Announcements:
- Drop deadline is this Sunday Nov 5th.
- All lecture notes needed for T3 posted (L13,…,L17).
- T3 sample questions posted.
- A3 posted.

Bayesian Networks
- A BN over variables \{X_1, X_2,…, X_n\} consists of:
  - a DAG (directed acyclic graph) whose nodes are the variables
  - a set of CPTs (conditional probability tables) \(\text{Pr}(X_i \mid \text{Par}(X_i))\) for each \(X_i\)
- Key notions (see text for defn’s, all are intuitive):
  - parents of a node: \(\text{Par}(X_i)\)
  - children of node
  - descendants of a node
  - ancestors of a node
  - family: set of nodes consisting of \(X_i\) and its parents
    - CPTs are defined over families in the BN

Example (Binary valued Variables)
- A couple CPTS are "shown".
- Explicit joint requires \(2^{11} - 1 = 2047\) parmrts
- BN requires only 27 parmrts (the number of entries for each CPT is listed)

Semantics of Bayes Nets.
- A Bayes net specifies that the joint distribution over the variable in the net can be written as the following product decomposition.
- \[\text{Pr}(X_1, X_2,…, X_n) = \text{Pr}(X_n \mid \text{Par}(X_n)) \times \text{Pr}(X_{n-1} \mid \text{Par}(X_{n-1})) \times \cdots \times \text{Pr}(X_1 \mid \text{Par}(X_1))\]
- This equation hold for any set of values \(d_1, d_2,…, d_n\) for the variables \(X_1, X_2,…, X_n\).
Semantics of Bayes Nets.

E.g., say we have $X_1, X_2, X_3$ each with domain $\text{Dom}[X_i] = \{a, b, c\}$ and we have

$$\text{Pr}(X_1, X_2, X_3) = \text{P}(X_3 | X_2) \text{P}(X_2) \text{P}(X_1)$$

Then

$$\text{Pr}(X_1=a, X_2=a, X_3=a) = \text{P}(X_3=a | X_2=a) \text{P}(X_2=a) \text{P}(X_1=a)$$
$$\text{Pr}(X_1=a, X_2=a, X_3=b) = \text{P}(X_3=b | X_2=a) \text{P}(X_2=a) \text{P}(X_1=a)$$
$$\text{Pr}(X_1=a, X_2=a, X_3=c) = \text{P}(X_3=c | X_2=a) \text{P}(X_2=a) \text{P}(X_1=a)$$

Example (Binary valued Variables)

$$\text{Pr}(a, b, c, d, e, f, g, h, i, j, k) = \text{Pr}(a) \times \text{Pr}(b) \times \text{Pr}(c | a) \times \text{Pr}(d | a, b) \times \text{Pr}(e | c) \times \text{Pr}(f | d) \times \text{Pr}(g) \times \text{Pr}(h | e, f) \times \text{Pr}(i | f, g) \times \text{Pr}(j | h, i) \times \text{Pr}(k | i)$$

Semantics of Bayes Nets.

Note that this means we can compute the probability of any setting of the variables using only the information contained in the CPTs of the network.

Constructing a Bayes Net

It is always possible to construct a Bayes net to represent any distribution over the variables $X_1, X_2, \ldots, X_n$, using any ordering of the variables.

- Take any ordering of the variables (say, the order given). From the chain rule we obtain.

  $$\text{Pr}(X_1, \ldots, X_n) = \text{Pr}(X_n | X_1, \ldots, X_{n-1}) \text{Pr}(X_{n-1} | X_1, \ldots, X_{n-2}) \ldots \text{Pr}(X_1)$$

- Now for each $X_i$ go through its conditioning set $X_1, \ldots, X_{i-1}$, and iteratively remove all variables $X_j$ such that $X_i$ is conditionally independent of $X_j$ given the remaining variables. Do this until no more variables can be removed.

  The final product will specify a Bayes net.
Constructing a Bayes Net

- The end result will be a product decomposition/Bayes net
  \[ \Pr(X_n | \text{Par}(X_n)) \Pr(X_{n-1} | \text{Par}(X_{n-1})) \ldots \Pr(X_1) \]

- Now we specify the numeric values associated with each term \( \Pr(X_i | \text{Par}(X_i)) \) in a CPT.

- Typically we represent the CPT as a table mapping each setting of \( \{X_i, \text{Par}(X_i)\} \) to the probability of \( X_i \) taking that particular value given that the variables in \( \text{Par}(X_i) \) have their specified values.

- If each variable has \( d \) different values.
  - We will need a table of size \( d^{|\{X_i, \text{Par}(X_i)\}|} \).
  - That is, exponential in the size of the parent set.

- Note that the original chain rule
  \[ \Pr(X_1, \ldots, X_n) = \Pr(X_n | X_1, \ldots, X_{n-1}) \Pr(X_{n-1} | X_1, \ldots, X_{n-2}) \ldots \Pr(X_1) \]
  requires as much space to represent as specifying the probability of each individual event.

Causal Intuitions

- The BN can be constructed using an arbitrary ordering of the variables.

- However, some orderings will yield BN’s with very large parent sets. This requires exponential space, and (as we will see later) exponential time to perform inference.

- Empirically, and conceptually, a good way to construct a BN is to use an ordering based on causality. This often yields a more natural and compact BN.

Causal Intuitions

- Malaria, the flu and a cold all “cause” aches. So use the ordering that causes come before effects
  Malaria, Flu, Cold, Aches

  \[ \Pr(M, F, C, A) = \Pr(A | M, F, C) \Pr(C | M, F) \Pr(F | M) \Pr(M) \]

- Each of these disease affects the probability of aches, so the first conditional probability does not change.

- It is reasonable to assume that these diseases are independent of each other: having or not having one does not change the probability of having the others. So \( \Pr(C | M, F) = \Pr(C) \) \( \Pr(F | M) = \Pr(F) \)
Suppose we build the BN for distribution \( P \) using the opposite ordering:

- i.e., we use ordering Aches, Cold, Flu, Malaria

\[
Pr(A,C,F,M) = Pr(M|A,C,F) \cdot Pr(F|A,C) \cdot Pr(C|A) \cdot Pr(A)
\]

We can't reduce \( Pr(M|A,C,F) \). Probability of Malaria is clearly affected by knowing aches. What about knowing aches and Cold, or aches and Cold and Flu?

- Probability of Malaria is affected by both of these additional pieces of knowledge

Knowing Cold and of Flu lowers the probability of Aches indicating Malaria since they "explain away" Aches!

Similarly, we can't reduce \( Pr(F|A,C) \).

\( Pr(C|A) \neq Pr(C) \)

Obtain a much more complex Bayes net. In fact, we obtain no savings over explicitly representing the full joint distribution (i.e., representing the probability of every atomic event).

I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

- Variables: Burglary, Earthquake, Alarm, JohnCalls, MaryCalls

- Network topology reflects "causal" knowledge:
  - A burglar can set the alarm off
  - An earthquake can set the alarm off
  - The alarm can cause Mary to call
  - The alarm can cause John to call
Burglary Example

- A burglary can set the alarm off
- An earthquake can set the alarm off
- The alarm can cause Mary to call
- The alarm can cause John to call

\[
\begin{array}{c|c}
\text{B} & \text{E} & P(A|B,E) \\
T & T & .95 \\
T & F & .94 \\
F & T & .29 \\
F & F & .001 \\
\end{array}
\]

- # of Params: \(1 + 1 + 4 + 2 + 2 = 10\) (vs. \(2^5 - 1 = 31\))

Example of Constructing Bayes Network

- Suppose we choose the ordering \(M, J, A, B, E\)

\[
P(J | M) = P(J)\? \\
P(B | A, J, M) = P(B | A)\? \\
P(B | A, J, M) = P(B)\?
\]
Example continue...

- Suppose we choose the ordering M, J, A, B, E

\[ P(J \mid M) = P(J)? \text{ No} \]
\[ P(A \mid J, M) = P(A \mid J)? P(A \mid J, M) = P(A)? \text{ No} \]
\[ P(B \mid A, J, M) = P(B \mid A)? \text{ Yes} \]
\[ P(B \mid A, J, M) = P(B)? \text{ No} \]
\[ P(E \mid B, A, J, M) = P(E \mid A)? \text{ No} \]
\[ P(E \mid B, A, J, M) = P(E \mid A, B)? \text{ Yes} \]

Example continue...

- Suppose we choose the ordering M, J, A, B, E

\[ P(J \mid M) = P(J)? \text{ No} \]
\[ P(A \mid J, M) = P(A \mid J)? P(A \mid J, M) = P(A)? \text{ No} \]
\[ P(B \mid A, J, M) = P(B \mid A)? \text{ Yes} \]
\[ P(B \mid A, J, M) = P(B)? \text{ No} \]
\[ P(E \mid B, A, J, M) = P(E \mid A)? \text{ No} \]
\[ P(E \mid B, A, J, M) = P(E \mid A, B)? \text{ Yes} \]

Example continue...

- Deciding conditional independence is **hard** in non-causal directions!

- (Causal models and conditional independence seem hardwired for humans!)

- Network is **less compact**: \(1 + 2 + 4 + 2 + 4 = 13\) numbers needed!