Un
Uncertainty

Markov Models

CSE 415: Introduction to Artificial Intelligence
University of Washington
Winter 2019

Most of the slides for this lecture were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All CS188 materials are available at http://ai.berkeley.edu.

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

sequence of random variables

probability distribution of our random variables give us our transition model

$X_1 \to X_2 \to X_3 \to X_4 \dashrightarrow$

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

- 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

RANDOM actions

transition distribution does NOT change over time; depends solely on current state

Joint Distribution of a Markov Model

Transition probabilities are ONLY a factor of our current state.

$X_1 \to X_2 \to X_3 \to X_4$

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

- 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:

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

special type of joint prob distribution

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

- 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?

NO! Many dependent events cannot be written this way

Chain Rule and Markov Models

X_i is a sequence of random variables that depend only on X_{i-1}, the random variable immediately preceeding it.
REPRESENTS THE STATE OF OUR SYSTEM OVER TIME

$X_1 \to X_2 \to X_3 \to X_4$

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

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|\underline{X_1}, X_2)P(X_4|\underline{X_1, X_2}, X_3)$$

- Assuming that

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

X3 is independent from X1 given I know X2

simplifies to 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

$X_1 \to X_2 \to X_3 \to X_4 \dashrightarrow$

- 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:

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

future is conditional independent of the past given the present (states)

simplifies to 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

$X_1 \to X_2 \to X_3 \to X_4$

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

Future events does NOT tell us any information about the preceeding state

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

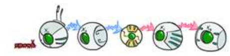
$$
\begin{aligned}
P(X_1 \mid X_2, X_3, X_4) &= \frac{P(X_1, X_2, X_3, X_4)}{P(X_2, X_3, X_4)} \\
&= \frac{P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_2)P(X_4 \mid X_3)}{\sum_{x_1} P(x_1)P(X_2 \mid x_1)P(X_3 \mid X_2)P(X_4 \mid X_3)} \\
&= \frac{P(X_1, X_2)}{P(X_2)} \\
&= P(X_1 \mid X_2)
\end{aligned}
$$

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:

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

- Implied conditional independencies:



Past independent of future given the present

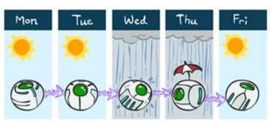
i.e., if $t_1 < t_2 < t_3$ then: $X_{t_1} \perp\!\!\!\perp X_{t_3} \mid X_{t_2}$

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

Conditional distribution (transition model)
does not change as our model goes on;
THIS MIGHT NOT BE TRUE IN GENERAL

Example Markov Chain: Weather

- States: X = {rain, sun}



- Initial distribution: 1.0 sun

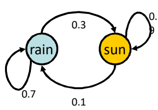
Where do we start?

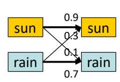
- CPT $P(X_t \mid X_{t-1})$:

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

yesterday

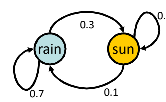
Two new ways of representing the same CPT





Example Markov Chain: Weather

- Initial distribution: 1.0 sun



- What is the probability distribution after one step?

$$
\begin{aligned}
P(X_2 = \text{sun}) = \;& P(X_2 = \text{sun}|X_1 = \text{sun})P(X_1 = \text{sun}) + \\
& P(X_2 = \text{sun}|X_1 = \text{rain})P(X_1 = \text{rain})
\end{aligned}
$$

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

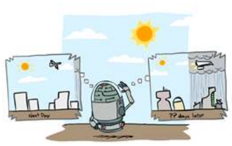
Way easier if we calculate this with the matrix algebra of markov chains.

With T = our transition matrix

Mini-Forward Algorithm

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

$X_1 \to X_2 \to X_3 \to X_4 \dashrightarrow$



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

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

Forward simulation

Example Run of Mini-Forward Algorithm

- From initial observation of sun

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

P(X_1) P(X_2) P(X_3) P(X_4) P(X_∞)

- From initial observation of rain

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

P(X_1) P(X_2) P(X_3) P(X_4) P(X_∞)

- From yet another initial distribution P(X_1):

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

P(X_1) P(X_∞)

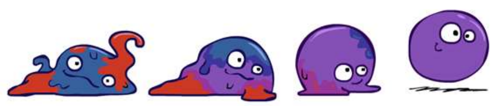
NOT guaranteed to converge.

Stationary Distributions

Convergence condition: markov chain must be irreducible; transition matrix T is power positive.
Then stationary distribution is unit eigenvector associated to eigenvalue of 1

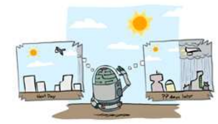
- For most chains:
  - Influence of the initial distribution gets less and less over time.
  - The distribution we end up in is independent of the initial distribution
- Stationary distribution: P_∞
  - The distribution we end up with is called the stationary distribution of the chain
  - It satisfies

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



Example: Stationary Distributions

- Question: What's P(X) at time t = infinity?



$$X_1 \to X_2 \to X_3 \to X_4 \dashrightarrow$$

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

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

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

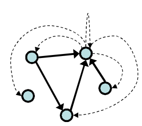
X_{t-1}	X_t	$P(X_t\|X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

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

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

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)