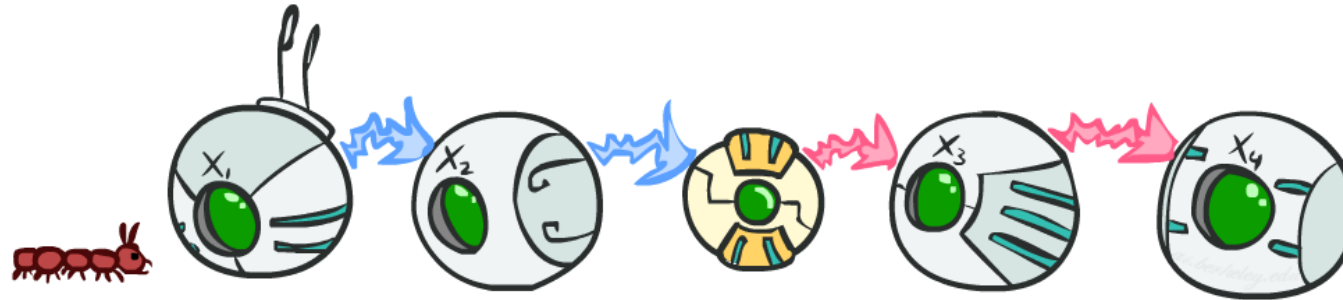


CS 5522: Artificial Intelligence II

Markov Models



Instructor: Wei Xu

Ohio State University

[These slides were adapted from CS188 Intro to AI at UC Berkeley.]

Probabilistic Inference

- Probabilistic inference: compute a desired probability from other known probabilities (e.g. conditional from joint)
- We generally compute conditional probabilities
 - $P(\text{on time} \mid \text{no reported accidents}) = 0.90$
 - These represent the agent's *beliefs* given the evidence
- Probabilities change with new evidence:
 - $P(\text{on time} \mid \text{no accidents, 5 a.m.}) = 0.95$
 - $P(\text{on time} \mid \text{no accidents, 5 a.m., raining}) = 0.80$
 - Observing new evidence causes *beliefs to be updated*



Inference by Enumeration

- General case:


- Evidence variables: $E_1 \dots E_k = e_1 \dots e_k$
 - Query* variable: Q
 - Hidden variables: $H_1 \dots H_r$
- $\left. \begin{array}{l} E_1 \dots E_k = e_1 \dots e_k \\ Q \\ H_1 \dots H_r \end{array} \right\} \begin{array}{l} X_1, X_2, \dots X_n \\ \text{All variables} \end{array}$

- We want:

* Works fine with multiple query variables, too

$$P(Q|e_1 \dots e_k)$$

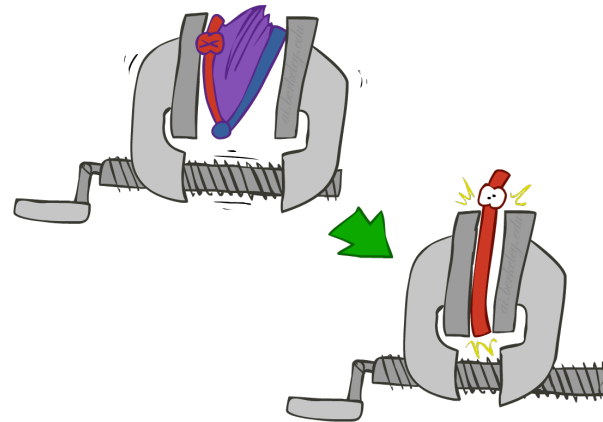
- Step 1: Select the entries consistent with the evidence



x	P(x)
-3	0.05
-1	0.25
0	0.07
1	0.2
5	0.01

2 0.15

- Step 2: Sum out H to get joint of Query and evidence



$$P(Q, e_1 \dots e_k) = \sum_{h_1 \dots h_r} \underbrace{P(Q, h_1 \dots h_r, e_1 \dots e_k)}_{X_1, X_2, \dots X_n}$$

- Step 3: Normalize

$$\times \frac{1}{Z}$$

$$Z = \sum_q P(Q, e_1 \dots e_k)$$

$$P(Q|e_1 \dots e_k) = \frac{1}{Z} P(Q, e_1 \dots e_k)$$

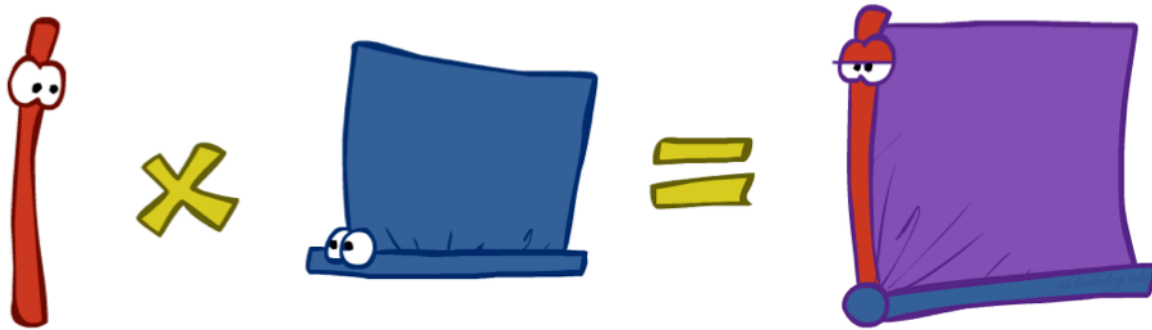
Inference by Enumeration

- Obvious problems:
 - Worst-case time complexity $O(d^n)$
 - Space complexity $O(d^n)$ to store the joint distribution

The Product Rule

- Sometimes have conditional distributions but want the joint

$$P(y)P(x|y) = P(x, y) \quad \longleftrightarrow \quad P(x|y) = \frac{P(x, y)}{P(y)}$$



The Product Rule

$$P(y)P(x|y) = P(x, y)$$

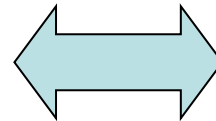
- Example:

$P(W)$

W	P
sun	0.8
rain	0.2

$P(D|W)$

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3



$P(D, W)$

D	W	P
wet	sun	
dry	sun	
wet	rain	
dry	rain	

The Chain Rule

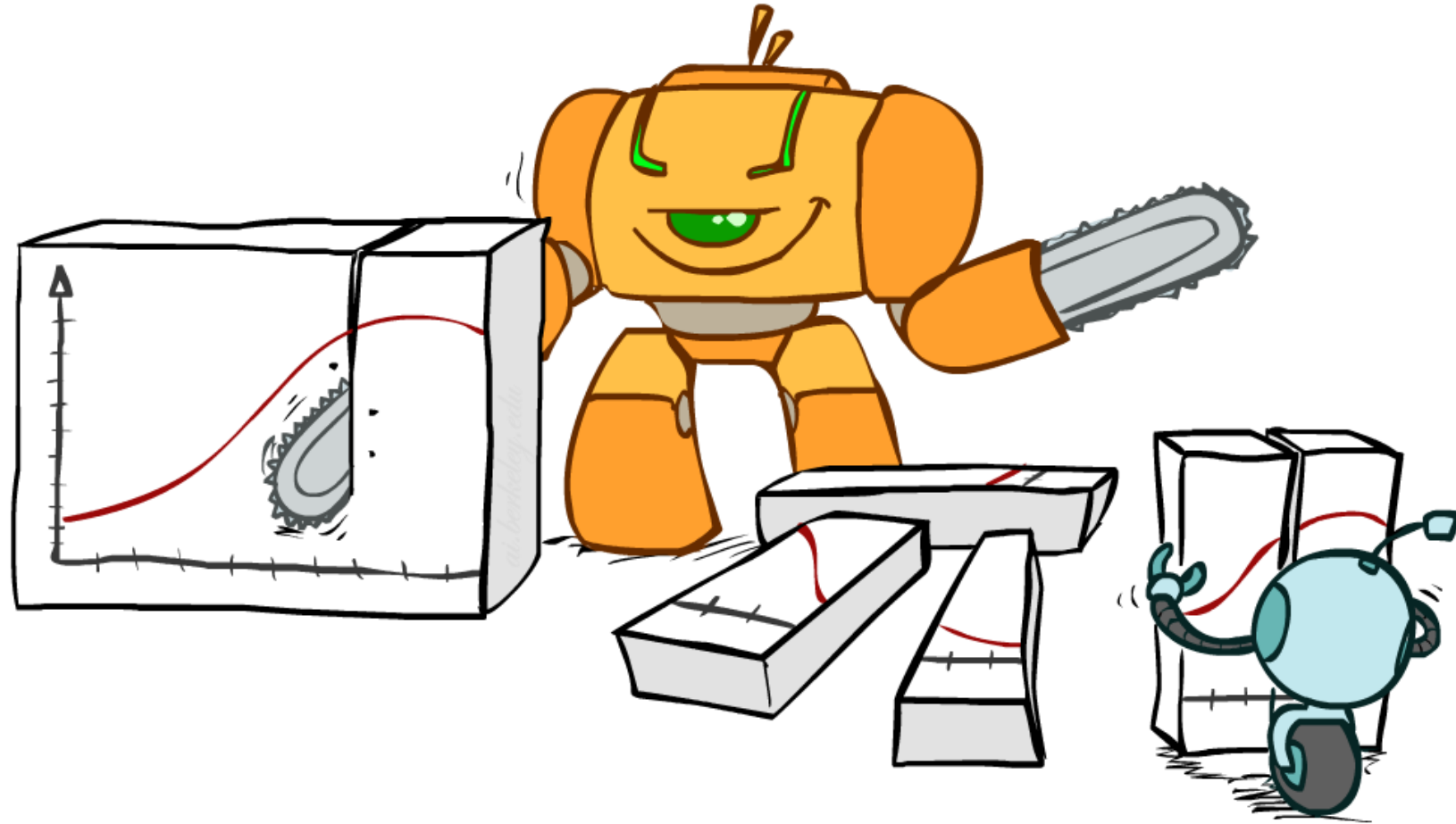
- More generally, can always write any joint distribution as an incremental product of conditional distributions

$$P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)$$

$$P(x_1, x_2, \dots x_n) = \prod_i P(x_i|x_1 \dots x_{i-1})$$

- Why is this always true?

Bayes Rule



Bayes' Rule

- Two ways to factor a joint distribution over two variables:

$$P(x, y) = P(x|y)P(y) = P(y|x)P(x)$$

That's my rule!

- Dividing, we get:

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}$$

- Why is this at all helpful?

- Lets us build one conditional from its reverse
- Often one conditional is tricky but the other one is simple
- Foundation of many systems we'll see later (e.g. ASR, MT)

- In the running for most important AI equation!



Inference with Bayes' Rule

- Example: Diagnostic probability from causal probability:

$$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause})P(\text{cause})}{P(\text{effect})}$$

- Example:

- M: meningitis, S: stiff neck

$$\left. \begin{array}{l} P(+m) = 0.0001 \\ P(+s|+m) = 0.8 \\ P(+s|-m) = 0.01 \end{array} \right\} \text{Example givens}$$

$$P(+m|+s) = \frac{P(+s|+m)P(+m)}{P(+s)} = \frac{P(+s|+m)P(+m)}{P(+s|+m)P(+m) + P(+s|-m)P(-m)} = \frac{0.8 \times 0.0001}{0.8 \times 0.0001 + 0.01 \times 0.999}$$

- Note: posterior probability of meningitis still very small
- Note: you should still get stiff necks checked out! Why?

Quiz: Bayes' Rule

- Given:

$P(W)$

R	P
sun	0.8
rain	0.2

$P(D|W)$

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3

- What is $P(W \mid \text{dry})$?

Independence

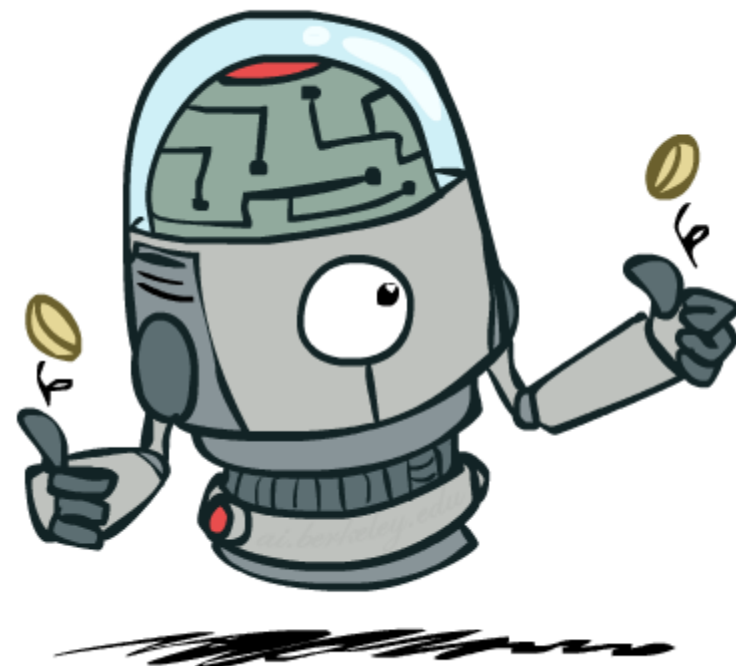
- Two variables are *independent* in a joint distribution if:

$$P(X, Y) = P(X)P(Y)$$

$$X \perp\!\!\!\perp Y$$

$$\forall x, y \ P(x, y) = P(x)P(y)$$

- Says the joint distribution *factors* into a product of two simple ones
 - Usually variables aren't independent!
- Can use independence as a *modeling assumption*
 - Independence can be a simplifying assumption
 - Empirical* joint distributions: at best “close” to independent
 - What could we assume for {Weather, Traffic, Cavity}?



Example: Independence?

$P_1(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P(T)$

T	P
hot	0.5
cold	0.5

$$P_2(T, W) = P(T)P(W)$$

T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

$P(W)$

W	P
sun	0.6
rain	0.4

Example: Independence

- N fair, independent coin flips:

$P(X_1)$

H	0.5
T	0.5

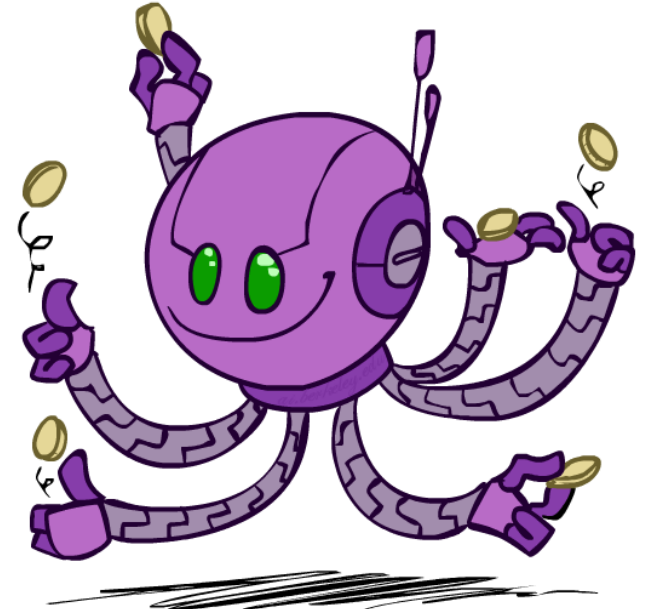
$P(X_2)$

H	0.5
T	0.5

...

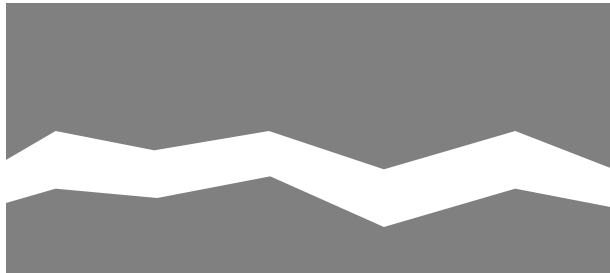
$P(X_n)$

H	0.5
T	0.5

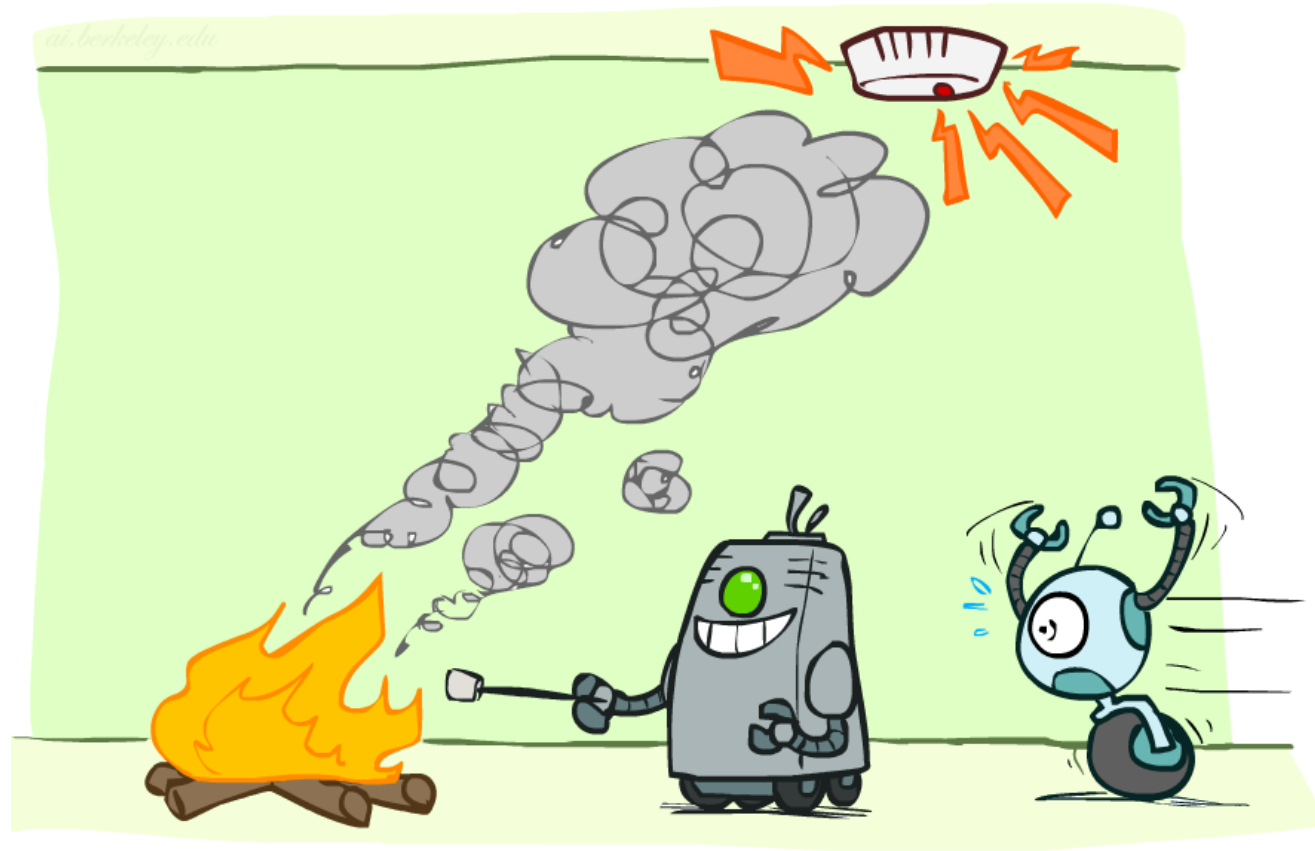


$P(X_1, X_2, \dots, X_n)$

2^n

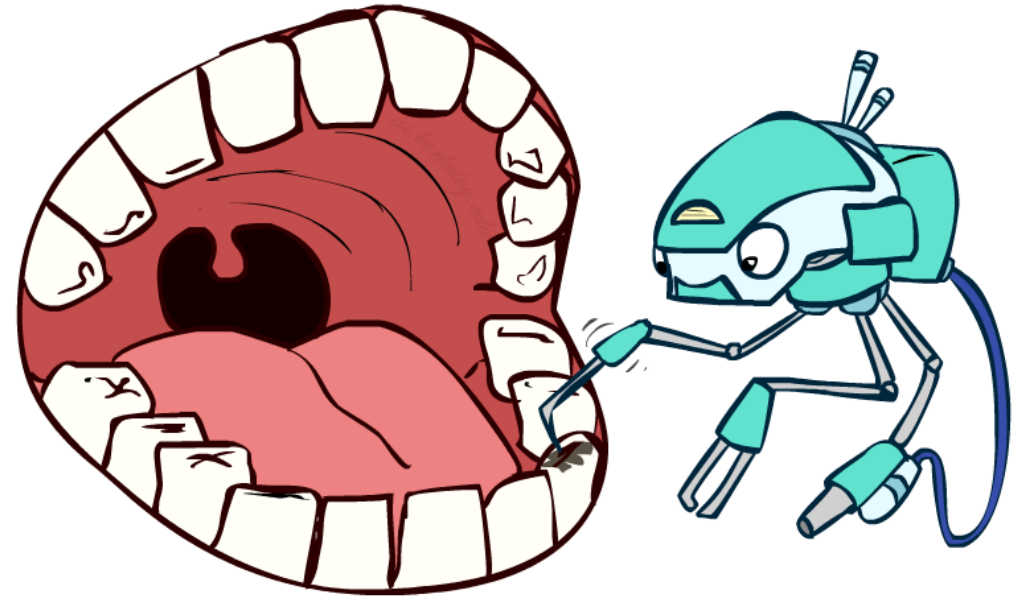


Conditional Independence



Conditional Independence

- $P(\text{Toothache}, \text{Cavity}, \text{Catch})$
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
 - $P(+\text{catch} \mid +\text{toothache}, +\text{cavity}) = P(+\text{catch} \mid +\text{cavity})$
- The same independence holds if I don't have a cavity:
 - $P(+\text{catch} \mid +\text{toothache}, -\text{cavity}) = P(+\text{catch} \mid -\text{cavity})$
- Catch is *conditionally independent* of Toothache given Cavity:
 - $P(\text{Catch} \mid \text{Toothache}, \text{Cavity}) = P(\text{Catch} \mid \text{Cavity})$
- Equivalent statements:
 - $P(\text{Toothache} \mid \text{Catch}, \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity})$
 - $P(\text{Toothache}, \text{Catch} \mid \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) P(\text{Catch} \mid \text{Cavity})$
 - One can be derived from the other easily



Conditional Independence

- Unconditional (absolute) independence very rare (why?)
- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments.
- X is conditionally independent of Y given Z $X \perp\!\!\!\perp Y | Z$

if and only if:

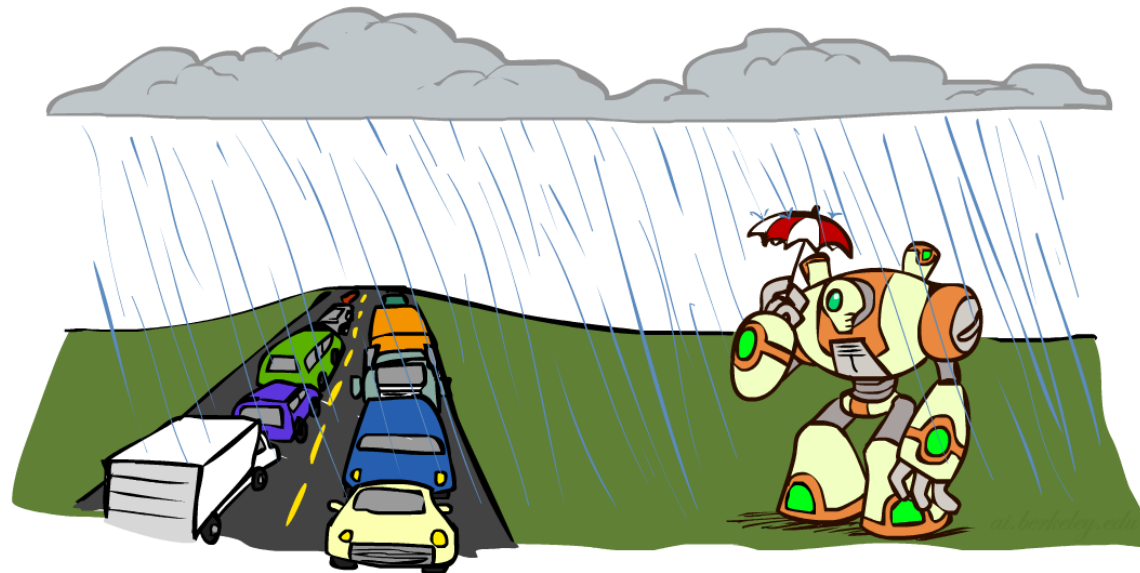
$$\forall x, y, z : P(x, y | z) = P(x | z)P(y | z)$$

or, equivalently, if and only if

$$\forall x, y, z : P(x | z, y) = P(x | z)$$

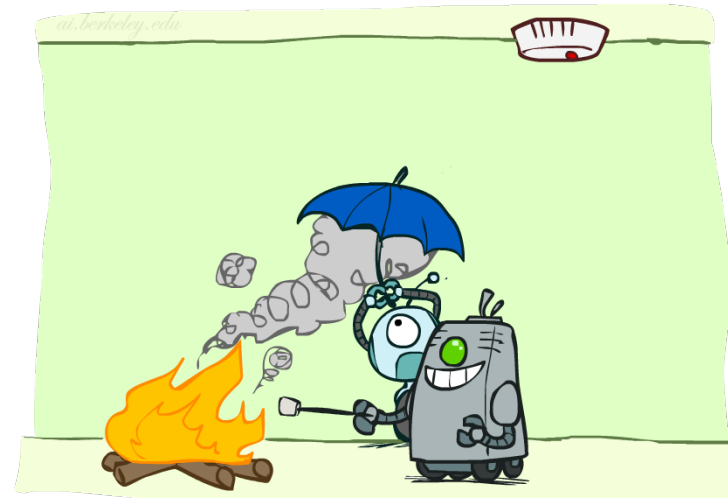
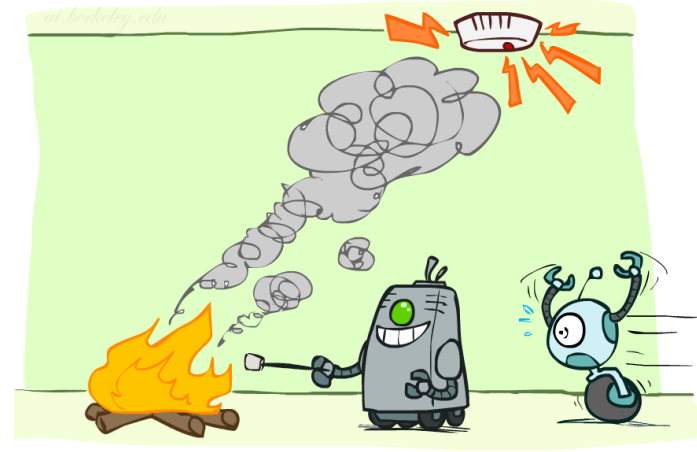
Conditional Independence

- What about this domain:
 - Traffic
 - Umbrella
 - Raining



Conditional Independence

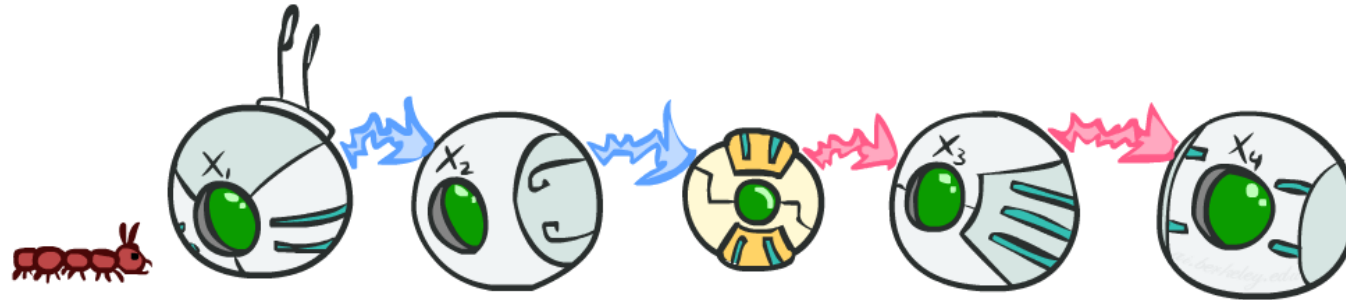
- What about this domain:
 - Fire
 - Smoke
 - Alarm



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
$$\begin{aligned} P(X_1, X_2, \dots, X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \dots \\ &= \prod_{i=1}^n P(X_i|X_1, \dots, X_{i-1}) \end{aligned}$$
- 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 $X \perp\!\!\!\perp Y | Z$
 $\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$

Markov Models

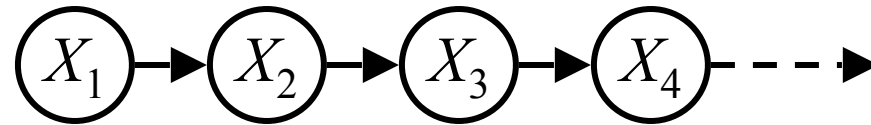


Reasoning over Time or Space

- 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

Markov Models

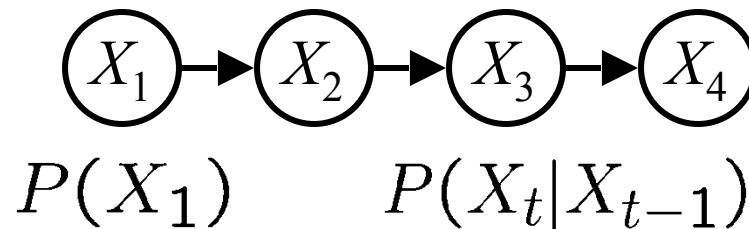
- Value of X at a given time is called the **state**



$$P(X_1) \quad P(X_t|X_{t-1})$$

- 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

Joint Distribution of a Markov Model



- Joint distribution:

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)$$

- More generally:

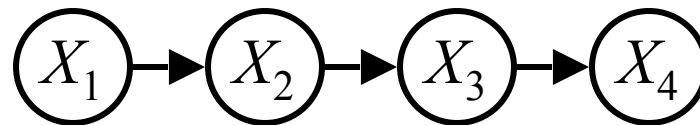
$$P(X_1, X_2, \dots, X_T) = P(X_1)P(X_2|X_1)P(X_3|X_2) \dots P(X_T|X_{T-1})$$

$$= P(X_1) \prod_{t=2}^T P(X_t|X_{t-1})$$

- Questions to be resolved:

- Does this indeed define a joint distribution?
- Can every joint distribution be factored this way, or are we making some assumptions about the joint distribution by using this factorization?

Chain Rule and Markov Models



- From the chain rule, every joint distribution over X_1, X_2, X_3, X_4 can be written as:

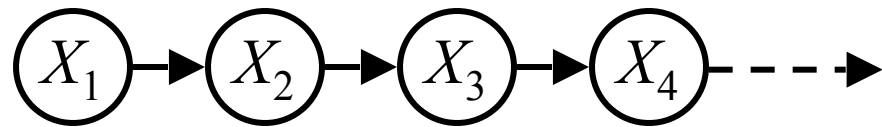
$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)P(X_4|X_1, X_2, X_3)$$

- Assuming that $X_3 \perp\!\!\!\perp X_1 \mid X_2$ and $X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3$

results in the expression posited on the previous slide:

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)$$

Chain Rule and Markov Models



- From the chain rule, every joint distribution over X_1, X_2, \dots, X_T can be written as:

$$P(X_1, X_2, \dots, X_T) = P(X_1) \prod_{t=2}^T P(X_t | X_1, X_2, \dots, X_{t-1})$$

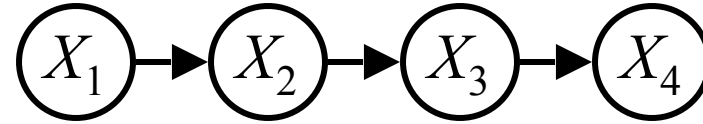
- Assuming that for all t :

$$X_t \perp\!\!\!\perp X_1, \dots, X_{t-2} \mid X_{t-1}$$

gives us the expression posited on the earlier slide:

$$P(X_1, X_2, \dots, X_T) = P(X_1) \prod_{t=2}^T P(X_t | X_{t-1})$$

Implied Conditional Independencies



- We assumed: $X_3 \perp\!\!\!\perp X_1 \mid X_2$ and $X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3$

- Do we also have $X_1 \perp\!\!\!\perp X_3, X_4 \mid X_2$?

- Yes!

- Proof:

$$\begin{aligned} P(X_1 \mid X_2, X_3, X_4) &= \frac{P(X_1, X_2, X_3, X_4)}{P(X_2, X_3, X_4)} \\ &= \frac{P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_2)P(X_4 \mid X_3)}{\sum_{x_1} P(x_1)P(X_2 \mid x_1)P(X_3 \mid X_2)P(X_4 \mid X_3)} \\ &= \frac{P(X_1, X_2)}{P(X_2)} \\ &= P(X_1 \mid X_2) \end{aligned}$$

Markov Models Recap

- Explicit assumption for all t : $X_t \perp\!\!\!\perp X_1, \dots, X_{t-2} \mid X_{t-1}$

- Consequence, joint distribution can be written as:

$$\begin{aligned} P(X_1, X_2, \dots, X_T) &= P(X_1)P(X_2|X_1)P(X_3|X_2) \dots P(X_T|X_{T-1}) \\ &= P(X_1) \prod_{t=2}^T P(X_t|X_{t-1}) \end{aligned}$$

- Implied conditional independencies: (try to prove this!)

- Past variables independent of future variables given the present

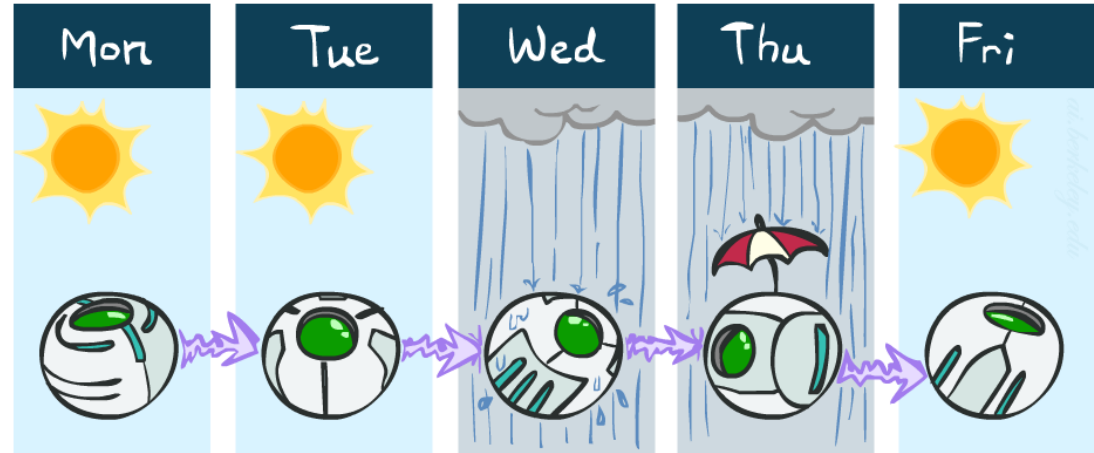
i.e., if $t_1 < t_2 < t_3$ or $t_1 > t_2 > t_3$ then: $X_{t_1} \perp\!\!\!\perp X_{t_3} \mid X_{t_2}$

- Additional explicit assumption: $P(X_t \mid X_{t-1})$ is the same for all t

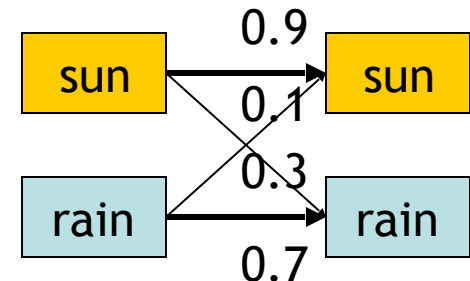
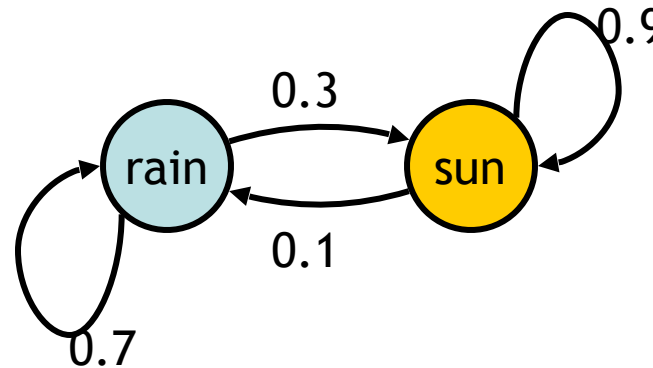
Example Markov Chain: Weather

- States: $X = \{\text{rain}, \text{sun}\}$
- Initial distribution: 1.0 sun
- CPT $P(X_t \mid X_{t-1})$:

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

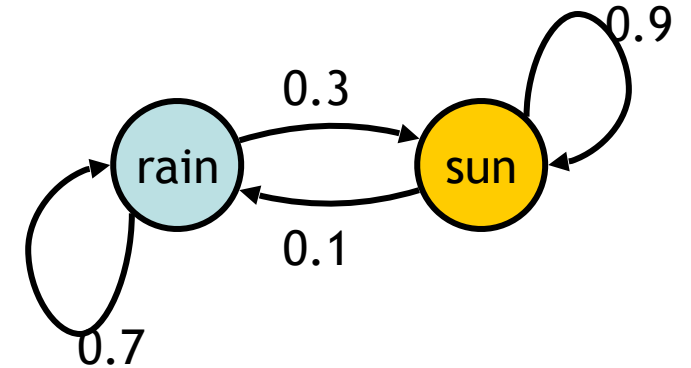


Two new ways of representing the same CPT



Example Markov Chain: Weather

- Initial distribution: 1.0 sun



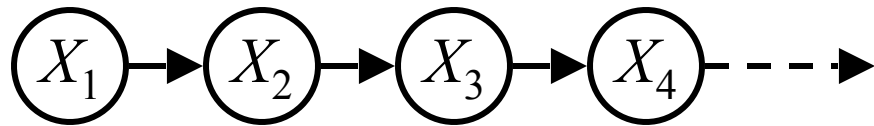
- What is the probability distribution after one step?

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

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

Mini-Forward Algorithm

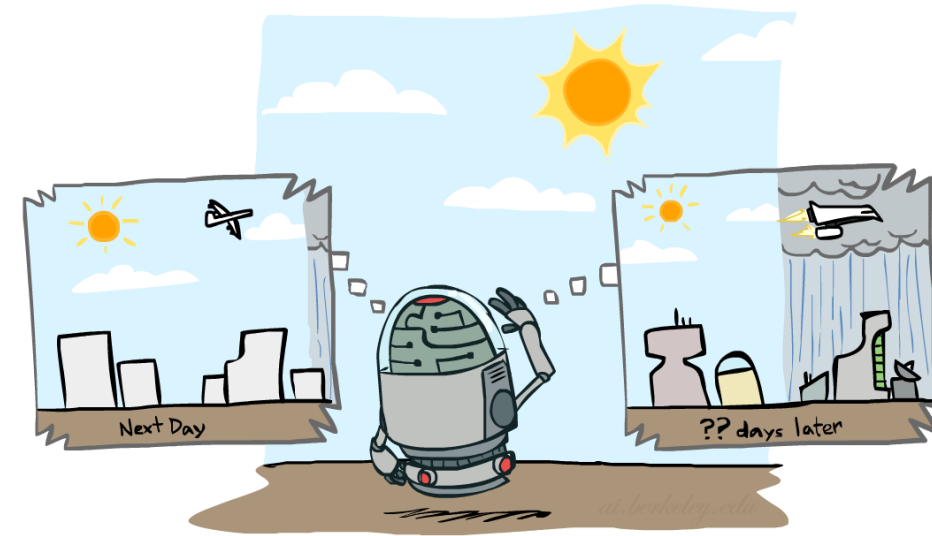
- Question: What's $P(X)$ on some day t ?



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

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

Forward simulation



Example Run of Mini-Forward Algorithm

- From initial observation of sun

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

- From initial observation of rain

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

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

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

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:

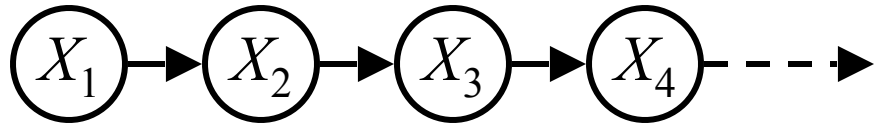
- The distribution we end up with is called the **stationary distribution** P_∞ of the chain
- It satisfies

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



Example: Stationary Distributions

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



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

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

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

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

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

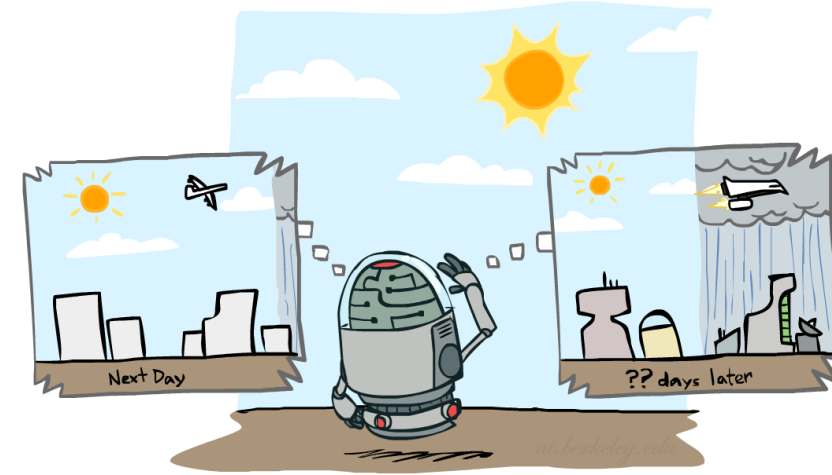
$$P_{\infty}(\text{rain}) = 1/3P_{\infty}(\text{sun})$$

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

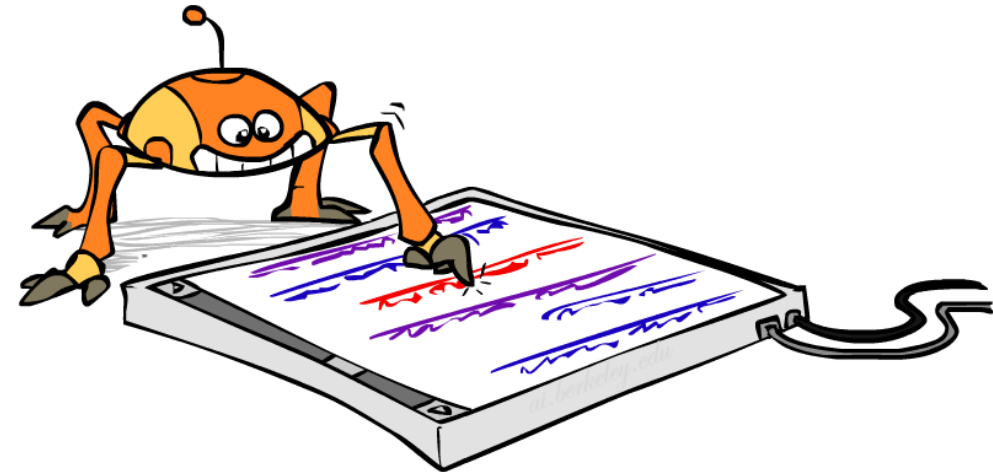
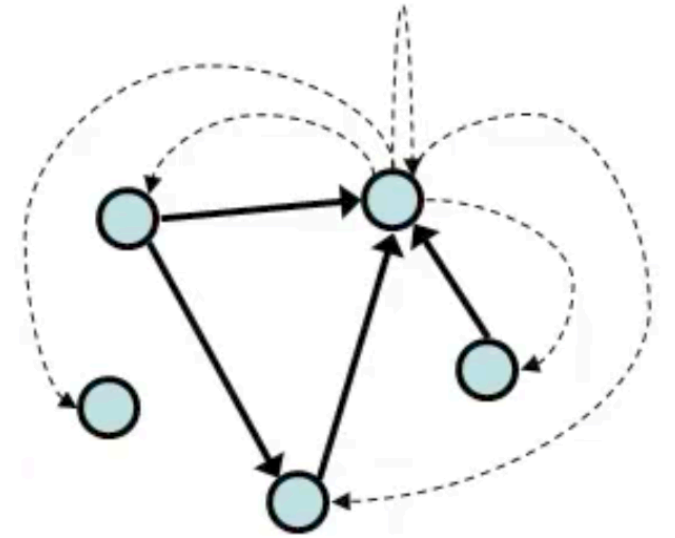
Application of Stationary Distribution: Web Link Analysis

- 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 (rank actually getting less important over time)



Next Time: Hidden Markov Models!
