CSE 473: Artificial Intelligence
Markov Models

Dieter Fox — University of Washington

(These slides were created by Sam Sime and Peter Abbeel for CS188 team 1 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

  \[
P(X_1) \quad P(X_t | X_{t-1})
\]
- Parameters: called transition probabilities or dynamics, specify how the state evolves over time (i.e., 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:
  \[
P(X_1, X_2, \ldots, X_T) = P(X_1)P(X_2 | X_1)P(X_3 | X_2) \ldots P(X_T | X_{T-1})
  = P(X_1) \prod_{t=2}^{T} P(X_t | X_{t-1})
\]
- 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:

  \[
P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2 | X_1)P(X_3 | X_2)P(X_4 | X_3)
\]
- Assuming that

  \( X_3 \perp X_1 \mid X_2 \quad \text{and} \quad X_4 \perp X_2, X_3 \mid X_3 \)

simplifies 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:

  \[
P(X_1, X_2, \ldots, X_T) = P(X_1) \prod_{t=2}^{T} P(X_t | X_{t-1})
\]
- Assuming that for all \( t \):

  \( X_t \perp X_1, \ldots, X_{t-2} \mid X_{t-1} \)

simplifies 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

\[ X_3 \perp X_1 \mid X_2 \quad \text{and} \quad X_4 \perp X_1, X_2 \mid X_3 \]

Do we also have \( X_1 \perp X_3, X_4 \mid X_2 \)?

Yes!

Proof:

\[
P(X \mid X_1, X_2, X_3) = \frac{P(X)P(X_1)P(X_2)P(X_3)}{P(X_1)P(X_2)} = \frac{P(X_1)P(X_2)P(X_3)}{P(X_1)P(X_2)}
\]

\[
= \frac{P(X_3)}{P(X_2)}
\]

\[
P(X \mid X_1, X_3) = \frac{P(X)P(X_1)P(X_3)}{P(X_1)} = \frac{P(X_1)P(X_3)P(X_3)}{P(X_1)P(X_3)}
\]

\[
= \frac{P(X_3)}{P(X_1)}
\]

\[
P(X \mid X_1, X_4) = \frac{P(X)P(X_1)P(X_4)}{P(X_1)} = \frac{P(X_1)P(X_4)P(X_4)}{P(X_1)P(X_4)}
\]

\[
= \frac{P(X_4)}{P(X_1)}
\]

\[
P(X \mid X_3, X_4) = \frac{P(X)P(X_3)P(X_4)}{P(X_3)} = \frac{P(X_3)P(X_4)P(X_4)}{P(X_3)P(X_4)}
\]

\[
= \frac{P(X_4)}{P(X_3)}
\]

\[
P(X \mid X_2) = \frac{P(X)P(X_2)}{P(X_2)} = \frac{P(X_2)P(X_2)P(X_2)}{P(X_2)P(X_2)}
\]

\[
= \frac{P(X_2)}{P(X_2)}
\]

\[
P(X \mid X_3) = \frac{P(X)P(X_3)}{P(X_3)} = \frac{P(X_3)P(X_3)P(X_3)}{P(X_3)P(X_3)}
\]

\[
= \frac{P(X_3)}{P(X_3)}
\]

\[
P(X \mid X_4) = \frac{P(X)P(X_4)}{P(X_4)} = \frac{P(X_4)P(X_4)P(X_4)}{P(X_4)P(X_4)}
\]

\[
= \frac{P(X_4)}{P(X_4)}
\]

\[
P(X \mid X_1) = \frac{P(X)P(X_1)}{P(X_1)} = \frac{P(X_1)P(X_1)P(X_1)}{P(X_1)P(X_1)}
\]

\[
= \frac{P(X_1)}{P(X_1)}
\]

Markov Models Recap

- Explicit assumption for all \( t \): \( X_t \perp X_1, \ldots, X_{t-2} \mid X_{t-1} \)

- Consequence, joint distribution can be written as:

\[
P(X_t, X_{t-1}, \ldots, X_2) = P(X_t)P(X_{t-1})P(X_2|X_1) \ldots P(X_2|X_{t-1})
\]

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

- Implied conditional independencies:

  - Past independent of future given the present

  \( \text{i.e., if } t_1 < t_2 \text{ then: } X_{t_1} \perp X_{t_2} \mid X_{t_2} \)

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

Example Markov Chain: Weather

- States: \( X = \) [rain, sun]

- Initial distribution: 1.0 sun

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

\[
\begin{array}{c|ccc}
X_t & X_{t-1} & P(X_t) \\
\hline
\text{rain} & \text{sun} & 0.9 \\
\text{sun} & \text{sun} & 0.7 \\
\text{sun} & \text{rain} & 0.3 \\
\end{array}
\]

Mini-Forward Algorithm

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

\[
P(x_t) = \text{known}
\]

\[
P(x_t) = \sum_{x_{t-1}} P(x_{t-1}, x_t) = \sum_{x_{t-1}} P(x_t | x_{t-1}) P(x_{t-1})
\]

Example Run of Mini-Forward Algorithm

- From initial observation of sun

\[
\begin{array}{c|ccc}
P(X_1) & P(X_2) & P(X_3) & P(X_4) \\
\hline
1.0 & 0.9 & 0.84 & 0.804 & 0.75 \\
0.1 & 0.16 & 0.196 & 0.25 \\
\end{array}
\]

- From initial observation of rain

\[
\begin{array}{c|ccc}
P(X_1) & P(X_2) & P(X_3) & P(X_4) \\
\hline
1.0 & 0.3 & 0.5 & 0.52 & 0.412 & 0.75 \\
0.7 & 0.1 & 0.04 & 0.25 \\
\end{array}
\]

- From yet another initial distribution \( P(X_0) \):

\[
\begin{array}{c|ccc}
P(X_0) & P(X_1) & P(X_2) \\
\hline
1 & p & 0.75 \\
1 - p & 0.25 \\
\end{array}
\]

[Demo: L13D1,2,3]
Video of Demo Ghostbusters Basic Dynamics

Video of Demo Ghostbusters Circular Dynamics

Video of Demo Ghostbusters Whirlpool Dynamics

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 \( P_\infty \) of the chain.
  - It satisfies
    \[
    P_\infty(X) = P_\infty(x) = \sum_x P(x|\omega)P_\infty(\omega)
    \]

Example: Stationary Distributions

- Question: What’s \( P(X) \) at time \( t = \infty \)?

\[
\begin{align*}
P_\infty(\text{sun}) &= P(\text{sun}|\text{sun})P_\infty(\text{sun}) + P(\text{sun}|\text{rain})P_\infty(\text{rain}) \\
&= 0.9P_\infty(\text{sun}) + 0.3P_\infty(\text{rain}) \\
&P_\infty(\text{rain}) = 0.1P_\infty(\text{sun}) + 0.7P_\infty(\text{rain}) \\
P_\infty(\text{sun}) &= 0.3P_\infty(\text{sun}) \\
P_\infty(\text{rain}) &= 0.7P_\infty(\text{sun})
\end{align*}
\]

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

\[
\begin{align*}
P_\infty(\text{sun}) &= 3/4 \\
P_\infty(\text{rain}) &= 1/4
\end{align*}
\]

Application of Stationary Distribution: Web Link Analysis

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

- 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 2.0 returned the set of pages containing all your keywords in decreasing rank, now all search engines use link analysis along with many other factors.

- Example:
  - Uniform distribution over pages.
  - Transitions:
    - Uniform jump to a random page.
    - Follow a random outlink.

- 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 2.0 returned the set of pages containing all your keywords in decreasing rank, now all search engines use link analysis along with many other factors.

- Web Link Analysis:
  - Somewhat robust to link spam.
  - Google 2.0 returned the set of pages containing all your keywords in decreasing rank, now all search engines use link analysis along with many other factors.