10701 Recitation: Bayesian Net (D-Seperation) & BN Inference

Min Chi

Nov 11\textsuperscript{th}, 2010

NSH 1507
Graphical Models

• Probabilistic graphical models are graphs in which nodes represent random variables, and the (lack of) arcs represent conditional independence assumptions.
  – Directed graphical models, such as Bayesian Networks or Belief Networks (BNs)
  – Undirected graphical models.
  – Chain graph: both directed and undirected arcs.
Bayesian Network

- A probabilistic graphical model that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG).
- A DAG is a collection of vertices (random variables) and directed edges (i.e. Conditional Dependencies), each edge connecting one vertex to another, such that there is no way to start at some vertex \( v \) and follow a sequence of edges that eventually loops back to \( v \) again.
Bayesian networks

- A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

- Syntax:
  - a set of nodes, one per variable
  - a directed, acyclic graph (link ≈ "directly influences")
  - In addition to the graph structure, it is necessary to specify the parameters of the model. For a directed model, we must specify the Conditional Probability Distribution (CPD) at each node. If the variables are discrete, this can be represented as a table (CPT), which lists the probability that the child node takes on each of its different values for each combination of values of its parents: \( P(X_i | \text{Parents}(X_i)) \)

- In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over \( X_i \) for each combination of parent values
An Example

• Four random variables, each has possible values denote by T (true) and F (false):
  – Cloudy (C)
  – SprinklerOn (S)
  – Rain (R)
  – Wet Grass (W)

• The full joint distribution for the four variable require $2^n - 1$ parameter (15).
• By the chain rule of probability, the joint probability of all the four random variable:
  \[ P(C, S, R, W) = P(C) \times P(S|C) \times P(R|C,S) \times P(W|C,S,R) \]

• By using conditional independence relationships:
  \[ P(C, S, R, W) = P(C) \times P(S|C) \times P(R|C) \times P(W|S,R) \]
  
  \( (1+2+2+4) = 9 \) parameters.

An Example
An Example

<table>
<thead>
<tr>
<th>C</th>
<th>P(S=F)</th>
<th>P(S=T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>T</td>
<td>0.9</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S</th>
<th>R</th>
<th>P(W=F)</th>
<th>P(W=T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>F</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>0.01</td>
<td>0.99</td>
</tr>
</tbody>
</table>
Another Example

\[ P(C,A,R,E,B) = P(B)P(E|B)P(R|E,B)P(A|R,B,E)P(C|A,R,B,E)1+2+4+8+16 = 31 \]

versus

\[ P(C,A,R,E,B) = P(B)P(E)P(R|E)P(A|B,E)P(C|A)1+1+2+4+2 = 10 \]

• For binary variable,
• For \( n \) nodes, the full joint would require \( (2^n - 1) \) parameters.
  • \( n=1000 \quad 2^{1000}-1 \)
• For BN, the maximum number of parameters need to learn \( n \times 2^k \), where \( k \) is the maximum number of parents nodes in a graph.
  • \( n=1000, \quad k=5, \quad 32000 \)
Causal Markov assumption

• Bayesian networks encode the dependencies and independencies between variables.

• Under the causal Markov assumption, each variable in a Bayesian network is independent of its ancestors given the values of its parents.
D-separation

• D-separation is about more general conditional independence in a BN.

• Three Patterns for paths through a node in a Bayesian network.
• In a serial connection from X1 to X3 via X2, evidence from X1 to X3 is **blocked** only when we have hard evidence about X2.

• Intermediate cause.
A diverging connection

- In a diverging connection where $X_1$ and $X_3$ have the common parent $X_2$, evidence from $X_1$ to $X_3$ is blocked only when we have hard evidence about $X_2$.
- Common cause.
• In a converging connection where \( X_2 \) has parents \( X_1 \) and \( X_3 \), any evidence about \( X_2 \) results in evidence transmitted between \( X_1 \) and \( X_3 \).

• Common Effect.
d-separation

• Definition: If $X_1$, $X_2$ and $X_3$ are three disjoint subsets of nodes in a DAG, then $E$ is said to d-separate $X_1$ from $X_3$ if **every undirected path** from $X_1$ to $X_3$ is blocked by $X_2$. A path is blocked if it contains a node $Z$ such that:

• (1) $Z$ has one incoming and one outgoing arrow and $Z$ is in $X_2$; or
• (2) $Z$ has two outgoing arrows and $Z$ is in $X_2$; or
• (3) $Z$ has two incoming arrows and neither $Z$ nor any of its descendants is in $X_2$. 
Path

• Intuition: dependency must “flow” along paths in the graph.
• A path is a sequence of neighboring variables.

• Examples:
  R ← E → A ← B
  C ← A ← E → R
Example 1

- $d$-sep$(R, B)$?
Example 1

- d-sep(X1, X3 | X2)
- d-sep(R, B)?
  - X1 = {R}, X3 = {B}, X2 = {}
  - Find all the path between R, B
  - Check the node:
    - Earthquake.
      (diverging, not in X2). Not blocking.
    - Alarm
      (Converging, A or C are not in X2). Block!
Example 1:

• $d$-sep$(R, B)$? YES.
Example 2

- $d$-sep$(R,B \mid A)$?
Example 2

- d-sep(X1, X3|X2)
- d-sep(R,B|A)?
  - X1 = \{R\}, X3={B}, X2 ={A}
  - Find all the path between R, B
  - Check the node:
    - Earthquake.
      (diverging, not in X2). Not blocking.
    - Alarm
      (Converging, A or C are IN X2). Not blocking!
Example 2

- $d$-sep($R, B | A$)? No.
Example 3

- $d$-sep($R,B | E,A$)?
Example 3

- $d$-sep($X_1$, $X_3 | X_2$)
- $d$-sep($R$, $B | E$, $A$)?
  - $X_1 = \{R\}$, $X_3 = \{B\}$, $X_2 = \{E,A\}$
  - Find all the path between $R$, $B$
  - Check the node:
    - Earthquake.
      (diverging, IN $X_2$). Blocking!
    - Alarm
      (Converging, $A$ or $C$ are IN $X_2$). Not blocking.
Example 3

• $d$-sep($R,B \mid E,A$)? Yes.
Example 4

- $d$-sep($\text{Radio}(R), \text{Gas}(G) | \text{Ignition}(I))$?
Example 4

- \( d\text{-sep}(\text{Radio}(R), \text{Gas}(G) | \text{Ignition}(I)) \)? Yes.

- \( P(R, G | I) = \)
- \( P(G | I, R) = \)
Example 4

• d-sep(Radio(R), Gas(G) | Ignition(I))? Yes.

• \( P(R, G/I) = P(R/I) \cdot P(G/I) \)

• \( P(G/I, R) = P(G/I) \)
Example 5

• $d$-sep($\text{Radio}(R), \text{Gas}(G) | \text{Battery}$)? Yes.

\[ P(R/B, G) = P(R/B) \]

\[ P(G/B, R) = P(G/B) \]
More d-seperation

• *Gas and Radio* are independent given no evidence at all. \( P(\text{Gas}/\text{Radio})=P(\text{Gas}); P(\text{Radio}/\text{Gas})=P(\text{Radio}) \)

• *Gas and Radio* are dependent given evidence about whether the car *Starts*. *For example, if* the car does not start, then the radio playing is increased evidence that we are out of gas. \( P(\text{Gas}/\text{Radio, Start}) \neq P(\text{Gas}/\text{Start}) \)

• *Gas and Radio* are also dependent given evidence about whether the car *Moves*, because that is enabled by the car starting. \( P(\text{Gas}/\text{Radio, Move}) \neq P(\text{Gas}/\text{Move}) \)
Re-visit d-separation

• Definition: If $X_1$, $X_2$ and $X_3$ are three disjoint subsets of nodes in a DAG, then $E$ is said to d-separate $X_1$ from $X_3$ if every undirected path from $X_1$ to $X_3$ is blocked by $X_2$. A path is blocked if it contains a node $Z$ such that:
  
  • (1) $Z$ has one incoming and one outgoing arrow and $Z$ is in $X_2$; or
  
  • (2) $Z$ has two outgoing arrows and $Z$ is in $X_2$; or
  
  • (3) $Z$ has two incoming arrows and neither $Z$ nor any of its descendants is in $X_2$. 
D-separation on Sets

• $d$-sep($\{Q\}, \{X, Y, Z, P\} | W)$?
D-seperation on Sets

• $d$-sep($\{Q\}$, $\{X, Y, Z, P\}$|W)? Yes!

Blue path is a divergent path that is closed since we condition on W.
D-seperation on Sets

• \( d\text{-}sep(\{Z\}, \{X, W, Q\}|\emptyset) \)?
D-separation on Sets

• $d$-sep($\{Z\}, \{X, W, Q\} | \emptyset$)? Yes!

Blue path is a closed convergent path since we do not condition on $Y$ or it’s descendent $P$. 
D-separation on Sets

• $d$-sep($\{Z\}, \{X, W, Q\}|\{P\})$?
D-separation on Sets

• $d$-sep($\{Z\}, \{X, W, Q\} \mid \{P\}$)? No.

Blue path is an open convergent path since we condition on $P$, a descendent of $Y$. 
D-separation on Sets

• $d$-sep($\{Z, Y, P\}$, $\{W, Q\} | \{X\}$)?
D-seperation on Sets

- $d$-sep($\{Z, Y, P\}$, $\{W, Q\} | \{X\}$)? YES

Blue path is a closed sequential path since we condition on $X$. 
D-separation on Sets

• $d$-sep($\{Z, Y, P\}, \{W, Q\}|\emptyset)$?
D-seperation on Sets

- $d$-sep($\{Z, Y, P\}, \{W, Q\}|\emptyset)$? No.

Blue path is a closed sequential path since we do not condition on $X$. 
D-separation: Multiple Paths

- \( \text{d-sep}({\{C\}, \{F\}} | {\{D\}}) \)?
D-separation: Multiple Paths

- $d$-sep($\{C\}, \{F\} | \{D\})$?
D-separation: Multiple Paths

- $d$-sep($\{C\}, \{F\} | \{D\})$?

Red path is blocked by D.
Blue path is blocked by E not in evidence.
D-separation: Multiple Paths

- \( d\text{-sep}({C}, \{F\}|\{D\}) \)? YES.

**Red path** is blocked by D.

**Blue path** is blocked by E not in evidence.
D-separation: Multiple Paths

• $d$-sep({C}, {F} | {D, E})?
D-separation: Multiple Paths

- \( d\text{-}sep({C}, {F}|{D, E}) \)?
D-seperation: Multiple Paths

- $d\text{-sep}(\{C\}, \{F\} \mid \{D, E\})$?

Red path is blocked by D.
Blue path is NOT blocked.
D-separation: Multiple Paths

- $d$-sep\($\{C\}, \{F\}|\{D, E\}\) \text{? No.}$

Red path is blocked by D.
Blue path is NOT blocked.
D-separation

- D-separation is about more general conditional independence in a Bayesian network.
- If two sets of nodes X and Y are d-separated in Bayesian networks by a third set Z (excluding X and Y), the corresponding variable sets X and Y are independent given the variables in Z.
**Temporal Models - Dynamic Bayesian Networks (DBNs)**

- Dynamic Bayesian Networks (DBNs) are directed graphical models of stochastic processes.
- It is assumed that the model structure does not change and also normally assume that the parameters do not change, i.e., the model is time-invariant.
HMM

• The simplest kind of DBN is a Hidden Markov Model (HMM), which has one discrete hidden node and one discrete or continuous observed node per slice.

• **Markov property:** if the conditional probability distribution of future states of the process depend only upon the present state; that is, the past is irrelevant because it doesn't matter how the current state was obtained.
Bayesian Network Inference Example

The most common task we wish to solve using Bayesian networks is probabilistic inference.

\[
\begin{array}{cc|cc}
C & \text{P}(C=F) & \text{P}(C=T) \\
\hline
F & 0.5 & 0.5 \\
T & 0.9 & 0.1 \\
\end{array}
\]

\[
\begin{array}{c|cc|cc}
\text{SprinklerOn} & \text{P}(S=F) & \text{P}(S=T) \\
\hline
F & 0.5 & 0.5 \\
T & 0.9 & 0.1 \\
\end{array}
\]

\[
\begin{array}{c|cc|cc}
\text{Rain} & \text{P}(R=F) & \text{P}(R=T) \\
\hline
F & 0.8 & 0.2 \\
T & 0.2 & 0.8 \\
\end{array}
\]

\[
\begin{array}{c|cc|cc}
\text{WetGrass} & \text{P}(W=F) & \text{P}(W=T) \\
\hline
F & 1.0 & 0.0 \\
T & 0.1 & 0.9 \\
\end{array}
\]

\[
\begin{array}{c|cc|cc}
S & R & \text{P}(W=F) & \text{P}(W=T) \\
\hline
\text{F} & \text{F} & 0.1 & 0.9 \\
\text{T} & \text{F} & 0.1 & 0.9 \\
\text{F} & \text{T} & 0.1 & 0.9 \\
\text{T} & \text{T} & 0.01 & 0.99 \\
\end{array}
\]
BN Inference Example

• Suppose we observe: the grass is wet. What is the probability it is raining?
More Details

\[
P(S = 1 | W = 1) = \frac{P(S = 1, W = 1)}{P(W)}
\]

\[
\sum_{c,r} P(C = c, S = 1, R = r, W = 1)
\]

\[
= \frac{\sum_{c,r,s} P(C = c, S = s, R = r, W = 1)}{\sum_{c,r,s} P(C = c, S = s, R = r, W = 1)}
\]

\[
\sum_{c,r} P(C = c) P(S = 1 | C = c) P(R = r | C = c) P(W = 1 | S = 1, R = r)
\]

\[
= \frac{\sum_{c,r,s} P(C = c) P(S = s | C = c) P(R = r | C = c) P(W = 1 | S = s, R = r)}{\sum_{c,r,s} P(C = c) P(S = s | C = c) P(R = r | C = c) P(W = 1 | S = s, R = r)}
\]

\[
= \frac{0.2781}{0.6471} = 0.43
\]