Markov Networks

Doug Downey
Northwestern EECS 395/495 Fall 2014
I-Maps, Perfect Maps, and I-Equivalence

- **I-Map for S**: A graph containing at most a set $S$ of independence assertions, i.e. statements of the form $(X \perp Y | Z)$.
- E.g., some I-Maps for $S = \{(A \perp B | C)\}$.
I-Maps: why they matter

- If $G$ is an I-Map for the independences in a distribution $P$, then we can represent $P$ as a Bayes Net with graph $G$.
  - Whereas we can’t do so if $G$ is not an I-Map for $P$

- A given distribution may have many different I-Maps
  - Minimal I-Map for $S$: An I-Map for $S$ for which the removal of any edge renders it not an I-Map for $S$
  - Perfect Map for $S$: A graph with exactly the set of independencies in $S$
Example

- Two Perfect Maps for $S = \{(A \perp B \mid C)\}$

```
A → C → B
```

```
A ← C ← B
```
I-Equivalence (1 of 2)

- Two graphs are *I-Equivalent* if they imply identical sets of independence assertions.

- **I-equivalent**
  - Graph 1: A → C → B
  - Graph 2: A → C → B, with an additional edge A → B

- **Not I-equivalent**
  - Graph 1: A → C → B
  - Graph 2: A → C, B → C
Two graphs are I-Equivalent iff they have the same

- **Skeleton**: graph ignoring edge direction
- **Immoralities**: v-structures without direct edge between parents
Naïve Bayes Net

- NB assumes features conditionally independent given the class:

  - P(Spam=true) = 0.3
  - P(Spam=true | "lottery") = 0.04
  - P(Spam=true | "with") = 0.6
  - P(Spam=true | "dear") = 0.24
  - P(Spam=false | "lottery") = 0.01
  - P(Spam=false | "with") = 0.59
  - P(Spam=false | "dear") = 0.30
Limitations of Bayesian Networks

- Perfect Map for \{(A \perp B \mid C, D), (C \perp D \mid A, B)\}?

- Not possible! Bayes Nets can’t express all possible sets of independence assertions.
Alternative: Markov Networks

- Undirected Graphical Model
  - No CPTs. Uses potential functions $\phi_c$ defined over cliques
  - $P(x) = \prod_c \phi_c(x_c) \quad \frac{Z}{Z} = \sum_x \prod_c \phi_c(x_c)$

<table>
<thead>
<tr>
<th>Grades</th>
<th>TV</th>
<th>$\phi_1(G, TV)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>bad</td>
<td>none</td>
<td>2.0</td>
</tr>
<tr>
<td>good</td>
<td>none</td>
<td>3.0</td>
</tr>
<tr>
<td>bad</td>
<td>lots</td>
<td>3.0</td>
</tr>
<tr>
<td>good</td>
<td>lots</td>
<td>1.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TV</th>
<th>Trivia Knowledge</th>
<th>$\phi_2(TV, K)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>weak</td>
<td>2.0</td>
</tr>
<tr>
<td>lots</td>
<td>weak</td>
<td>1.0</td>
</tr>
<tr>
<td>none</td>
<td>strong</td>
<td>1.5</td>
</tr>
<tr>
<td>lots</td>
<td>strong</td>
<td>3.0</td>
</tr>
</tbody>
</table>
## Markov Net Joint Distribution

<table>
<thead>
<tr>
<th>Grades</th>
<th>TV</th>
<th>Trivia Know.</th>
<th>(\phi_1(G, TV))</th>
<th>(\phi_2(TV, K))</th>
<th>(\phi_1(G, TV) \times \phi_2(TV, K))</th>
<th>(P(G, TV, K))</th>
</tr>
</thead>
<tbody>
<tr>
<td>bad</td>
<td>none</td>
<td>weak</td>
<td>2.0</td>
<td>2.0</td>
<td>4.0</td>
<td>0.12</td>
</tr>
<tr>
<td>good</td>
<td>none</td>
<td>weak</td>
<td>3.0</td>
<td>2.0</td>
<td>6.0</td>
<td>0.18</td>
</tr>
<tr>
<td>bad</td>
<td>lots</td>
<td>weak</td>
<td>3.0</td>
<td>1.0</td>
<td>3.0</td>
<td>0.09</td>
</tr>
<tr>
<td>good</td>
<td>lots</td>
<td>weak</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>0.03</td>
</tr>
<tr>
<td>bad</td>
<td>none</td>
<td>strong</td>
<td>2.0</td>
<td>1.5</td>
<td>3.0</td>
<td>0.09</td>
</tr>
<tr>
<td>good</td>
<td>none</td>
<td>strong</td>
<td>3.0</td>
<td>1.5</td>
<td>4.5</td>
<td>0.13</td>
</tr>
<tr>
<td>bad</td>
<td>lots</td>
<td>strong</td>
<td>3.0</td>
<td>3.0</td>
<td>9.0</td>
<td>0.27</td>
</tr>
<tr>
<td>good</td>
<td>lots</td>
<td>strong</td>
<td>1.0</td>
<td>3.0</td>
<td>3.0</td>
<td>0.09</td>
</tr>
</tbody>
</table>

\(Z = 33.5\)
Instead of D-separation, simply graph separation

So (Grades \perp Trivia Knowledge \mid TV)
Expressivity of Markov Networks

- Perfect Map for \{(A \perp B \mid C, D), (C \perp D \mid A, B)\}?
Expressivity of Markov Networks

- Perfect Map for \( \{(A \perp B \mid C, D), (C \perp D \mid A, B)\} \)?
Expressivity of Markov Networks

- Perfect Map for \{ (A \perp B \mid C, D), (C \perp D \mid A, B) \}?

- Markov Nets can capture these independence assertions
But...

- How about $(A \perp C) \in S$, but $(A \perp C \mid B) \notin S$?

- Can’t be captured perfectly in Markov Networks
- If graph separation $\rightarrow$ conditional independence, new knowledge can only **remove** dependencies
Bayesian Networks => Markov Networks

- Markov Nets can encode independences that Bayes Nets cannot, and vice-versa
- To convert from BN to MN, “moralize”:

![Diagram of a Bayesian Network with nodes A, B, and C connected by arrows representing independences.]
Bayesian Networks => Markov Networks

- Markov Nets can encode independences that Bayes Nets cannot, and vice-versa
- To convert from BN to MN, “moralize”: 

![Diagram](image-url)
Markov Net Applications

- Best when no clear, directed causal structure
  - E.g. statistical physics, text, social networks, image analysis (e.g. segmentation, below)