Bayes Nets Representation: joint distribution and conditional independence

Yi Zhang
10-701, Machine Learning, Spring 2011
February 9\textsuperscript{th}, 2011

Parts of the slides are from previous 10-701 lectures
Outline

- Conditional independence (C. I.)
- Bayes nets: overview
- Local Markov assumption of BNs
- Factored joint distribution of BNs
- Infer C. I. from factored joint distributions
- D-separation (motivation)
Conditional independence

- X is conditionally independent of Y given Z

\[(\forall x, y, z) P(X = x \mid Y = y, Z = z) = P(X = x \mid Z = z)\]

- In short:

\[P(X \mid Y, Z) = P(X \mid Z)\]

- Equivalent to:

\[P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)\]
Bayes nets

- Bayes nets: directed acyclic graphs express sets of conditional independence via graph structure
  - All about the joint distribution of variables!
  - Conditional independence assumptions are useful
  - Naïve Bayes model is an extreme example
Three key questions for BNs

- **Representation:**
  - *What joint distribution does a BN represent?*

- **Inference**
  - How to answer questions about the joint distribution?
    - Conditional independence
    - Marginal distribution
    - Most likely assignment

- **Learning**
  - How to learn the graph structure and parameters of a BN from data?
Local Markov assumptions of BNs

- A variable $X$ is independent of its non-descendants given (only) its parents
  - Intuition: “flu” and “allergy” causes “headache” only through “sinus”
Local Markov assumptions of BNs

- A variable $X$ is independent of its non-descendants given (only) its parents

<table>
<thead>
<tr>
<th>parents</th>
<th>non-desc</th>
<th>assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>F, A</td>
<td>-</td>
</tr>
<tr>
<td>H</td>
<td>S</td>
<td>F, A, N</td>
</tr>
<tr>
<td>N</td>
<td>S</td>
<td>F, A, H</td>
</tr>
<tr>
<td>F</td>
<td>-</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>-</td>
<td>F</td>
</tr>
</tbody>
</table>

$F \perp A$, $H \perp \{F, A\}|S$, $N \perp \{F, A, H\}|S$
Local Markov assumptions of BNs

- Local Markov assumptions only express a *subset* of C.I.s on a BN
  - Is $X_M$ conditionally independent of $X_1$ given $X_2$?

- But they are sufficient to infer all others
Factored joint distribution of a BN

- A BN can represent the joint distributions of the following form:

\[
p(x) = \prod_{k=1}^{K} p(x_k | p_{a_k})
\]
Factored joint distribution of a BN

- A BN can represent the joint distributions of the following form:

\[ p(x) = \prod_{k=1}^{K} p(x_k | p_a_k) \]

\[
P(F, A, S, H, N) = P(F) P(A) P(S | F, A) P(H | S) P(N | S)
\]
Factored joint distribution of a BN

- Local Markov assumptions imply the factored joint distribution

\[
P(F, A, S, H, N) = P(F) P(F|A) P(S|F,A) P(H|S,F,A) P(N|S,F,A,H)
\]

**Chain rule**

\[
P(F) P(A) P(S|F,A) P(H|S) P(N|S)
\]

**Markov Assumption**

\[F \perp A, \quad H \perp \{F,A\}|S, \quad N \perp \{F,A,H\}|S\]
Factored joint distribution of a BN

- Naïve Bayes
  - Local Markov assumptions: $X_i \perp X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n \mid Y$
  - Factored joint distribution:

$$P(X_1, \ldots, X_n, Y) = P(Y)P(X_1 \mid Y)\ldots P(X_1 \mid Y)$$
Infer C.I. from the factored joint distribution

- We already see: local Markov assumptions $\rightarrow$ factored joint distribution
- Also, factored joint distribution $\rightarrow$ all C.I. in the BN
Infer C.I. from the factored joint distribution

- Factored Joint distribution

\[ p(a, b, c) = p(a|c)p(b|c)p(c) \]

- Show that \( a \independent b \mid c \)

\[
\begin{align*}
p(a, b|c) &= \frac{p(a, b, c)}{p(c)} \\
&= \frac{p(a|c)p(b|c)p(c)}{p(c)} \\
&= p(a|c)p(b|c)
\end{align*}
\]
Infer C.I. from the factored joint distribution

- Factored Joint distribution

\[ p(a, b, c) = p(a|c)p(b|c)p(c) \]

- Do we have \( a \perp b \) ? In general, no.

\[
p(a, b) = \sum_c p(a, b, c) = \sum_c p(a|c)p(b|c)p(c)
\]

- Cannot be written into two separate terms of \( a \) and \( b \)
D-separation: motivation

- Is $X_M$ conditionally independent of $X_1$ given $X_2$?
  - Intuitively yes: $X_1$ affects $X_M$ only through $X_2$.
  - Method I: using factored joint distribution to derive
    $$p(x_1, x_M | x_2) = \frac{p(x_1, x_2, x_M)}{p(x_2)} = \frac{\sum_{x_3, x_4, \ldots, x_{M-1}} p(x_1, x_2, \ldots, x_M)}{\sum_{x_1, x_3, x_4, \ldots, x_{M-1}, x_M} p(x_1, x_2, \ldots, x_M)}$$
  - Method II: D-separation 😊 --- not today