CSEP 573: Artificial Intelligence
Spring 2014

Markov Models

Ali Farhadi

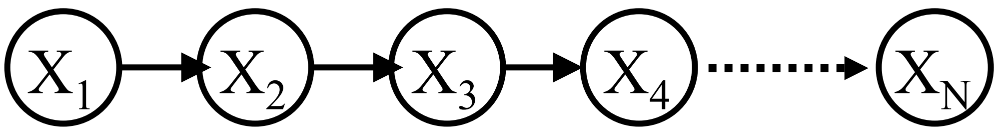
Many slides adapted from Pieter Abbeel, Dan Klein, Dan Weld,Stuart Russell, Andrew Moore & Luke Zettlemoyer

Markov Chains

- 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 (Markov Chains)

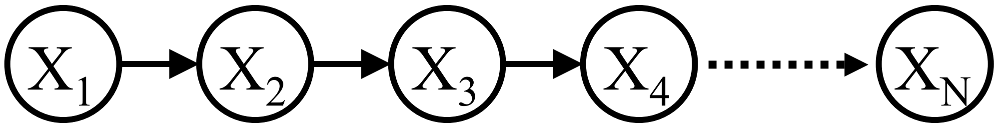
- A Markov model is:
  - a MDP with no actions (and no rewards)
  - Value of X is called the state



- A Markov model includes:
  - Random variables X_t for all time steps t (the state)
  - Parameters: called transition probabilities or dynamics, specify how the state evolves over time (also, initial probs)

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

Markov Models (Markov Chains)



- A Markov model defines
  - a joint probability distribution:

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

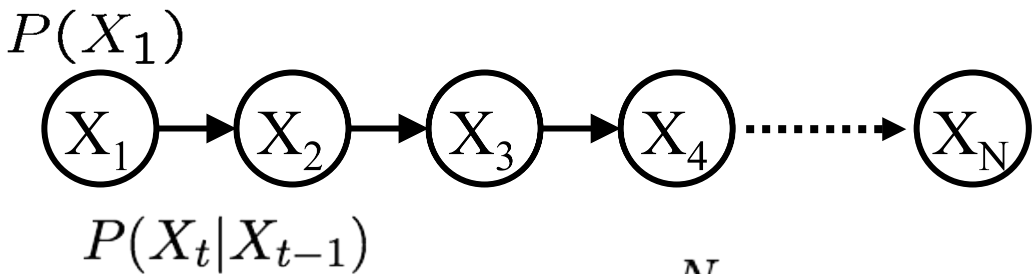
- More generally:

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

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

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

Markov Model



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

- 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

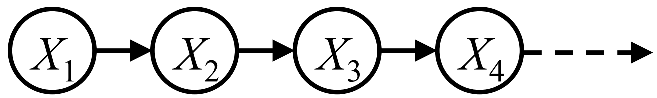
$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)$$

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

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

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

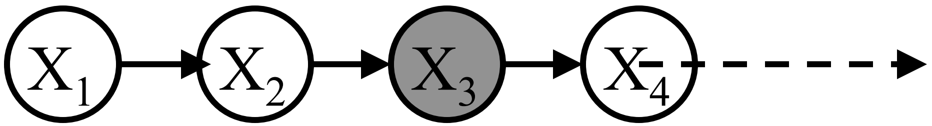
- Assuming that for all *t*:

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

gives us the expression posited on the earlier slide:

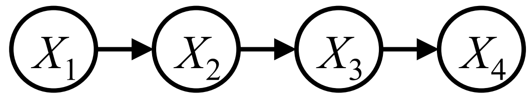
$$P(X_1, X_2, \ldots, X_T) = P(X_1) \prod_{t=2}^{T} P(X_t | X_{t-1})$$

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

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 $X_1 \perp\!\!\!\perp X_3, X_4 \mid X_2$?

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:

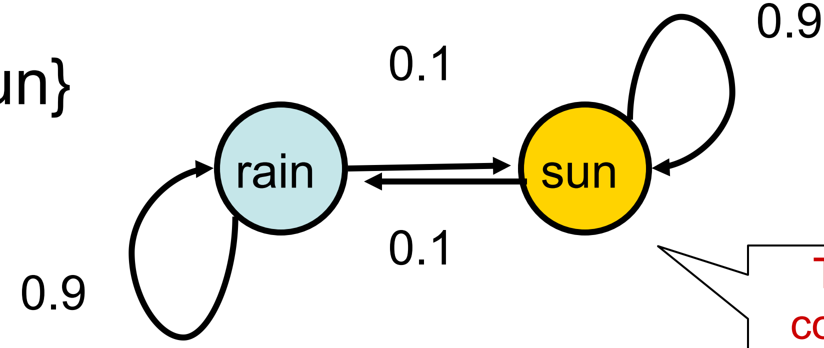
$$P(X_1, X_2, \ldots, X_T) = P(X_1)P(X_2|X_1)P(X_3|X_2)\ldots P(X_T|X_{T-1})$$
$$= P(X_1)\prod_{t=2}^{T} P(X_t|X_{t-1})$$

- 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 = {rain, sun}
  - Transitions:



This is a conditional distribution

  - Initial distribution: 1.0 sun
  - What's the probability distribution after one step?

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

$$0.9 \cdot 1.0 + 0.1 \cdot 0.0 = 0.9$$

Markov Chain Inference

- Question: probability of being in state x at time t?
- Slow answer:
  - Enumerate all sequences of length t which end in s
  - Add up their probabilities

$$P(X_t = sun) = \sum_{x_1 \ldots x_{t-1}} P(x_1, \ldots x_{t-1}, sun)$$

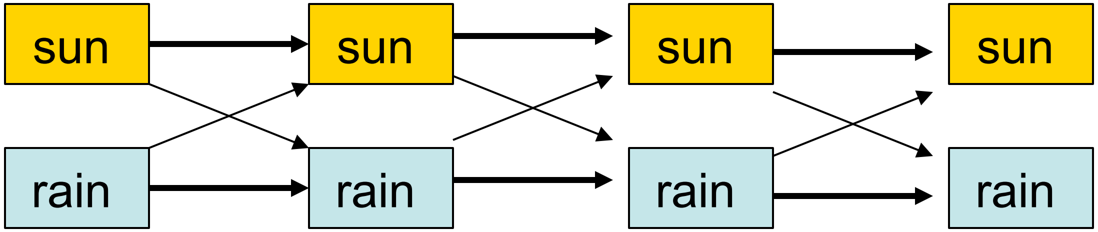
$$P(X_1 = sun)P(X_2 = sun|X_1 = sun)P(X_3 = sun|X_2 = sun)P(X_4 = sun|X_3 = sun)$$

$$P(X_1 = sun)P(X_2 = rain|X_1 = sun)P(X_3 = sun|X_2 = rain)P(X_4 = sun|X_3 = sun)$$

$$\vdots$$

Mini-Forward Algorithm

- Question: What's P(X) on some day t?
  - We don't need to enumerate every sequence!



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

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

Forward simulation

Example

- From initial observation of sun

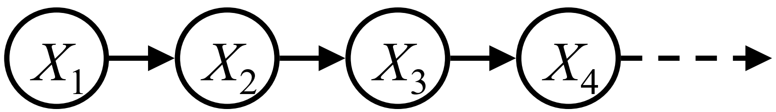
$$\underbrace{\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle}_{P(X_1)} \quad \underbrace{\left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle}_{P(X_2)} \quad \underbrace{\left\langle \begin{matrix} 0.82 \\ 0.18 \end{matrix} \right\rangle}_{P(X_3)} \quad \Longrightarrow \quad \underbrace{\left\langle \begin{matrix} 0.5 \\ 0.5 \end{matrix} \right\rangle}_{P(X_\infty)}$$

- From initial observation of rain

$$\underbrace{\left\langle \begin{matrix} 0.0 \\ 1.0 \end{matrix} \right\rangle}_{P(X_1)} \quad \underbrace{\left\langle \begin{matrix} 0.1 \\ 0.9 \end{matrix} \right\rangle}_{P(X_2)} \quad \underbrace{\left\langle \begin{matrix} 0.18 \\ 0.82 \end{matrix} \right\rangle}_{P(X_3)} \quad \Longrightarrow \quad \underbrace{\left\langle \begin{matrix} 0.5 \\ 0.5 \end{matrix} \right\rangle}_{P(X_\infty)}$$

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

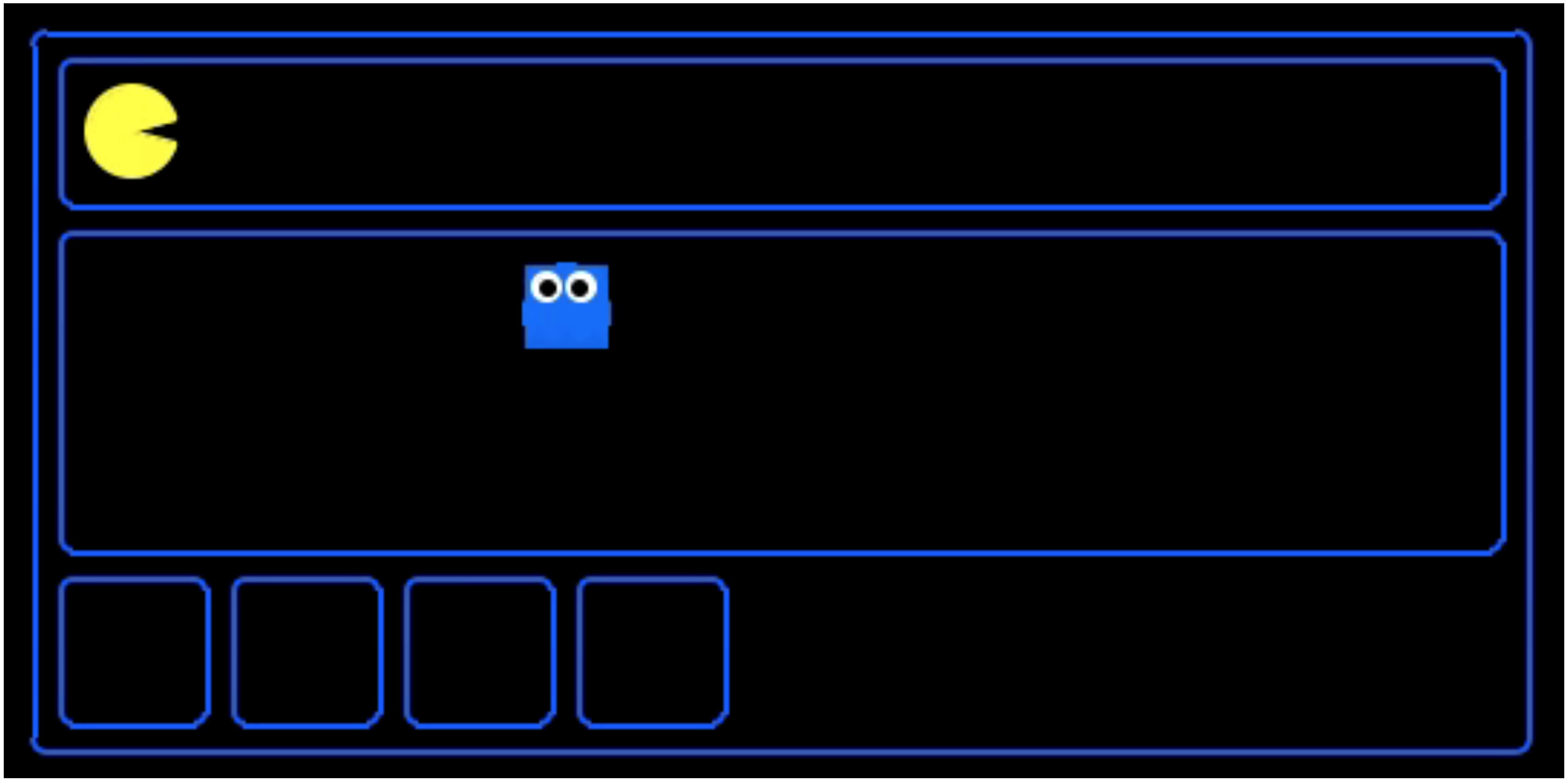
$$P_\infty(sun) = 3/4$$
$$P_\infty(rain) = 1/4$$

Stationary Distributions

- If we simulate the chain long enough:
  - What happens?
  - Uncertainty accumulates
  - Eventually, we have no idea what the state is!
- Stationary distributions:
  - For most chains, the distribution we end up in is independent of the initial distribution
  - Called the stationary distribution of the chain
  - Usually, can only predict a short time out

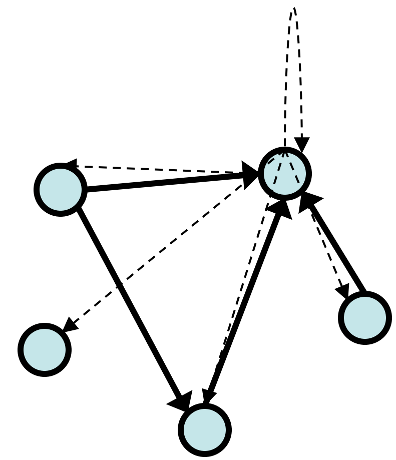
Pac-man Markov Chain

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



Web Link Analysis

- PageRank over a web graph
  - Each web page is a state
  - Initial distribution: uniform over pages
  - Transitions:
    - With prob. c, follow a random outlink (solid lines)
    - With prob. 1-c, uniform jump to a random page (dotted lines, not all shown)



- 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)