Inference

• Suppose you are given a Bayesian network with the graph structure and the parameters all figured out
• Now you would like to use it to do inference
• You need inference to make predictions or classifications with a Bayes net

Another Example

• You are very sick and you visit your doctor.
• The doctor is able to get the following information from you:
  – HasFever = true
  – HasCough = true
  – HasBreathingProblems = true
  – AteBaconRecently = true
• What’s the probability you have SwineFlu given the above?

Another Example

• Need to compute
  $P(SwineFlu = true \mid HasFever = true, HasCough = true, HasBreathingProblems = true, AteBaconRecently = true)$
• Suppose you pass out before you say a word to the doctor. The doctor is only able to determine you have a fever. What is $P(SwineFlu = true \mid HasFever = true)$?

Query Example

$P(SwineFlu = true \mid HasFever = true)$

Query Variable

Evidence Variable

Unobserved variables: HasCough, HasBreathingProblems, AteBaconRecently

Queries Formalized

We will use the following notation:

• $X$ = query variable
• $E = \{E_1, \ldots, E_m\}$ is the set of evidence variables
• $e$ = observed event
• $Y = \{Y_1, \ldots, Y_l\}$ are the non-evidence (or hidden) variables
• The complete set of variables $X = \{X\} \cup E \cup Y$

Need to calculate the query $P(X \mid e)$
Recall that:

\[ P(X | e) = \alpha P(X, e) = \alpha \sum_y P(X, e, y) \]

\[ P(x_1, \ldots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(X_i)) \]

This means you can answer queries by computing sums of products of conditional probabilities from the network.

**Example #1**

Whenever you see a conditional like \( P(B = \text{true} | C = \text{true}) \), use the Chain Rule:

\[ P(B = \text{true} | C = \text{true}) = P(B = \text{true}, C = \text{true}) / P(C = \text{true}) \]

To express the joint probability distribution as the entries in the CPTs, use:

\[ P(X_1, \ldots, X_n) = \prod_{i=1}^n P(X_i | \text{Parents}(X_i)) \]

**Example #1**

Whenever you need to get a subset of the variables e.g. \( P(B, C) \) from the full joint distribution \( P(A, B, C, D) \), use marginalization:

\[ P(X) = \sum_y P(X, Y = y) \]

**Example #1**

Take the probabilities that don’t depend on the terms in the summation and move them outside the summation.
Example #1

Take the probabilities that don’t depend on the terms in the summation and move them outside the summation.

Sums to 1

\[
\sum_{a} \sum_{d} P(A = a) P(B = b | A = a) P(C = true | A = a) P(D = d | C = true) = \sum_{a} \sum_{d} P(A = a) P(B = b | A = a) P(C = true | A = a) P(D = d | C = true) = \sum_{a} P(A = a) \sum_{d} P(B = b | A = a) P(C = true | A = a) P(D = d | C = true)
\]

Example #2

\[
P(B = true | J = true, M = true) = \frac{P(B = true, J = true, M = true)}{P(J = true, M = true)} = \frac{\sum_{a} \sum_{b} P(B = b, E = c, A = a, J = true, M = true)}{\sum_{a} \sum_{b} P(J = true, E = c, A = a, J = true, M = true)}
\]

Practice

Write out the equations for the following probabilities using probabilities you can obtain from the Bayesian network. You will have to leave it in symbolic form because the CPTs are not shown, but simplify your answer as much as possible.

1. \( P(A=true, B=true, C=true, D=true, E=true) \)

2. \( P(B=true | D=true) \)

3. \( P(A=true, D=true, E=true | B=true, C=true) \)
Complexity of Exact Inference

- The Burglary/Earthquake Bayesian network is an example of a polytree
- Singly connected networks (aka polytrees) have at most one undirected path between any two nodes in the network

Complexity of Exact Inference

- What about multiply connected networks?
- Exponential time and space complexity in the number of variables in the worst case
- Bad news: Inference in Bayesian networks is NP-hard
- Even worse news: inference is #P-hard (strictly harder than NP-complete problems)

The Good News

- Although exact inference is NP-hard, approximate inference is tractable
  - Lots of promising methods like sampling, MCMC, variational methods, etc.
- Approximate inference is a current research topic in Machine Learning

Practice

4. What is \( P(B=\text{false}, C=\text{false}) \)?

5. Can you come up with another Bayes net structure (using only the 3 nodes above) that represents the same joint probability distribution?
What You Should Know

• How to do exact inference in probabilistic queries of Bayes nets
• The complexity of inference for polytrees and multiply connected networks