Markov Models

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

\[
\begin{align*}
X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow \cdots \\
P(X_1) \quad P(X_1|X_{t-1})
\end{align*}
\]

- 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

\[
\begin{align*}
P(X_1, X_2, X_3, X_4) &= P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3) \\
&= P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)
\end{align*}
\]

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

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

simplifies to the expression posited on the previous slide:

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

- 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 to the expression posited on the earlier slide:

\[
P(X_1, X_2, \ldots, X_T) = P(X_1)\prod_{t=2}^{T} P(X_t|X_{t-1})
\]
Implicated Conditional Independencies

- We assumed: $X_3 \perp X_1 \mid X_3$ and $X_4 \perp X_1, X_2 \mid X_3$
- Do we also have $X_1 \perp X_3, X_4 \mid X_2$?
  - Yes!
  - Proof:
    $$P(X_1, X_2, X_3, X_4) = \frac{P(X_4, X_3, X_2, X_1)}{P(X_3, X_2, X_4)} \frac{P(X_1, X_3, X_2)}{P(X_3, X_2)}$$
    $$= \frac{P(X_4, X_3, X_2, X_1)}{P(X_3, X_2)} = P(X_1 \mid X_2)$$

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:
  $$P(X_1, X_2, \ldots, X_T) = P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1, X_2) \ldots P(X_T \mid X_{T-1})$$
  $$= P(X_1) \prod_{t=2}^T P(X_t \mid X_{t-1})$$
- Implied conditional independencies:
  Past independent of future given the present
  i.e., if $t_1 < t_2$ 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: $X = \{\text{rain, sun}\}$
- Initial distribution: 1.0 sun
- CPT $P(X_t \mid X_{t-1})$:

<table>
<thead>
<tr>
<th>$X_{t-1}$</th>
<th>$X_t$</th>
<th>$P(X_t \mid X_{t-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>sun</td>
<td>0.9</td>
</tr>
<tr>
<td>sun</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>rain</td>
<td>sun</td>
<td>0.3</td>
</tr>
<tr>
<td>rain</td>
<td>rain</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Two new ways of representing the same CPT

What is the probability distribution after one step?

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

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

Mini-Forward Algorithm

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

  $$P(X_1) = \text{known}$$
  $$P(X_t) = \sum_{x_{t-1}} P(x_{t-1}, x_t)$$
  $$= \sum_{x_{t-1}} P(x_{t-1})P(x_t \mid x_{t-1})$$

Example Run of Mini-Forward Algorithm

- From initial observation of sun
  $$\begin{array}{c|c|c|c|c|c}
  t & 1.0 & 0.9 & 0.84 & 0.804 & 0.75 \\
  P(X_t) & P(X_{t-1}) & P(X_{t-2}) & P(X_{t-3}) & P(X_{t-4}) \\
  \end{array}$$
- From initial observation of rain
  $$\begin{array}{c|c|c|c|c|c}
  t & 1.0 & 0.8 & 0.48 & 0.588 & 0.75 \\
  P(X_t) & P(X_{t-1}) & P(X_{t-2}) & P(X_{t-3}) & P(X_{t-4}) \\
  \end{array}$$
- From yet another initial distribution $P(X_1)$
  $$\begin{array}{c|c}
  t & 0.75 \\
  \end{array}$$
  $$\begin{array}{c|c|c|c|c}
  t & 1 - p & 0.25 \\
  P(X_t) & P(X_{t-1}) & P(X_{t-2}) & P(X_{t-3}) & P(X_{t-4}) \\
  \end{array}$$
Stationary distribution:
The distribution we end up with is called the stationary distribution of the chain.

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.

Example: Stationary Distributions

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

Stationary distribution:
The distribution we end up with is independent of the initial distribution.

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
- Will spend less time on less reachable pages
- Will spend the same amount of time on pages with equal reachability

Will spend more time on highly reachable pages.
Will spend less time on less reachable pages.
Will spend the same amount of time on pages with equal reachability.

Also:
- 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
- Will spend less time on less reachable pages
- Will spend the same amount of time on pages with equal reachability

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).