

CSE 573: Artificial Intelligence

Winter 2019

Markov Models & Hidden Markov Models

Hanna Hajishirzi

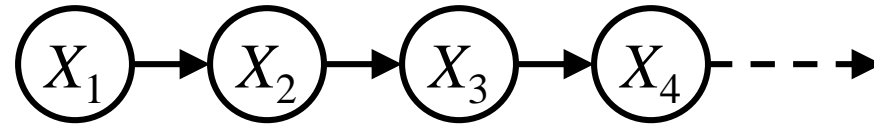
Many slides adapted from Pieter Abbeel, Dan Klein, Dan Weld, Stuart Russell, Andrew Moore & Luke Zettlemoyer

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

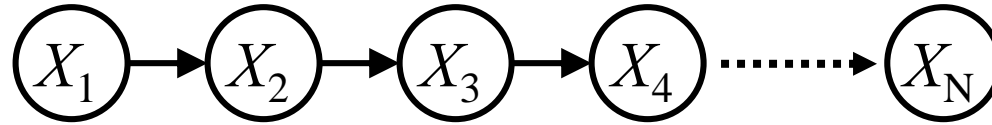
- Value of X at a given time is called the **state**



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

- 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

Markov Models (Markov Chains)



- A Markov model defines
 - a joint probability distribution:

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

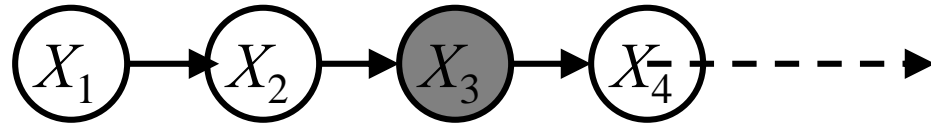
- More generally:

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

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

- One common inference problem:
 - Compute marginals $P(X_t)$ for all time steps t

Conditional Independence



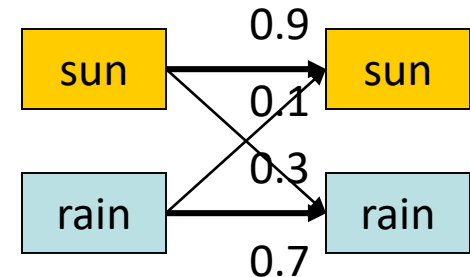
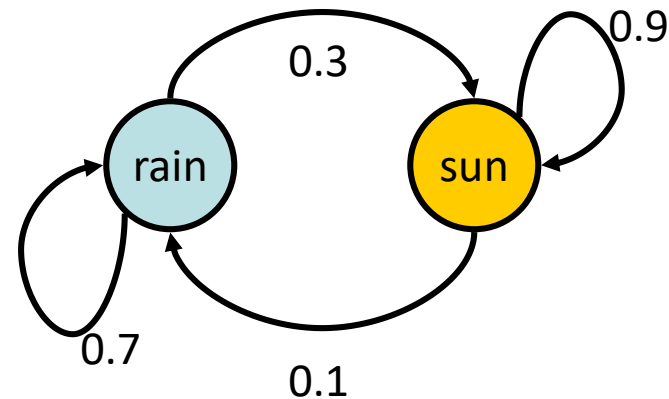
- **Basic conditional independence:**
 - Past and future independent of the present
 - Each time step only depends on the previous
 - This is called the (first order) Markov property

Example Markov Chain: Weather

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

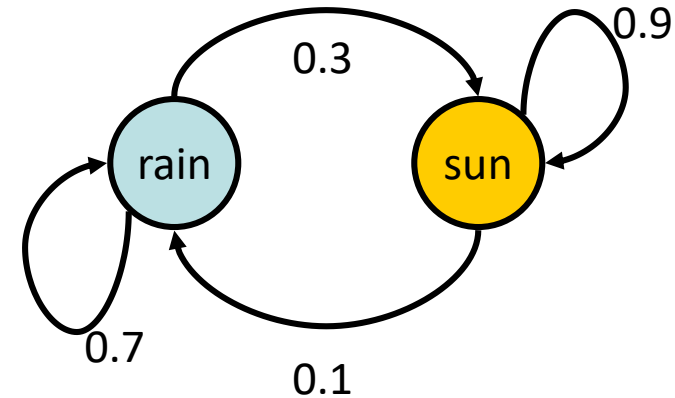
X_{t-1}	X_t	$P(X_t X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

Two new ways of representing the same CPT



Example Markov Chain: Weather

- Initial distribution: 1.0 sun



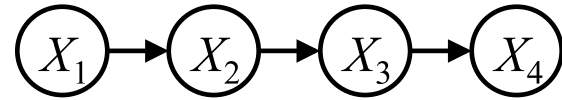
- What is the probability distribution after one step?

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

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

Chain Rule and Markov Models

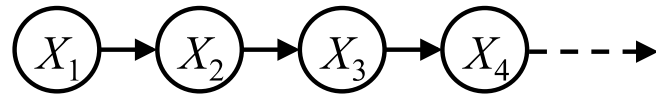
$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)$$



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

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

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 :

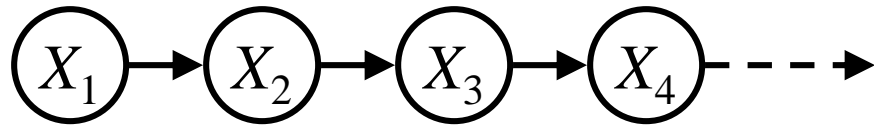
$$X_t \perp\!\!\!\perp X_1, \dots, X_{t-2} \mid X_{t-1}$$

gives us the expression posited on the earlier slide:

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

Mini-Forward Algorithm

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



$P(x_1)$ = 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

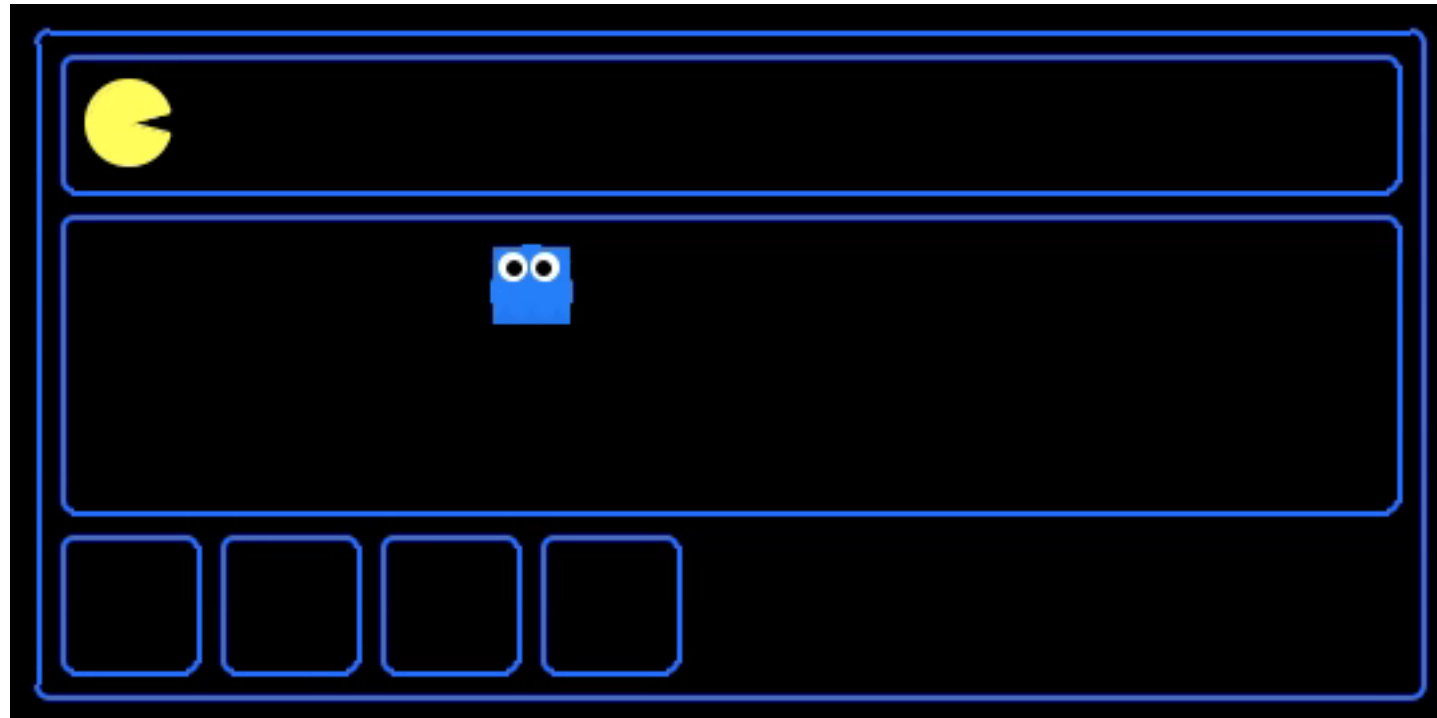
$$\begin{array}{ccccccc} \left\langle \begin{array}{c} 0.0 \\ 1.0 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.3 \\ 0.7 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.48 \\ 0.52 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.588 \\ 0.412 \end{array} \right\rangle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

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

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

Pac-man Markov Chain

Pac-man knows the ghost's initial position, but gets no observations!



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

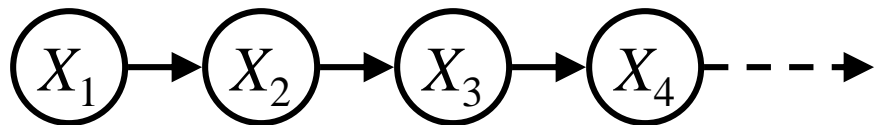
- Stationary distribution:

- 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 $t = \text{infinity}$?



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

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

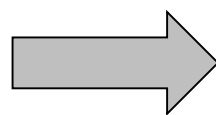
$$P_{\infty}(\text{sun}) = 0.9P_{\infty}(\text{sun}) + 0.3P_{\infty}(\text{rain})$$

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

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

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

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



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

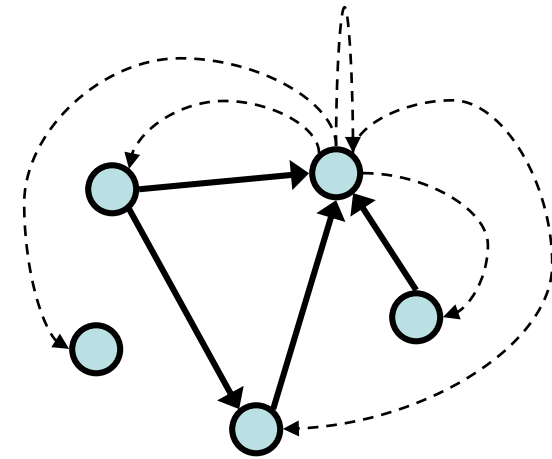
$$P_{\infty}(\text{rain}) = 1/4$$

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

