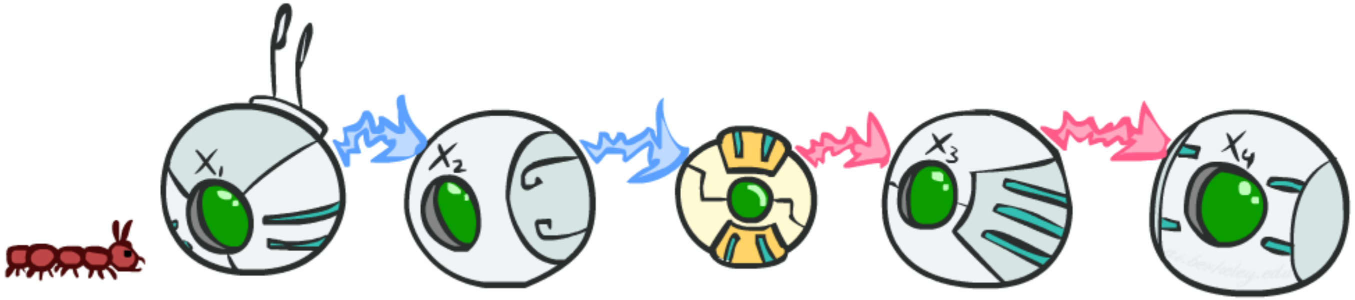
CS 4100 // artificial intelligence

instructor: byron wallace



Markov Models

Attribution: many of these slides are modified versions of those distributed with the *UC Berkeley CS188* materials
Thanks to John DeNero and Dan Klein

Today

- We'll cover conditional independence in some depth
- And build up to Markov models

Review: the product rule

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

Example:

$P(W)$

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

$\Leftrightarrow$

$P(D,W)$

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

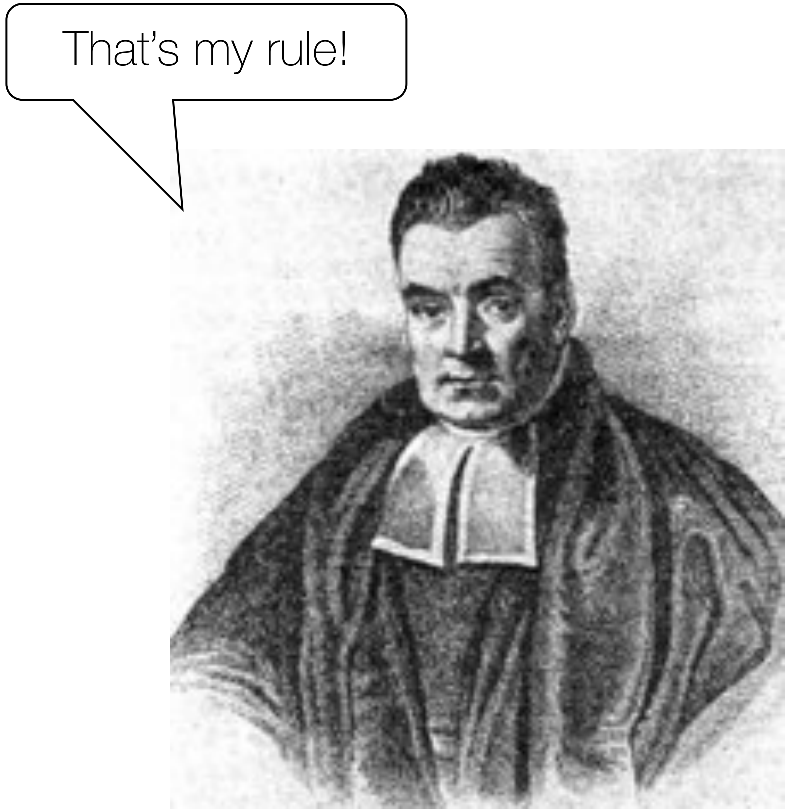
$$P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)$$

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

Bayes' rule

Two ways to factor a joint distribution over two variables:

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



Dividing, we get:

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

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)

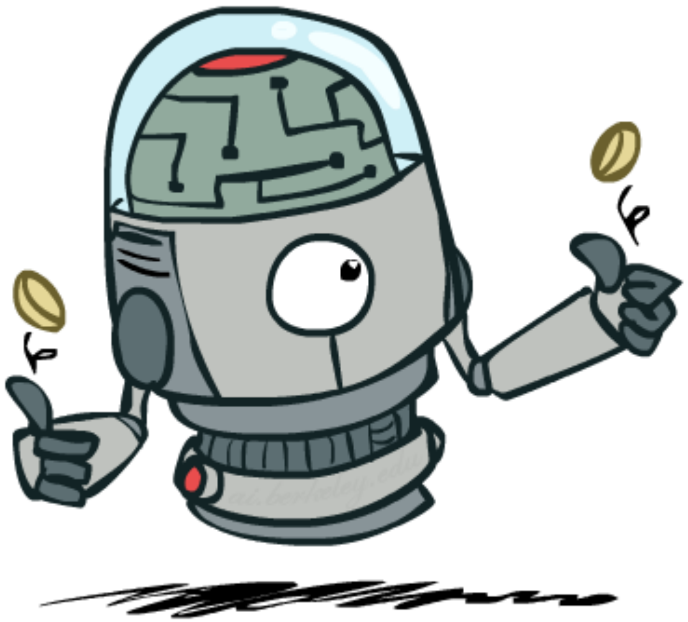
In the running for most important AI equation!

Independence

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

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

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



- Says the joint distribution *factors* into a product of two simple ones
- Usually variables aren't independent!

We 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}?

Independence is like something from CSPs: what?

Example: Independence?

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T, W)$

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

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

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

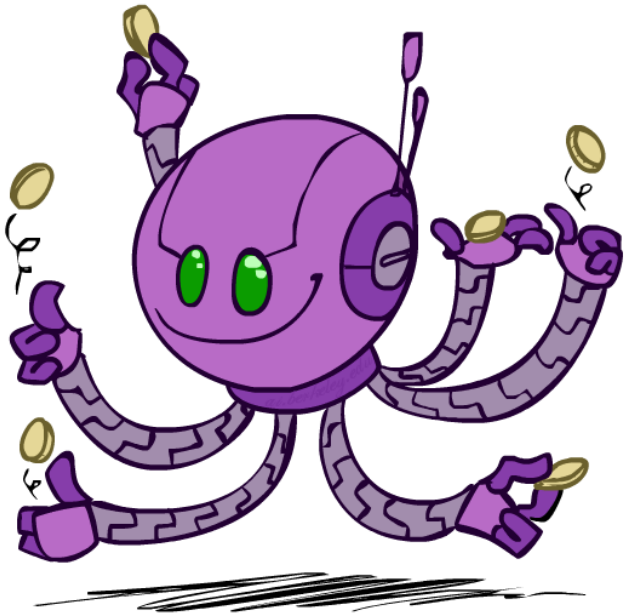
$P(X_2)$

H	0.5
T	0.5

$\dots$

$P(X_n)$

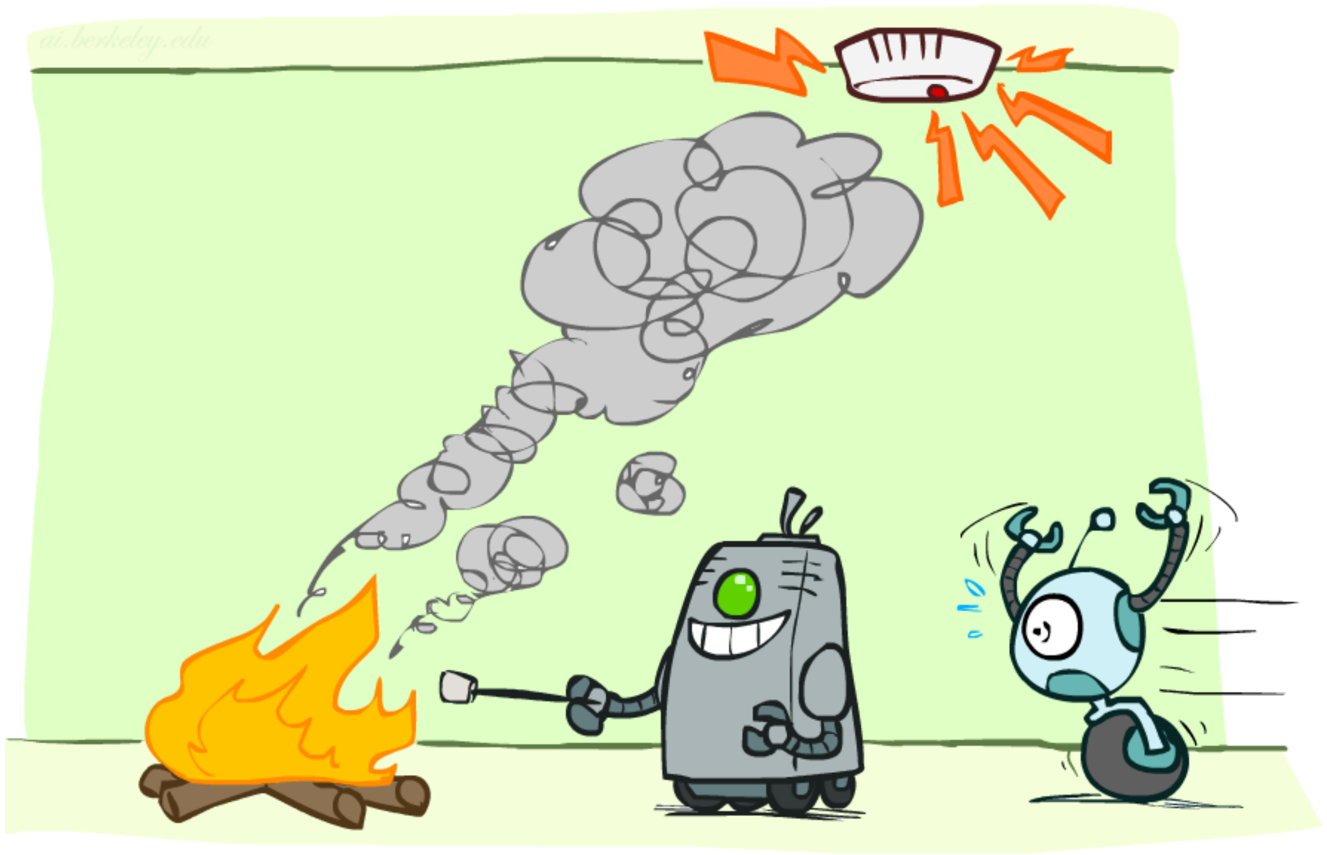
H	0.5
T	0.5



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

2^n

Conditional independence

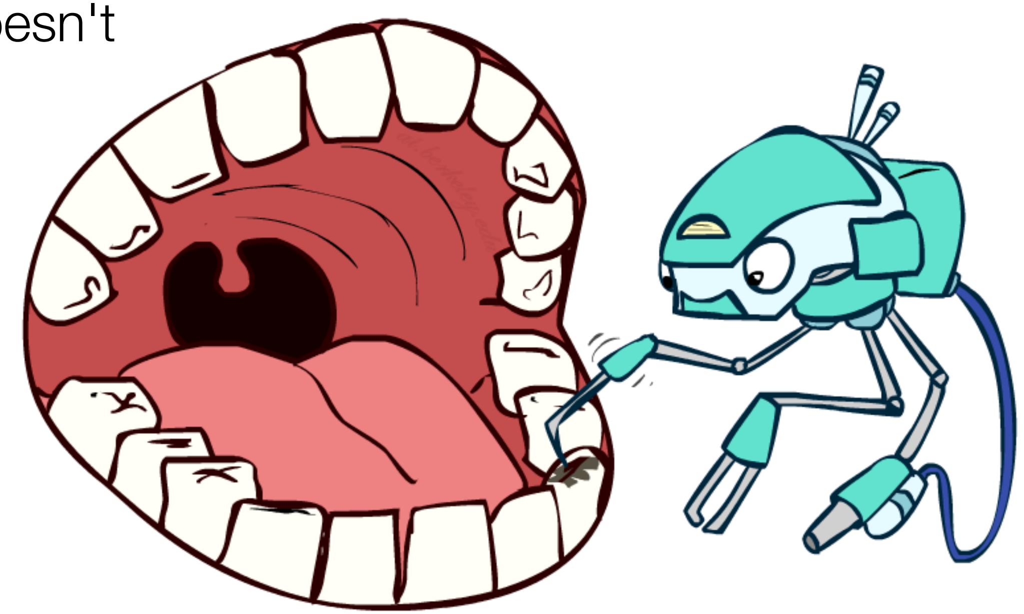


Conditional independence

P(Toothache, Cavity, Catch*) **catch* means probe finds/gets stuck in a cavity

If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

P(+catch | +toothache, +cavity) = P(+catch | +cavity)



The same independence holds if I don't have a cavity:

P(+catch | +toothache, -cavity) = P(+catch| -cavity)

Catch is *conditionally independent* of Toothache given Cavity:

P(Catch | Toothache, Cavity) = P(Catch | Cavity)

Equivalent statements:

- P(Toothache | Catch , Cavity) = P(Toothache | Cavity)
- P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)
- One can be derived from the other easily

Conditional independence (formal def)

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, written

$$X \perp\!\!\!\perp Y | Z$$

if and only if:

$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z) \qquad (1)$$

or, equivalently, if and only if

$$\forall x, y, z : P(x|z, y) = P(x|z) \qquad (2)$$

Proving equivalence (on board)

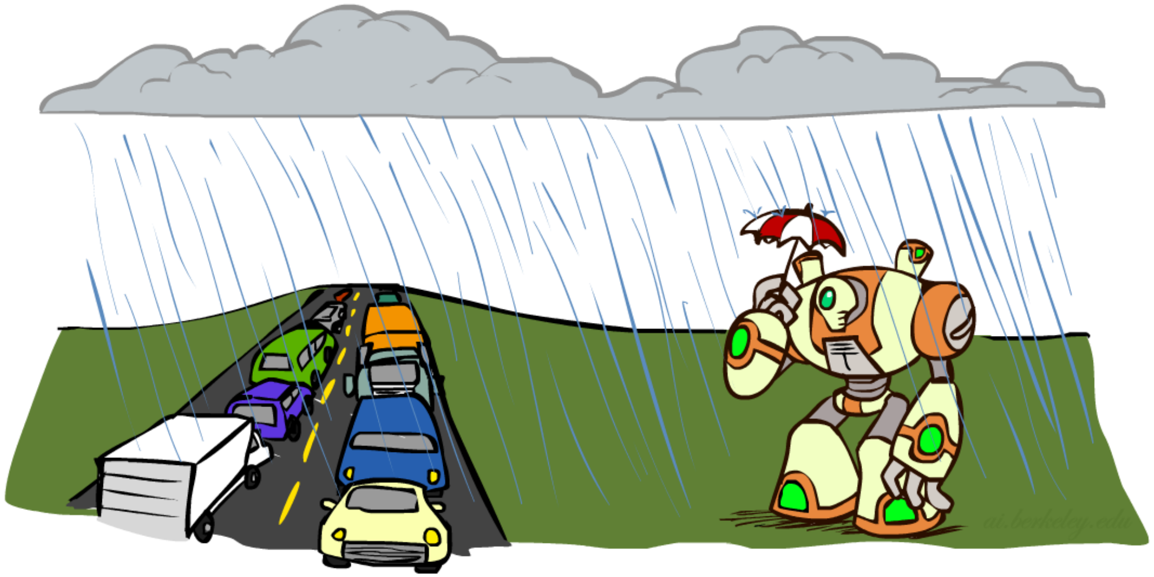
$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z) \longleftrightarrow \forall x, y, z : P(x|z, y) = P(x|z)$$

(1) (2)

Conditional independence

What about this domain:

- Traffic
- Umbrella
- Raining

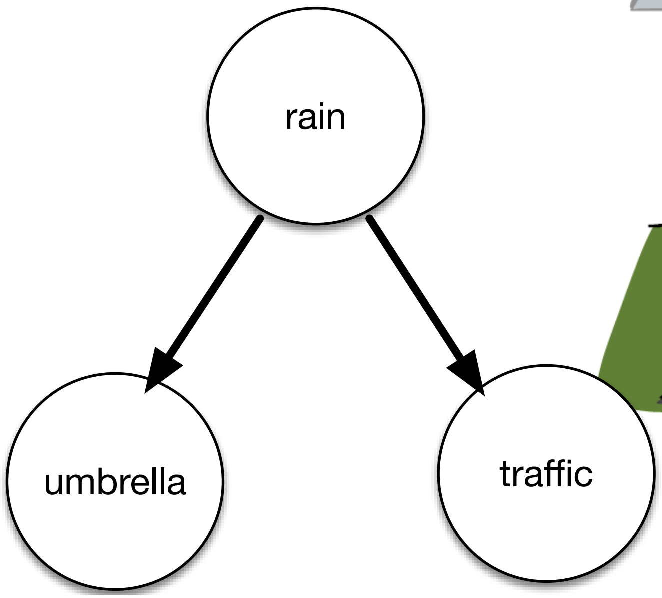


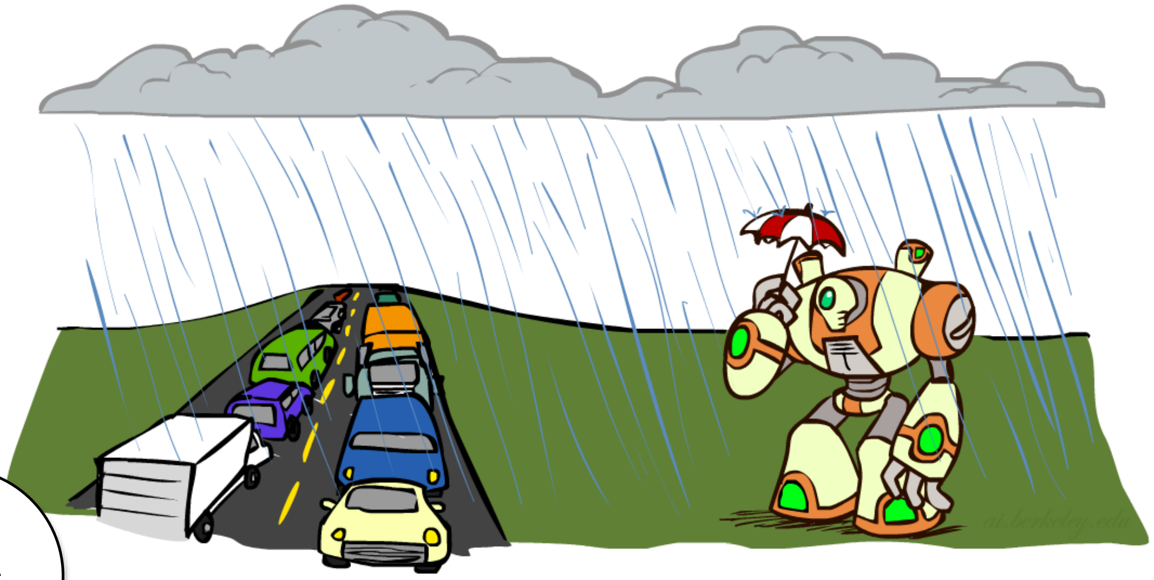
Reasonable independence assumption here?

Conditional independence

What about this domain:

- Traffic
- Umbrella
- Raining

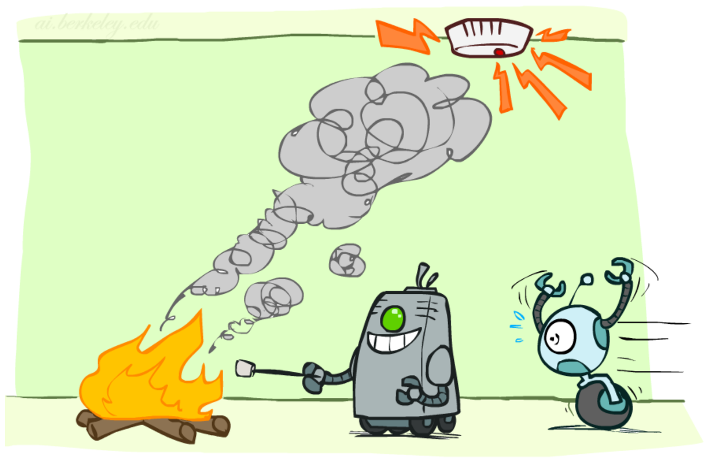


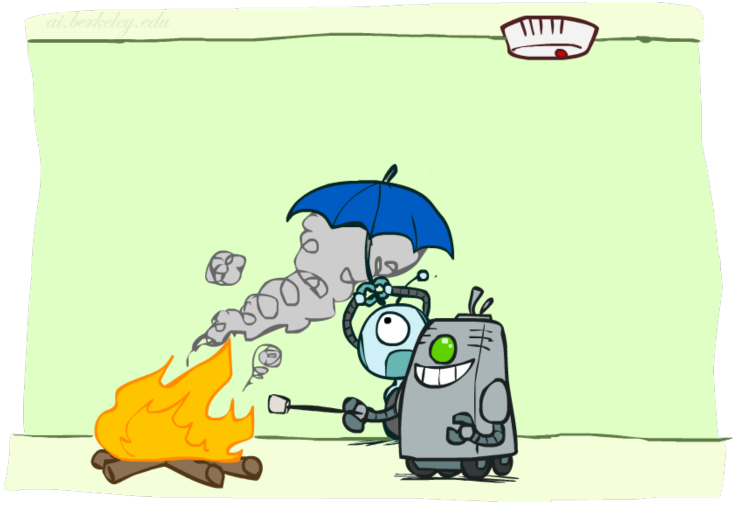


Conditional independence

What about this domain:

- Fire
- Smoke
- Alarm

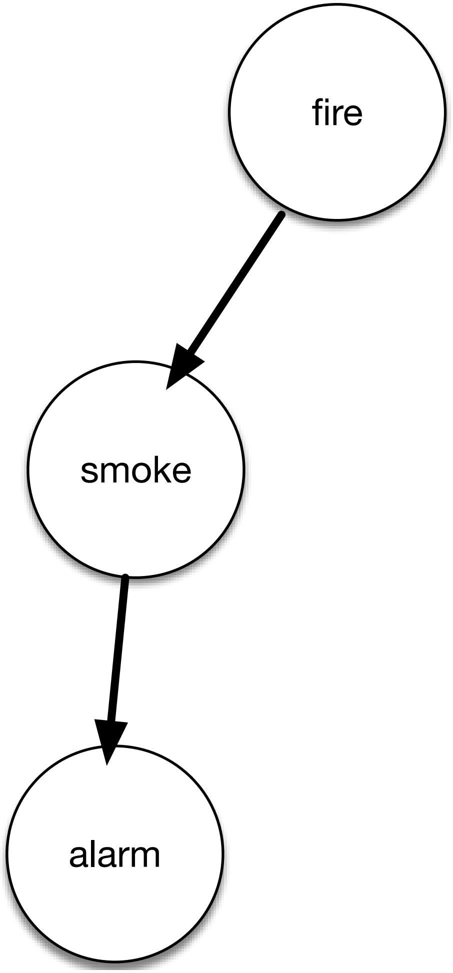


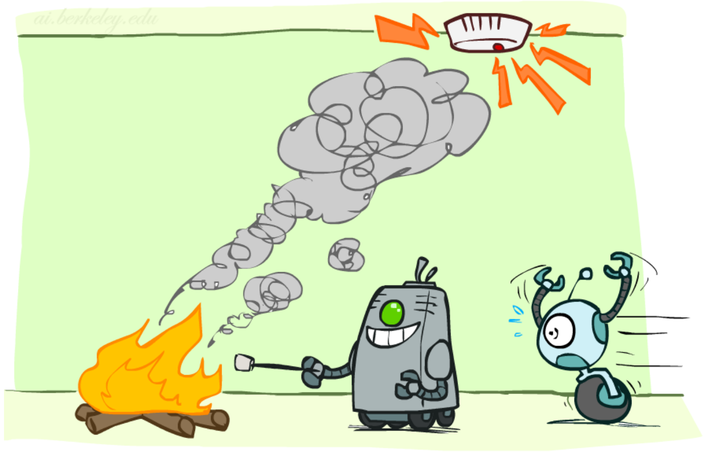


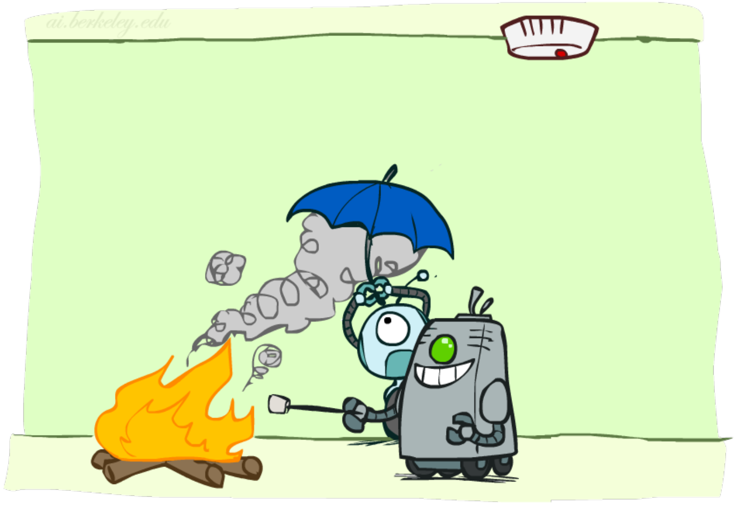
Conditional independence

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

$$\begin{aligned} P(X_1, X_2, \ldots X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots \\ &= \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1}) \end{aligned}$$

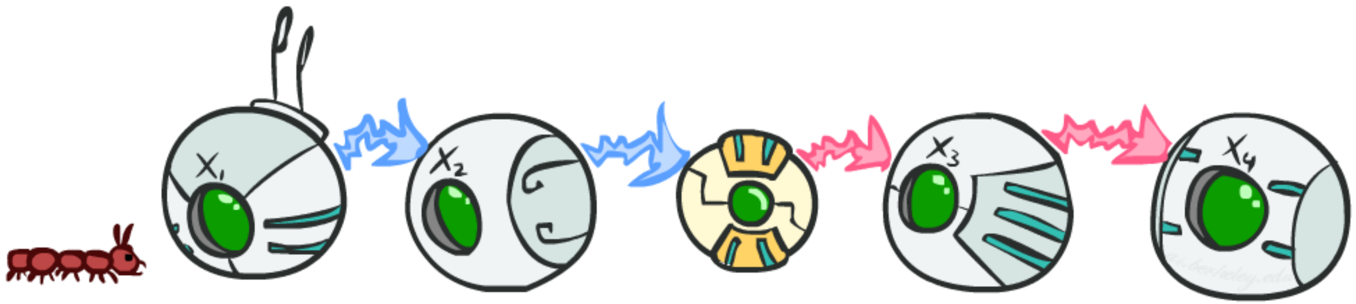
- X, Y independent if and only if:

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

- X and Y are conditionally independent given Z ($X \perp\!\!\!\perp Y \,|\, Z$) if and only if:

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

Markov models



Reasoning over *time* or *space*

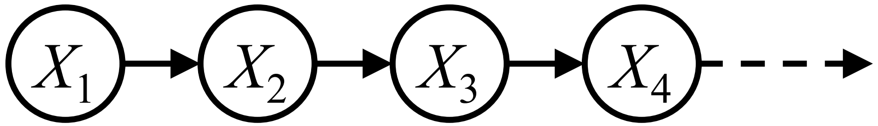
Often, we want to **reason about a sequence** of observations

- 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**



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

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

$$X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4$$

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

Joint distribution:

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

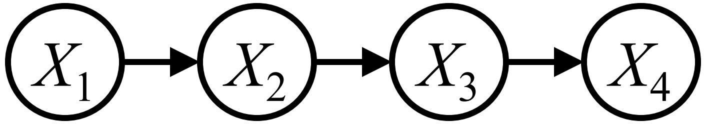
More generally:

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

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:

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

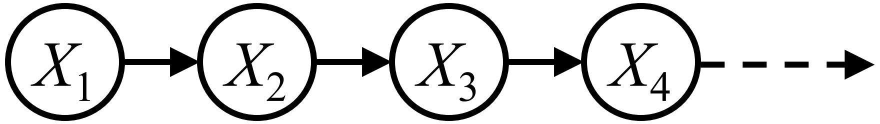
Assuming that

$$X_3 \perp\!\!\!\perp X_1 \mid X_2 \quad \text{and} \quad X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3$$

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 $X_1, X_2, \ldots, X_T$ can be written as:

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

Assuming that for all t:

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

Gives us the expression posited on the earlier slide:

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

Implied conditional independencies

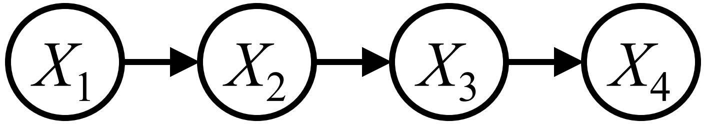
$$X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4$$

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! Though we did not explicitly make this assumption!
- Proof: *on board*

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*: $X_t \perp\!\!\!\perp X_1, \ldots, X_{t-2} \mid X_{t-1}$

Consequence, joint distribution can be written as:

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

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*

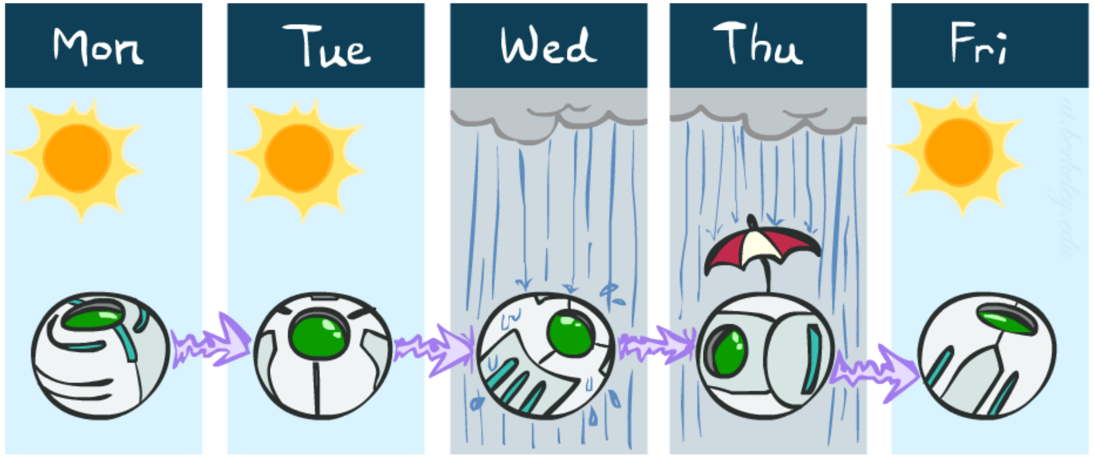
Example Markov chain: the weather

States: X = {rain, sun}

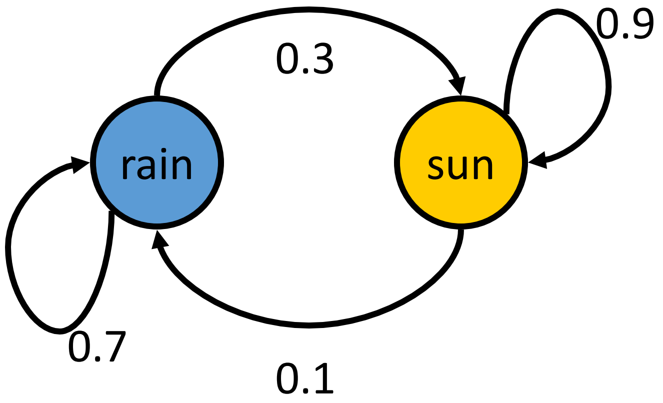
Initial distribution: 1.0 sun

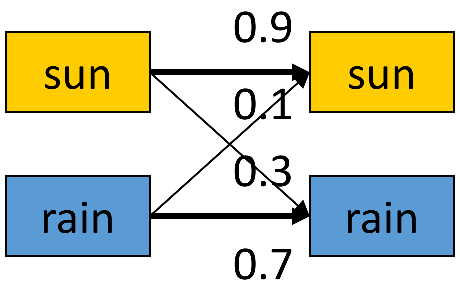
CPT $P(X_t \mid X_{t-1})$:

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



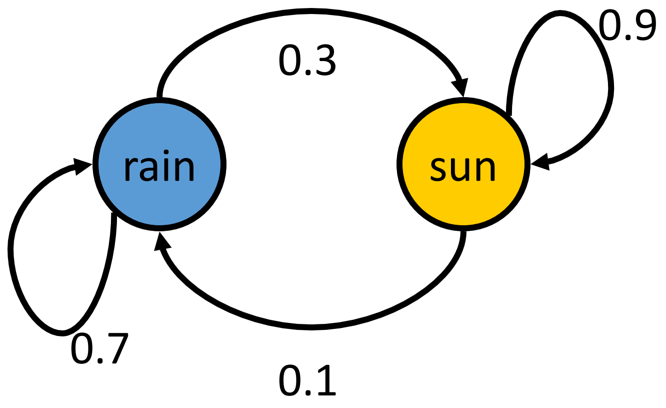
Two new ways of representing the same CPT





Example Markov chain: the weather

Initial distribution: 1.0 sun



What is the probability distribution after one step?

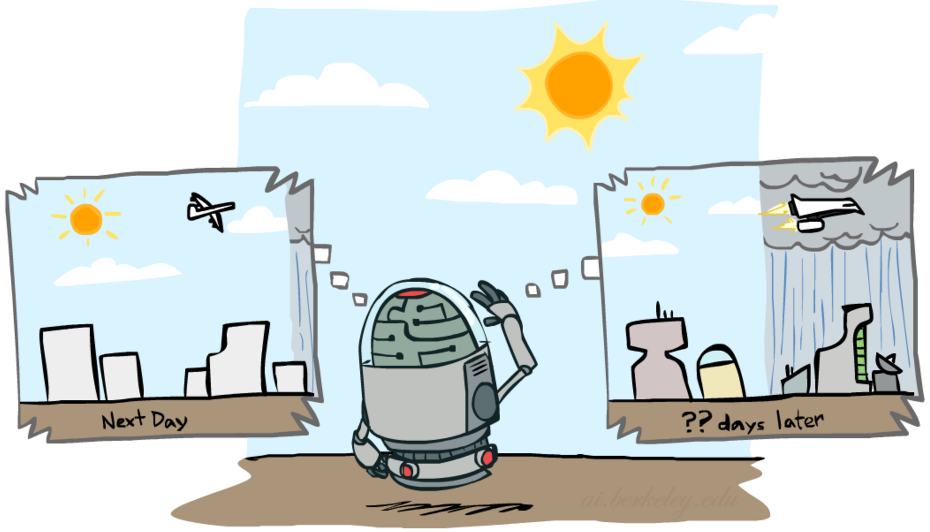
$$P(X_2 = \text{sun}) = \quad P(X_2 = \text{sun} | X_1 = \text{sun}) P(X_1 = \text{sun}) + P(X_2 = \text{sun} | X_1 = \text{rain}) P(X_1 = \text{rain})$$

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

Mini-forward algorithm

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

$$X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4 \dashrightarrow$$



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

$$P(x_t) = \sum_{x_{t-1}} P(x_{t-1}, x_t)$$

$$= \sum_{x_{t-1}} P(x_t \mid x_{t-1}) P(x_{t-1})$$

Forward simulation

Example run of mini-forward algorithm

From initial observation of sun

$$\left\langle \begin{array}{c} 1.0 \\ 0.0 \end{array} \right\rangle \quad \left\langle \begin{array}{c} 0.9 \\ 0.1 \end{array} \right\rangle \quad \left\langle \begin{array}{c} 0.84 \\ 0.16 \end{array} \right\rangle \quad \left\langle \begin{array}{c} 0.804 \\ 0.196 \end{array} \right\rangle \Rightarrow \left\langle \begin{array}{c} 0.75 \\ 0.25 \end{array} \right\rangle$$

$$\mathrm{P}(X_1) \quad \mathrm{P}(X_2) \quad \mathrm{P}(X_3) \quad \mathrm{P}(X_4) \quad \mathrm{P}(X_\infty)$$

From initial observation of rain

$$\left\langle \begin{array}{c} 0.0 \\ 1.0 \end{array} \right\rangle \quad \left\langle \begin{array}{c} 0.3 \\ 0.7 \end{array} \right\rangle \quad \left\langle \begin{array}{c} 0.48 \\ 0.52 \end{array} \right\rangle \quad \left\langle \begin{array}{c} 0.588 \\ 0.412 \end{array} \right\rangle \Rightarrow \left\langle \begin{array}{c} 0.75 \\ 0.25 \end{array} \right\rangle$$

$$\mathrm{P}(X_1) \quad \mathrm{P}(X_2) \quad \mathrm{P}(X_3) \quad \mathrm{P}(X_4) \quad \mathrm{P}(X_\infty)$$

From yet another initial distribution $\mathrm{P}(X_1)$:

$$\left\langle \begin{array}{c} p \\ 1-p \end{array} \right\rangle \quad \ldots \quad \Rightarrow \left\langle \begin{array}{c} 0.75 \\ 0.25 \end{array} \right\rangle$$

$$\mathrm{P}(X_1) \qquad \mathrm{P}(X_\infty)$$

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

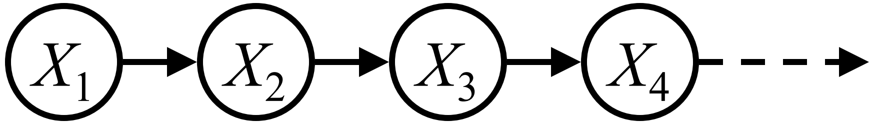
Stationary distribution:

- 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 = infinity?



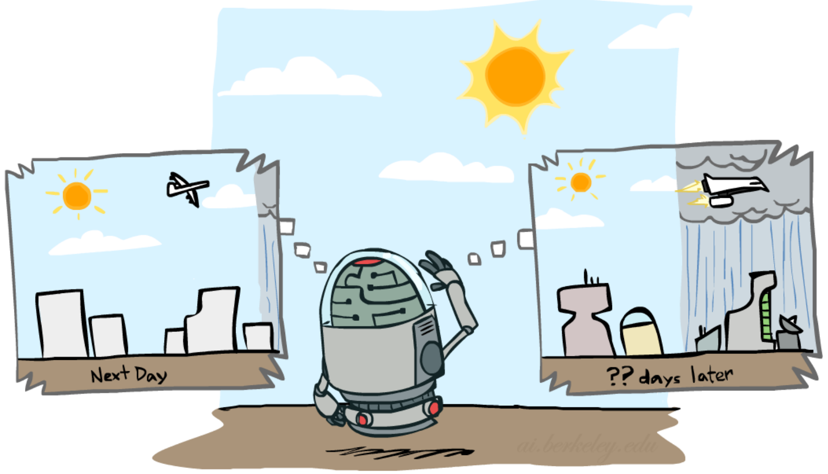
$$P_\infty(sun) = P(sun|sun)P_\infty(sun) + P(sun|rain)P_\infty(rain)$$
$$P_\infty(rain) = P(rain|sun)P_\infty(sun) + P(rain|rain)P_\infty(rain)$$

$$P_\infty(sun) = 0.9P_\infty(sun) + 0.3P_\infty(rain)$$
$$P_\infty(rain) = 0.1P_\infty(sun) + 0.7P_\infty(rain)$$

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

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

$$P_\infty(sun) = 3/4$$
$$P_\infty(rain) = 1/4$$



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

In-class exercise on Markov chains…

Let’s review

That's it for today!

- Next time: *Hidden Markov Models*

Some notes on the midterm

- In-class
- Closed book / laptop / calculator / notes / etc
- Probably **4-5 multi-part questions**, probing your understanding of the big approaches/models we've covered
- *Review in class next Tuesday*!