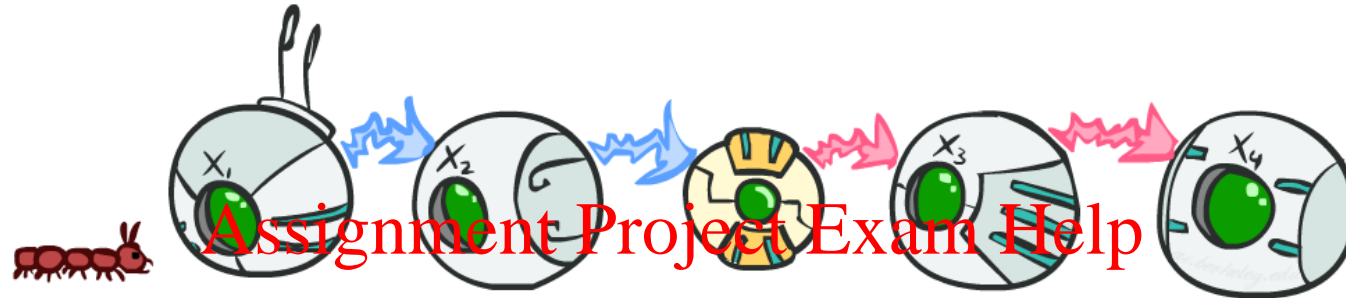
- Value of X at a given time is called the **state**



- Parameters: called **transition probabilities** or dynamics, specify how the state evolves over time (also, initial state probabilities)
- Stationarity assumption: transition probabilities the same at all times
- Same as MDP transition model, but no choice of action



Conditional Independence



- Basic conditional independence:
 - Past and future independent given the present
 - Each time step only depends on the previous
 - This is called the (first order) Markov property
- Note that the chain is just a (growable) BN
 - We can always use generic BN reasoning on it if we truncate the chain at a fixed length



Example Markov Chain: Weather

- States: $X = \{\text{rain}, \text{sun}\}$

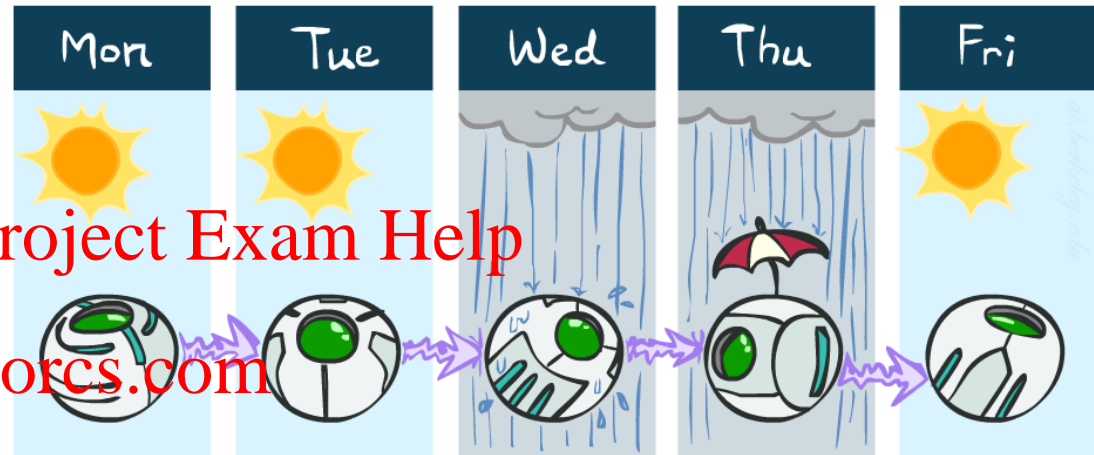
- Initial distribution: 1.0 sun

- CPT $P(X_t | X_{t-1})$:

X_{t-1}	X_t	$P(X_t X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

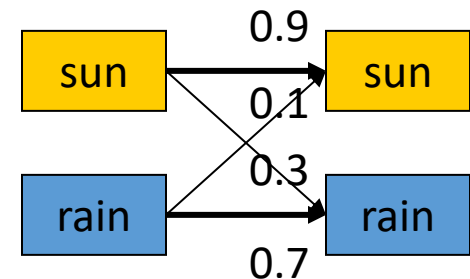
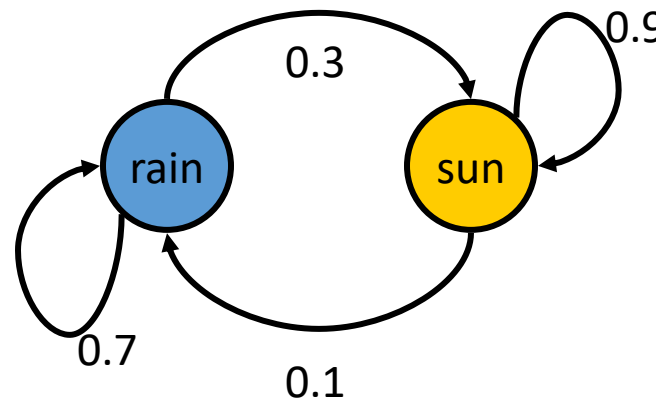
Assignment Project Exam Help

<https://tutorcs.com>



WeChat: cstutorcs

Two new ways of representing the same CPT



Example Markov Chain: Weather

- Initial distribution: 1.0 sun

Assignment Project Exam Help

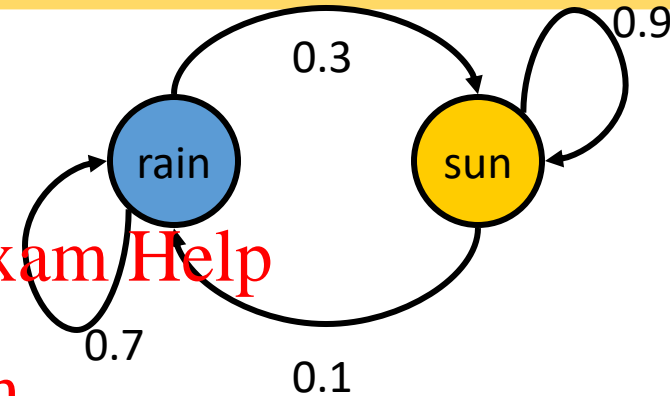
<https://tutorcs.com>

WeChat: cstutorcs

- What is the probability distribution after one step?

$$P(X_2 = \text{sun}) = P(X_2 = \text{sun} | X_1 = \text{sun})P(X_1 = \text{sun}) + P(X_2 = \text{sun} | X_1 = \text{rain})P(X_1 = \text{rain})$$

$$0.9 \cdot 1.0 + 0.3 \cdot 0.0 = 0.9$$



Mini-Forward Algorithm

- Question: What's $P(X)$ on some day t ?



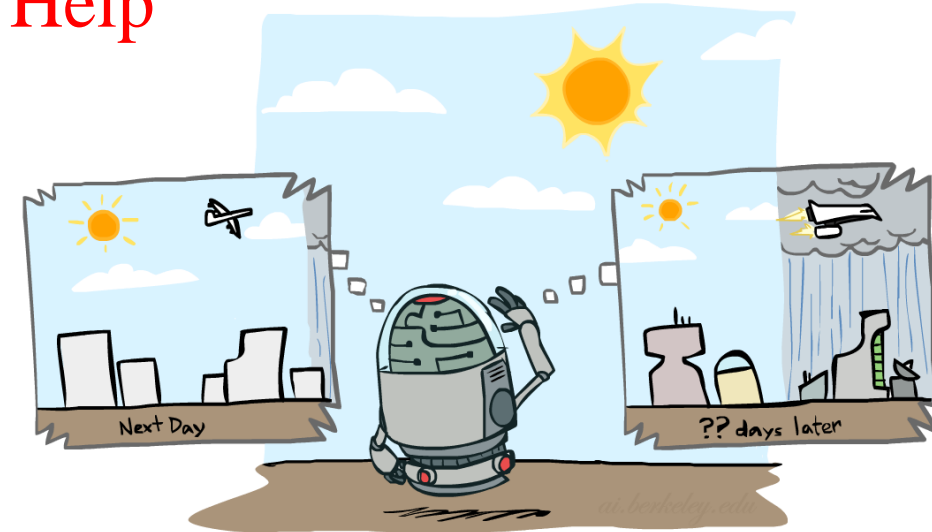
<https://tutorcs.com>

WeChat: cstutorcs

$P(x_1) = \text{known}$

$$\begin{aligned} P(x_t) &= \sum_{x_{t-1}} P(x_{t-1}, x_t) \\ &= \sum_{x_{t-1}} P(x_t \mid x_{t-1}) P(x_{t-1}) \end{aligned}$$

Forward simulation



Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\begin{array}{ccccc}
 \left\langle \begin{array}{c} 1.0 \\ 0.0 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.9 \\ 0.1 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.84 \\ 0.16 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.804 \\ 0.196 \end{array} \right\rangle & \Rightarrow & \left\langle \begin{array}{c} 0.75 \\ 0.25 \end{array} \right\rangle \\
 P(X_1) & P(X_2) & P(X_3) & P(X_4) & & P(X_\infty)
 \end{array}$$

- From initial observation of rain

$$\begin{array}{ccccc}
 \left\langle \begin{array}{c} 0.0 \\ 1.0 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.3 \\ 0.7 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.48 \\ 0.52 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.588 \\ 0.412 \end{array} \right\rangle & \Rightarrow & \left\langle \begin{array}{c} 0.75 \\ 0.25 \end{array} \right\rangle \\
 P(X_1) & P(X_2) & P(X_3) & P(X_4) & & P(X_\infty)
 \end{array}$$

- From yet another initial distribution $P(X_1)$:

$$\begin{array}{ccc}
 \left\langle \begin{array}{c} p \\ 1-p \end{array} \right\rangle & \dots & \Rightarrow \left\langle \begin{array}{c} 0.75 \\ 0.25 \end{array} \right\rangle \\
 P(X_1) & & P(X_\infty)
 \end{array}$$



Stationary Distributions

- For most chains:

- Influence of the initial distribution gets less and less over time
- The distribution we end up in is independent of the initial distribution

- Stationary distribution:

- The distribution we end up with is called the stationary distribution P_∞ of the chain

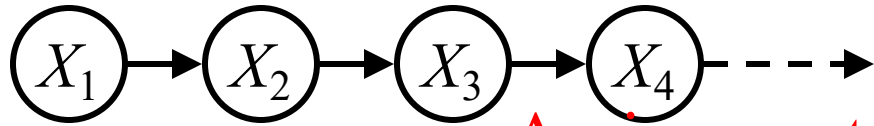
- It satisfies

$$P_\infty(X) = P_{\infty+1}(X) = \sum_x P(X|x)P_\infty(x)$$



Example: Stationary Distributions

- Question: What's $P(X)$ at time $t = \text{infinity}$?



Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

$$P_{\infty}(\text{sun}) = P(\text{sun}|\text{sun})P_{\infty}(\text{sun}) + P(\text{sun}|\text{rain})P_{\infty}(\text{rain})$$

$$P_{\infty}(\text{rain}) = P(\text{rain}|\text{sun})P_{\infty}(\text{sun}) + P(\text{rain}|\text{rain})P_{\infty}(\text{rain})$$

$$P_{\infty}(\text{sun}) = 0.9P_{\infty}(\text{sun}) + 0.3P_{\infty}(\text{rain})$$

$$P_{\infty}(\text{rain}) = 0.1P_{\infty}(\text{sun}) + 0.7P_{\infty}(\text{rain})$$

$$P_{\infty}(\text{sun}) = 3P_{\infty}(\text{rain})$$

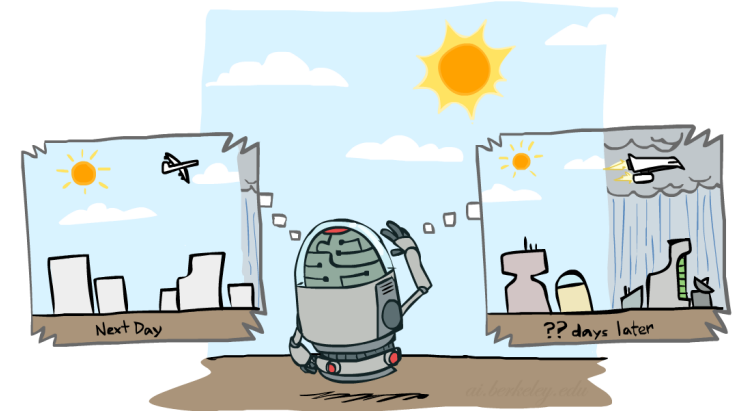
$$P_{\infty}(\text{rain}) = 1/3P_{\infty}(\text{sun})$$

Also: $P_{\infty}(\text{sun}) + P_{\infty}(\text{rain}) = 1$



$$P_{\infty}(\text{sun}) = 3/4$$

$$P_{\infty}(\text{rain}) = 1/4$$



X_{t-1}	X_t	$P(X_t X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7



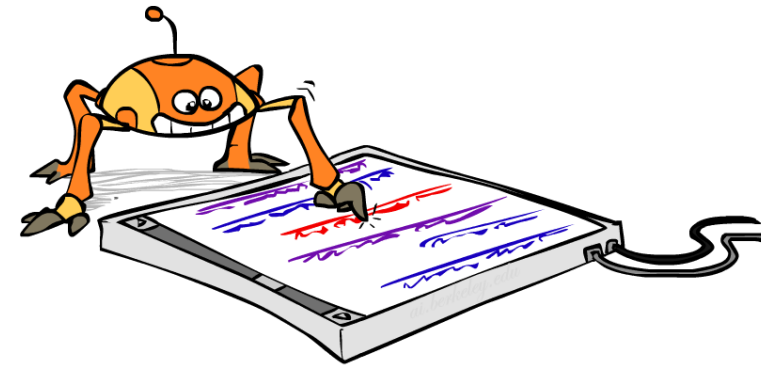
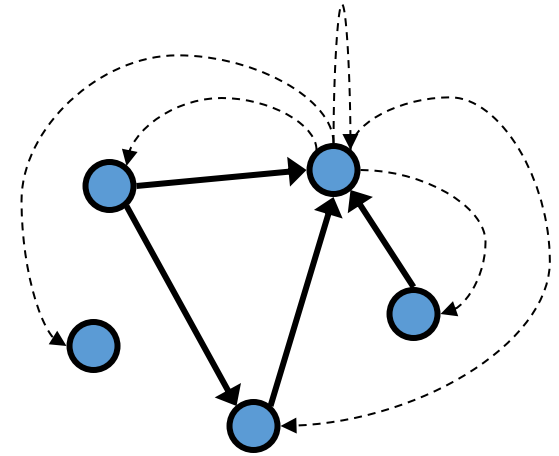
Application of Stationary Distribution: Web Link Analysis

- PageRank over a web graph
 - Each web page is a state
 - Initial distribution: uniform over pages
 - Transitions:
 - With prob. c , uniform jump to a random page (dotted lines, not all shown)
 - With prob. $1-c$, follow a random outlink (solid lines)
- Stationary distribution
 - Will spend more time on highly reachable pages
 - E.g. many ways to get to the Acrobat Reader download page
 - Somewhat robust to link spam
 - Google 1.0 returned the set of pages containing all your keywords in decreasing rank, now all search engines use link analysis along with many other factors

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



Application of Stationary Distributions: Gibbs Sampling*

- Each joint instantiation over all hidden and query variables is a state: $\{X_1, \dots, X_n\} = H \cup Q$

Assignment Project Exam Help

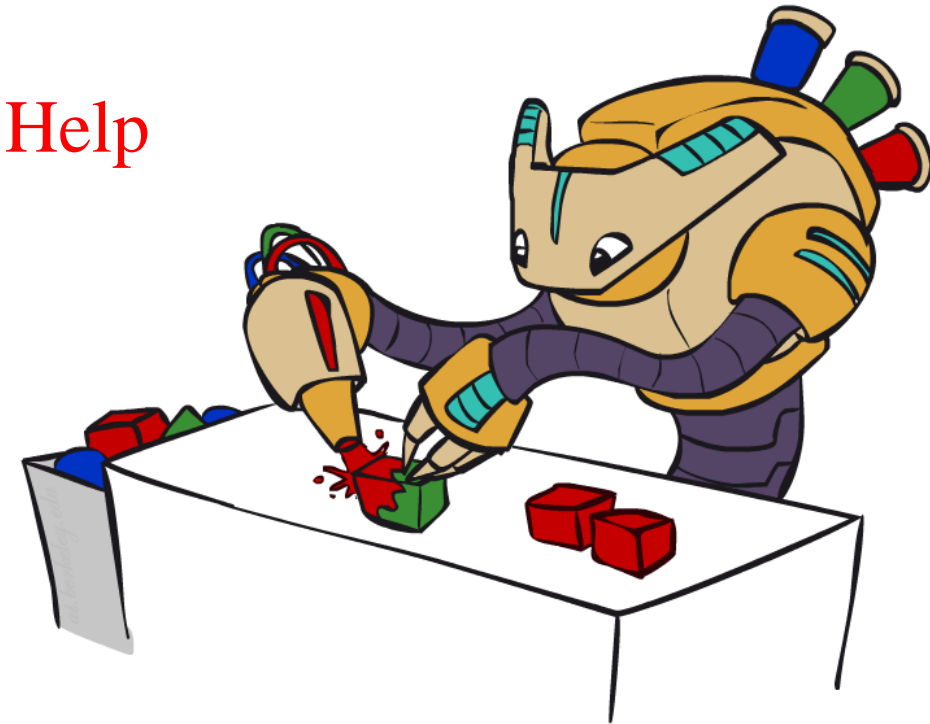
- Transitions:

- With probability $1/n$ resample variable X_i according to

$$P(X_i \mid x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n, e_1, \dots, e_m)$$

- Stationary distribution:

- Conditional distribution $P(X_1, X_2, \dots, X_n \mid e_1, \dots, e_m)$
 - Means that when running Gibbs sampling long enough we get a sample from the desired distribution
 - Requires some proof to show this is true!



Hidden Markov Models

Assignment Project Exam Help

<https://tutorcs.com>

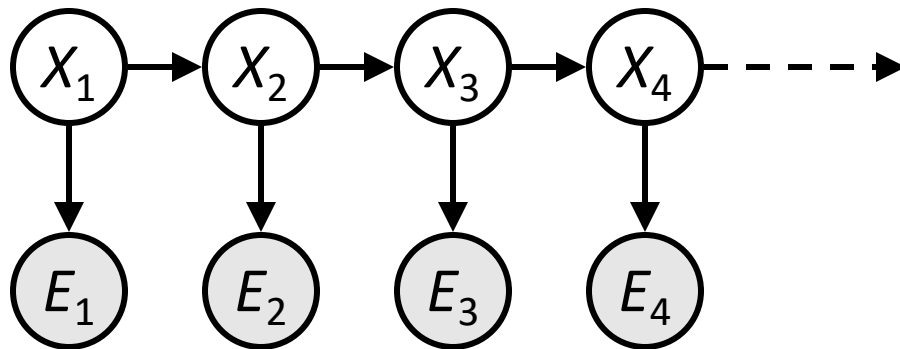
WeChat: cstutorcs



Hidden Markov Models

- Markov chains not so useful for most agent
 - Need observations to update your beliefs

- Hidden Markov models (HMMs)
 - Underlying Markov chain over states X
 - You observe outputs (effects) at each time step



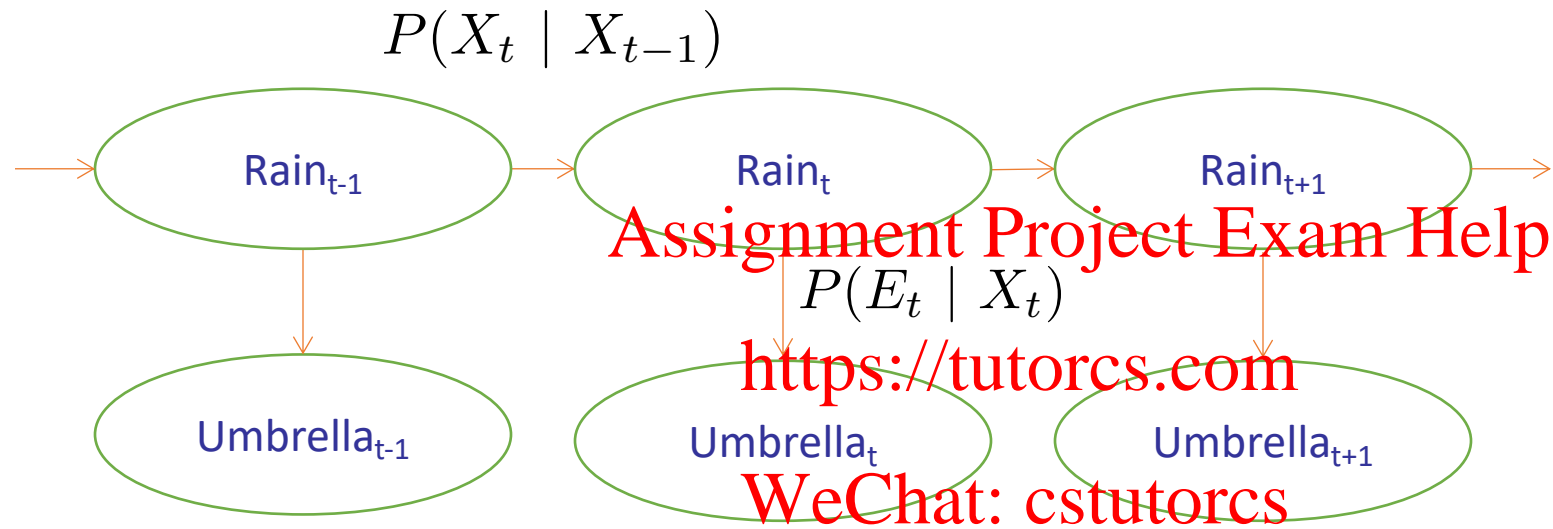
Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



Example: Weather HMM



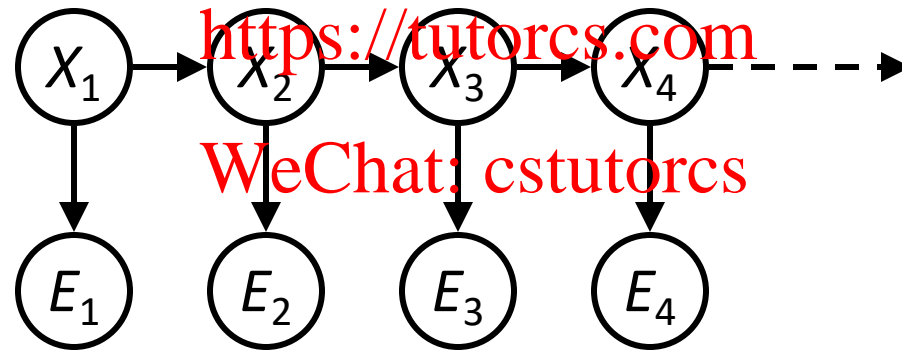
- An HMM is defined by:
 - Initial distribution: $P(X_1)$
 - Transitions: $P(X_t | X_{t-1})$
 - Emissions: $P(E_t | X_t)$

R_{t-1}	R_t	$P(R_t R_{t-1})$
+r	+r	0.7
+r	-r	0.3
-r	+r	0.3
-r	-r	0.7

R_t	U_t	$P(U_t R_t)$
+r	+u	0.9
+r	-u	0.1
-r	+u	0.2
-r	-u	0.8

Conditional Independence

- HMMs have two important independence properties:
 - Markov hidden process: future depends on past via the present
 - Current observation independent of all else given current state



- Quiz: does this mean that evidence variables are guaranteed to be independent?
 - [No, they tend to be correlated by the hidden state]



Real HMM Examples

- Speech recognition HMMs:
 - Observations are acoustic signals (continuous valued)
 - States are specific positions in specific words (so tens of thousands)
- Machine translation HMMs:
 - Observations are words (tens of thousands)
 - States are translation options
- Robot tracking:
 - Observations are range readings (continuous)
 - States are positions on a map (continuous)

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



Filtering / Monitoring

- Filtering, or monitoring, is the task of tracking the distribution $B_t(X) = P_t(X_t \mid e_1, \dots, e_t)$ (the belief state) over time

Assignment Project Exam Help

- We start with $B_1(X)$ in an initial setting, usually uniform

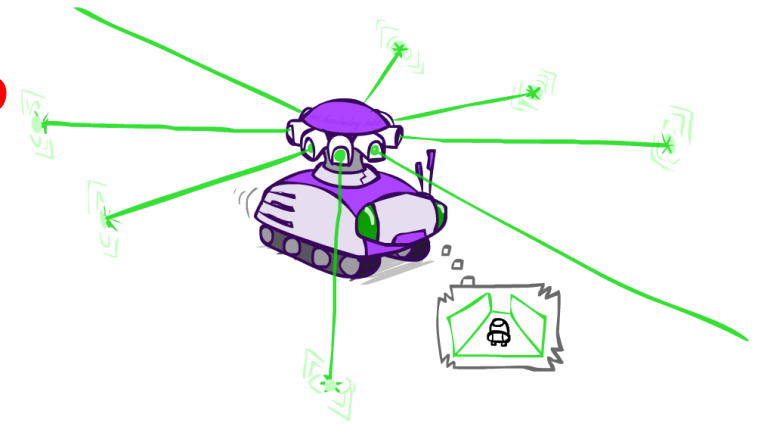
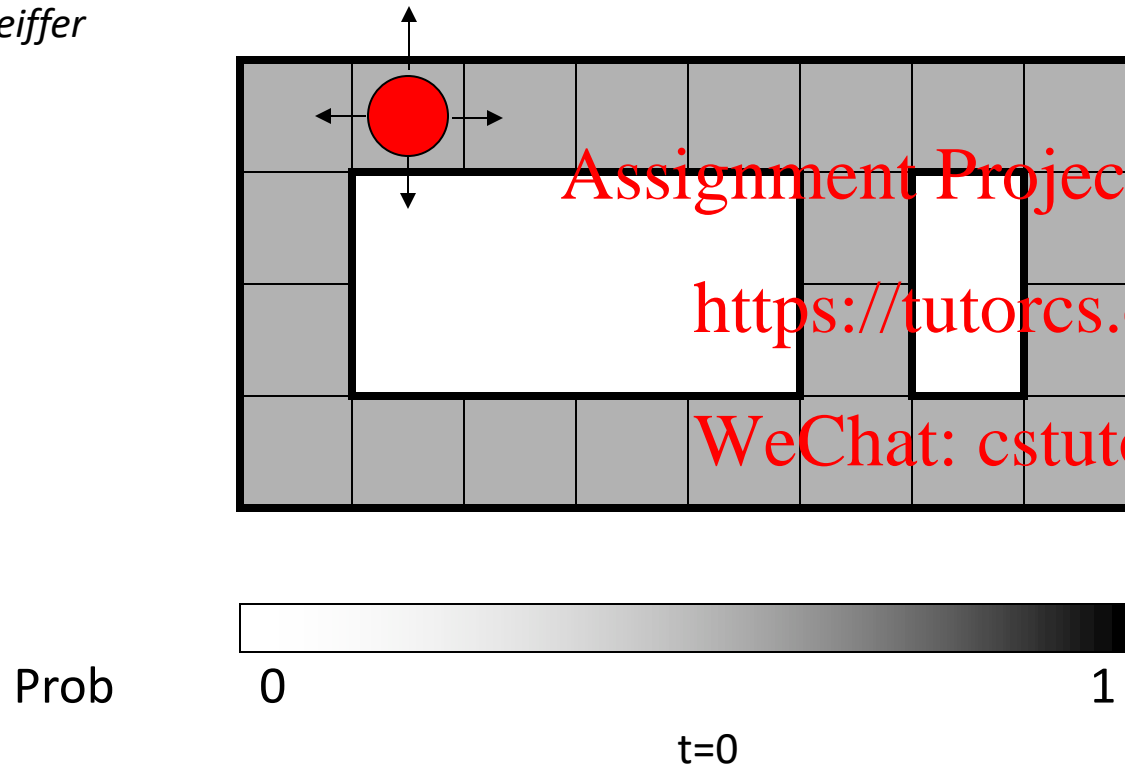
WeChat: cstutorcs

- As time passes, or we get observations, we update $B(X)$



Example: Robot Localization

*Example from
Michael Pfeiffer*

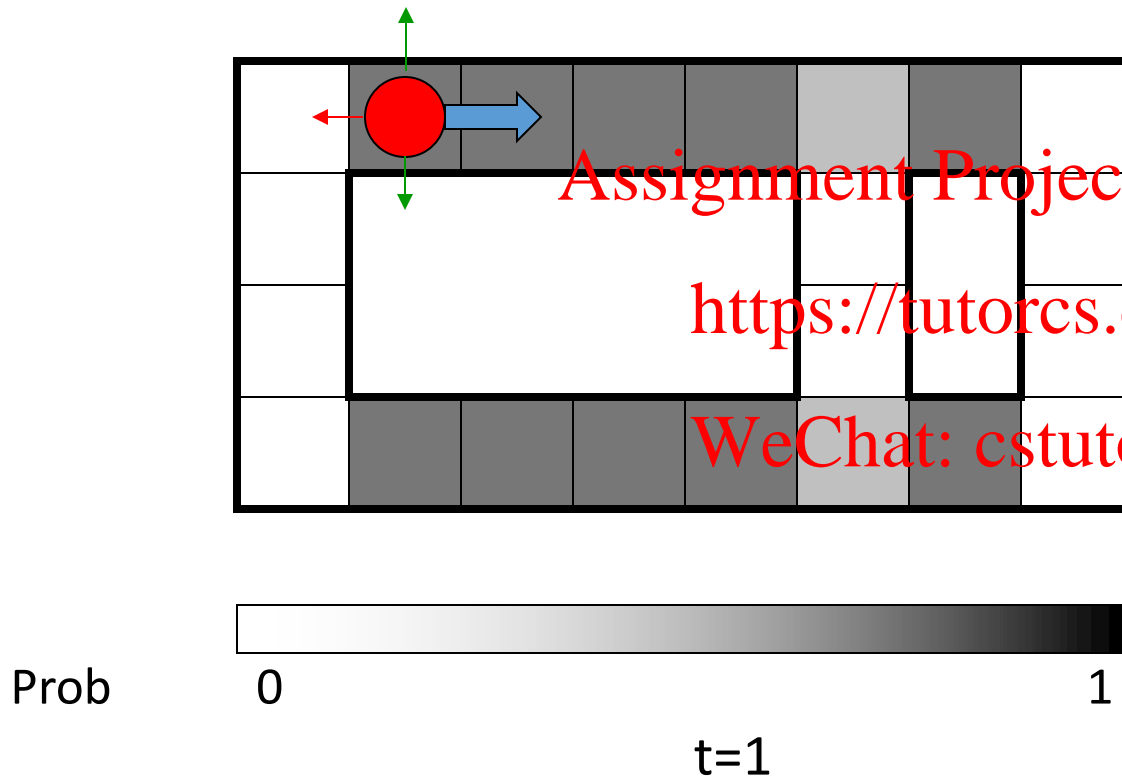


Sensor model: can read in which directions there is a wall, never more than 1 mistake

Motion model: may not execute action with small prob.



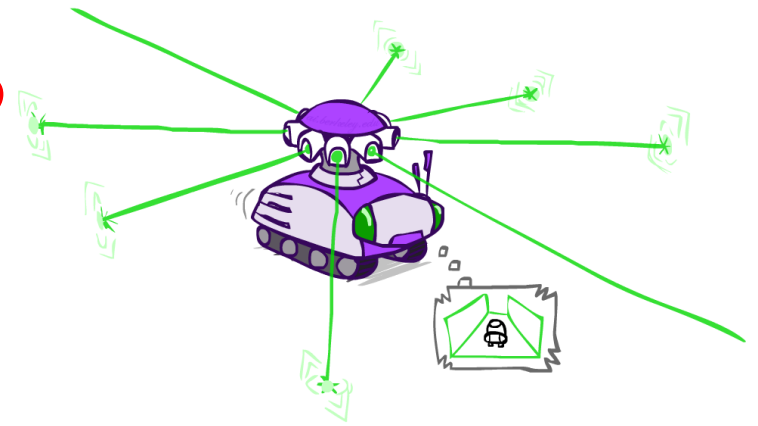
Example: Robot Localization



Assignment Project Exam Help

<https://tutorcs.com>

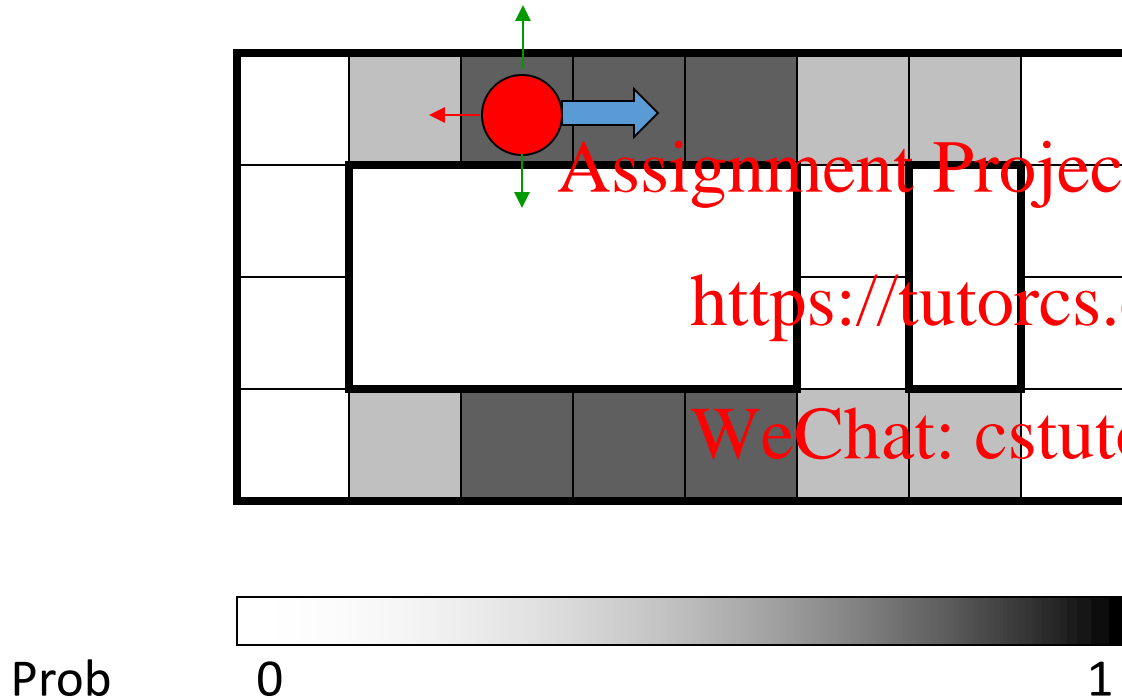
WeChat: cstutorcs



Lighter grey: was possible to get the reading, but less likely b/c required 1 mistake



Example: Robot Localization

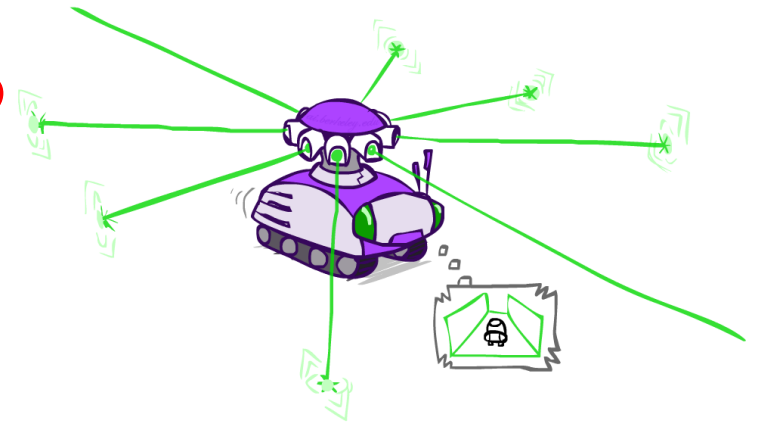


$t=2$

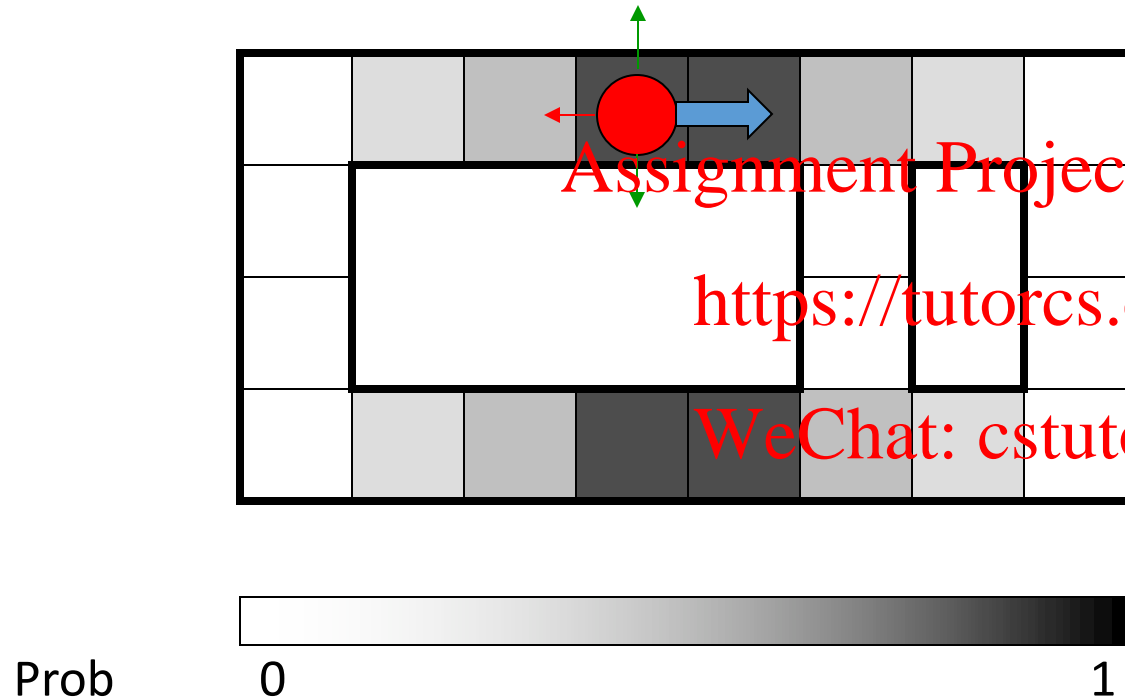
Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



Example: Robot Localization

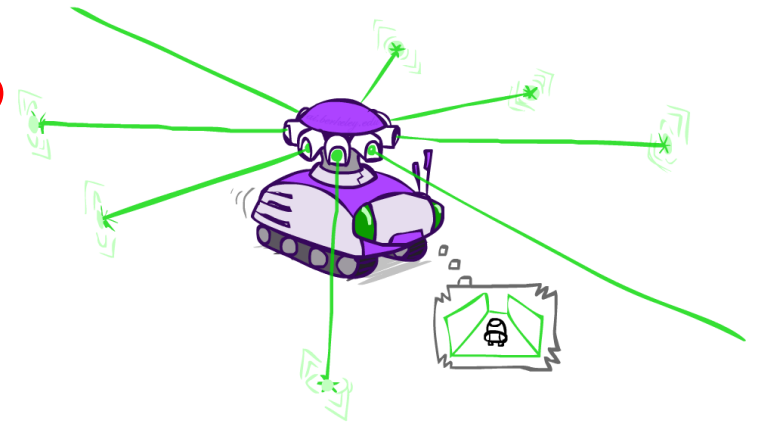


$t=3$

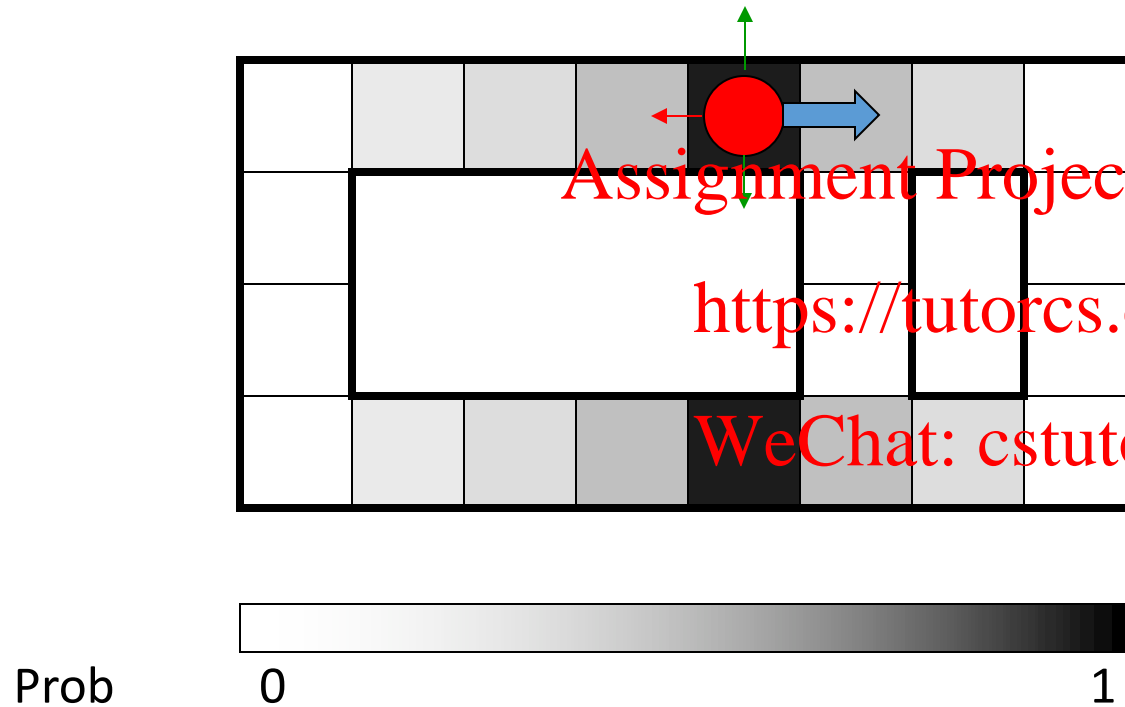
Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



Example: Robot Localization

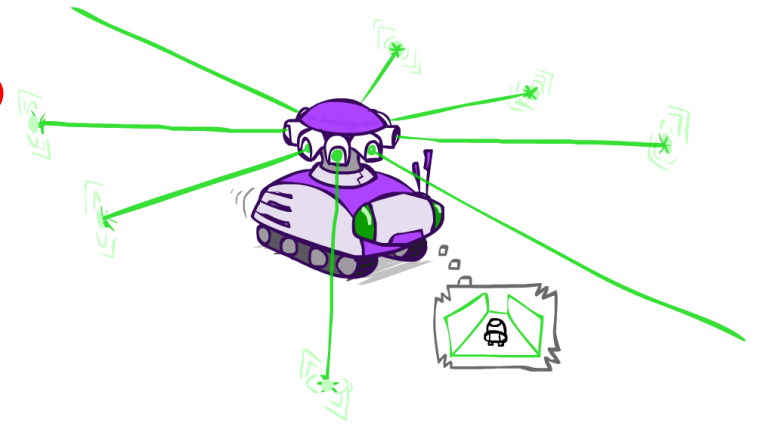


$t=4$

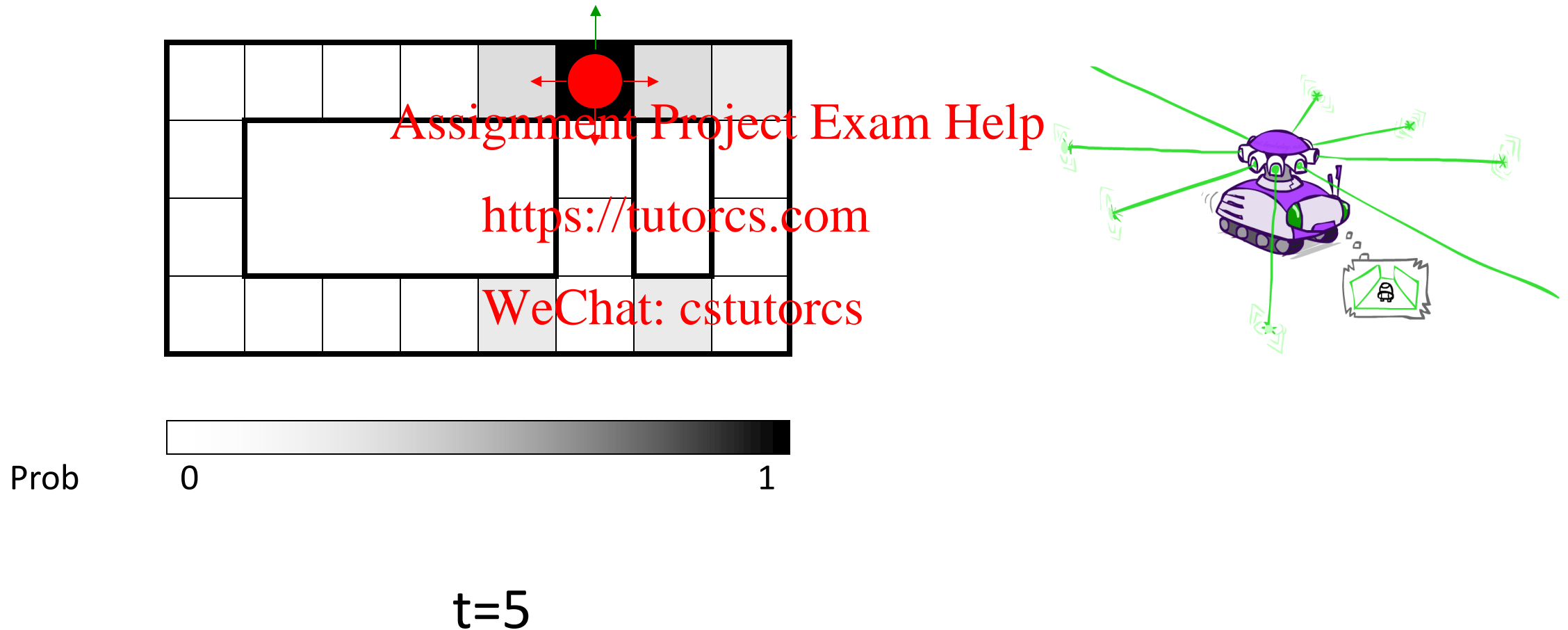
Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



Example: Robot Localization



The Forward Algorithm

- We are given evidence at each time and want to know

$$B_t(X) = P(X_t|e_{1:t})$$

Assignment Project Exam Help

- Induction: assuming we have current belief $B(X_t) = P(X_t|e_{1:t})$

$$P(X_{t+1}|e_{1:(t+1)}) \leftarrow P(X_{t+1}|e_{1:t}) \leftarrow P(X_t|e_{1:t})$$

Observation
update

Passage of time
update



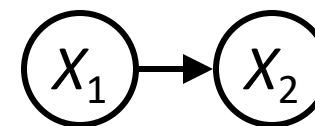
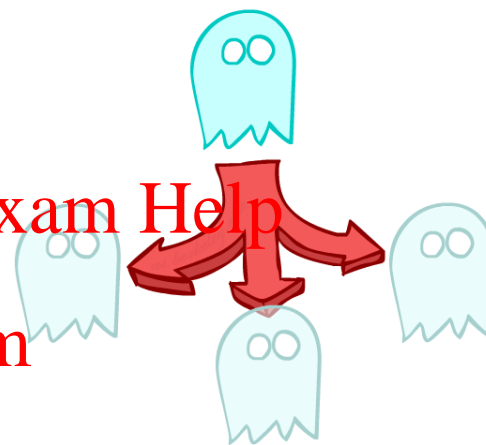
Inference: Base Cases



Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs



$$P(X_1|e_1)$$

$$\begin{aligned} P(x_1|e_1) &= P(x_1, e_1)/P(e_1) \\ &\propto_{X_1} P(x_1, e_1) \\ &= P(x_1)P(e_1|x_1) \end{aligned}$$

$$P(X_2)$$

$$\begin{aligned} P(x_2) &= \sum_{x_1} P(x_1, x_2) \\ &= \sum_{x_1} P(x_1)P(x_2|x_1) \end{aligned}$$



Passage of Time

- Assume we have current belief $P(X \mid \text{evidence to date})$

$$B(X_t) = P(X_t | e_{1:t})$$

- Then, after one time step passes:

Assignment Project Exam Help

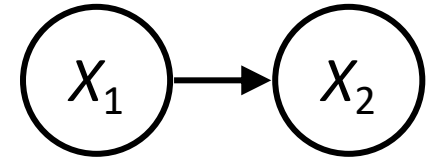
$$P(X_{t+1} | e_{1:t}) = \sum_{x_t} P(X_{t+1} | x_t, e_{1:t}) P(x_t | e_{1:t})$$

$$= \sum_{x_t} P(X_{t+1} | x_t, e_{1:t}) P(x_t | e_{1:t})$$

$$= \sum_{x_t} P(X_{t+1} | x_t) P(x_t | e_{1:t})$$

- Basic idea: beliefs get “pushed” through the transitions

- With the “B” notation, we have to be careful about what time step t the belief is about, and what evidence it includes



- Or compactly:

$$B'(X_{t+1}) = \sum_{x_t} P(X' | x_t) B(x_t)$$



Observation

- Assume we have current belief $P(X \mid \text{previous evidence})$:

$$B'(X_{t+1}) = P(X_{t+1} | e_{1:t})$$

- Then, after evidence comes in:

$$P(X_{t+1} | e_{1:t+1}) = P(X_{t+1}, e_{t+1} | e_{1:t}) / P(e_{t+1} | e_{1:t})$$

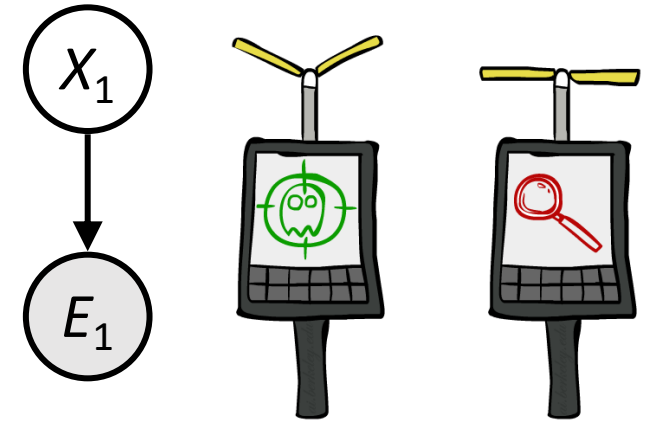
$$\propto_{X_{t+1}} P(X_{t+1}, e_{t+1} | e_{1:t})$$

$$= P(e_{t+1} | e_{1:t}, X_{t+1}) P(X_{t+1} | e_{1:t})$$

$$= P(e_{t+1} | X_{t+1}) P(X_{t+1} | e_{1:t})$$

- Or, compactly:

$$B(X_{t+1}) \propto_{X_{t+1}} P(e_{t+1} | X_{t+1}) B'(X_{t+1})$$



- Basic idea: beliefs “reweighted” by likelihood of evidence
- Unlike passage of time, we have to renormalize



Example: Weather HMM

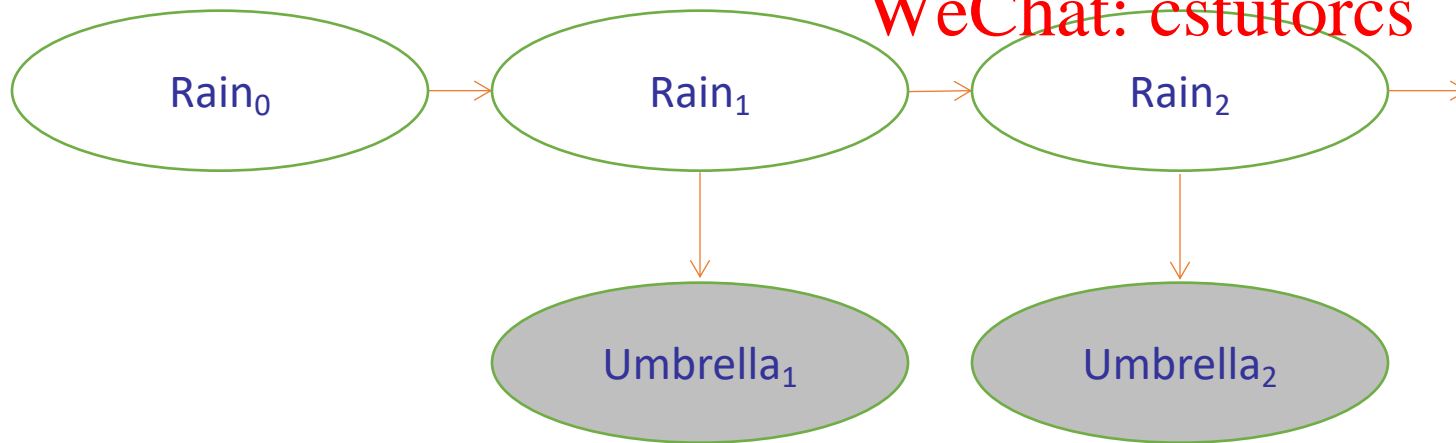


$$\begin{array}{l}
 B(+r) = 0.5 \\
 B(-r) = 0.5
 \end{array}
 \quad
 \begin{array}{l}
 B'(+r) = 0.5 \\
 B'(-r) = 0.5
 \end{array}
 \quad
 \begin{array}{l}
 B(+r) = 0.818 \\
 B(-r) = 0.182
 \end{array}
 \quad
 \begin{array}{l}
 B'(+r) = 0.627 \\
 B'(-r) = 0.373
 \end{array}
 \quad
 \begin{array}{l}
 B(+r) = 0.883 \\
 B(-r) = 0.117
 \end{array}$$

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: estutorcs



R_t	R_{t+1}	$P(R_{t+1} R_t)$
+r	+r	0.7
+r	-r	0.3
-r	+r	0.3
-r	-r	0.7

R_t	U_t	$P(U_t R_t)$
+r	+u	0.9
+r	-u	0.1
-r	+u	0.2
-r	-u	0.8



Online Belief Updates

- Every time step, we start with current $P(X \mid \text{evidence})$

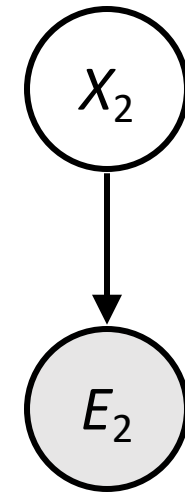
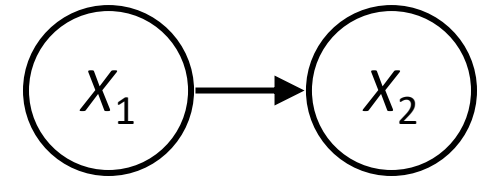
- We update for time:

$$P(x_t | e_{1:t-1}) = \sum_{x_{t-1}} P(x_{t-1} | e_{1:t-1}) \cdot P(x_t | x_{t-1})$$

WeChat: cstutorcs

- We update for evidence:

$$P(x_t | e_{1:t}) \propto_X P(x_t | e_{1:t-1}) \cdot P(e_t | x_t)$$



Next Time: Particle Filtering and Applications of HMMs

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

