Chain Rule and CI
Bayesian Networks
Ordering of Variables and Causality

ORIE 4742
Mar 22, 2021
Spencer Peters

Reading off CI's - d-separation
BN as convenient notations for complex models
Chain rule and conditional probabilities

\[ P(x_1, x_2, x_3) = P(x_1) P(x_2 | x_1) P(x_3 | x_1, x_2) \]
\[ \text{indep} \implies = P(x_1) P(x_2) P(x_3) \]

\[ P(x_1, x_2, x_3) = P(x_1) P(x_2 | x_1) P(x_3 | x_1, x_2) \]
\[ x_3 \perp x_1 \mid x_2 \implies = P(x_1) P(x_2 | x_1) P(x_3 | x_2) \]
Conditional independence

\[ P(x_1, x_2, x_3) = P(x_1) P(x_2 | x_1) P(x_3 | x_1, x_2) \]

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.4 0.3</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.6 0.7</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\begin{array}{c}
\text{X}_1 = 1 \\
\text{X}_2 = 1 \\
\text{X}_3 = 0
\end{array}
\end{align*}
\]

\[ P((1, 1, 0)) = 0.9 \times 0.7 \times 0.2 \]
Bayesian Networks (BNs)

Bayes Net $B$ is a directed acyclic graph (the nodes $N$ are variables) and conditional probabilities (CPT) $P(X_i \mid Pa(X_i))$ for each $X_i \in N$. 

$$P(x_1, x_2, x_3) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1, x_2)$$
Rain or Sprinkler?

W: Grass wet
R: raining
S: sprinklers on
C: cloudy

\[
P(W|S, C) > P(W|S) \]
\[
P(W|S, C, R) = P(W|S, R) \]
\[
W \perp C | S, R \]
Ordering of variables in BNs and causality (1/2)

\[ P(X=0) = 0.05 \]

\[ P(X=0) = 0.05 \]

in both cases.

\[ P(\text{checks}) = 0.1 \]

\[ 0.5 \times 0.1 \]

\[ 0.5 \times 0.1 \]
Ordering of variables in BNs and causality (1/2)

\[ \begin{align*}
X &= \text{coin flip } \{0, 1\} \\
Y &= \text{student's report } \{0, 1\}
\end{align*} \]

\[ P(x, y) = P(x) P(y | x) \]

\[ P(x, y) = P(y) P(x | y) \]
Ordering of variables in BNs and causality (2/2)
Reading off a BN’s conditional independences (d-separation)

\[
P(x, y) = P(x) \cdot P(y)
\]
Chains

\[ X \leftrightarrow Y \leftrightarrow Z \]

\[ P(z \mid x, y) = P(z \mid y) \iff z \perp \!
\perp x \mid y \]

\[ P(x \mid y, z) = P(x \mid y) \iff x \perp \!
\perp z \mid y \]
Forks

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

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

\[
P(S=1 \mid W=1, R=0) = 1
\]

\[
P(S=1 \mid W=1) = \frac{2}{3}
\]
Colliders/Joins

\[ P(S=1 | W=1) = 1 \]

\[ P(W=1 | S=1 \text{ or } R=1) = 1 \]

\[ P(W=1 | S=0 \text{ and } R=0) = 0 \]
d-Separation
d-Separation

T, Y | 1 × ?
False
M, X | Z

U → V → Y
Z
BNs as notation: Regression

\[(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\]

\[y_i = Ax_i + B + \varepsilon_i\]

\[\varepsilon_i \sim N(0, \sigma^2)\]

\[A \sim N(3, 1^2)\]

\[B \sim N(0, 100^2)\]
BNs as notation: Regression

\[(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\]

\[y_i = A x_i + B + \varepsilon_i\]
BNs as notation: Prediction

\[ \hat{y} = A \mathbf{x} + B + \mathbf{N}(0, \sigma^2) \]
Categories $C_1, \ldots, C_m$

$Y_i \sim \text{Dirichlet} (\beta_1, \ldots, \beta_c)$