Outline

• Logical inference and probabilistic inference

• Independence and conditional independence

• Bayes Nets
  – Semantics of Bayes Nets
  – How to construct a Bayes net
  – Conditional Independence in Bayes nets

• Variable elimination algorithm

• Naïve Bayes
Logical inference vs. probabilistic inference

- Problem: KB |= α?
- Model checking can determine entailment
  - Is M(KB) a subset of M(α)?
  - # of models: 2^n, n=3 here.
- Problem: P(X,Y)=? Or P(X|Y)=?
- Full joint probability distribution can be used to answer any query.
  - # of parameters: hmk > 2^n
  - How to answer the query?

<table>
<thead>
<tr>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>KB</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>?</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>?</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>?</td>
</tr>
</tbody>
</table>
Inference given full joint probability distribution

- Joint probability
  \[ P(x, y) = \sum_z P(x, y, z) \] (Marginalization)

- Conditional probability
  \[ P(x | y) = \frac{P(x, y)}{P(y)} = \frac{\sum_z P(x, y, z)}{\sum_{x,z} P(x, y, z)} \] (definition + marginalization)
  
  - Or \( P(x | y) = \alpha \sum_z P(x, y, z) \) (normalization)

  \[ \alpha = \frac{1}{\sum_{x,z} P(x, y, z)} \]

- Time and space: \( O(2^n) \)
Independence and conditional independence

• Independence of two events
  – Events $a$ and $b$ are independent if knowing $b$ tells us nothing about $a$
  – $P(a|b) = P(a)$ or $P(a|b) = P(a)P(b)$

• Independence of two random variables
  – Random variable $X$ and $Y$ are independent if for all $x,y$, $P(X=x, Y=y) = P(X=x)P(Y=y)$
  – Shorthand: $P(X,Y) = P(X)P(Y)$

• Conditional independence
  – $X$ and $Y$ are conditionally independent given $Z$ if $P(X,Y|Z) = P(X|Z)P(Y|Z)$
Bayesian Network/Bayes Net (1)

• Semantics
  – Nodes are random variables
  – Edges are directed. Edge $X \rightarrow Y$ indicates $x$ has a direct influence on $Y$
  – There is no cycles
  – Each node is associated with a conditional probability distribution: $P(x|\text{Parents}(x))$

• How to construct a Bayes Net?
  – Topology comes from human expert
  – Conditional probabilities: learned from data
Conditional independence in Bayes Nets

- A node is conditionally independent of non-descendants given its parents.
- A node conditionally independent of all other nodes given its Markov blanket.
  - A Markov blanked of a node is composed of its parents, its children, and its children’s other parents.
Bayesian Network/Bayes Net(3)

- Bayes nets represent the full joint probability

\[ P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\text{Parents}(X_i)) \]

- Exact inference (\( P(b|j,m) = ? \) example in the textbook)

\[
P(b, | j, m) = \alpha P(b, j, m) = \alpha \sum_{e,a} P(b, j, m, e, a)
\]

\[
= \alpha \sum_{e,a} P(b)P(e)P(a|e, b)P(j|a)P(m|a)
\]

\[
= \alpha P(b) \sum_{e} P(e) \sum_{a} P(a|e. b)P(j|a)P(m|a)
\]
Variable Elimination Algorithm (1)

- Variable elimination algorithm

  \[ P(b, |j, m) = \alpha P(b) \sum_e P(e) \sum_a P(a|e. b)P (j|a)P(m|a) \]

  - \( g_1(e, b) = \sum_a P(a|e. b)P (j|a)P(m|a) \)
  - \( g_2(b) = \sum_e P(e)g_1(e, b) \)
  - \( g_3(b) = P(b) g_2(b) \)

- Define and evaluate function for each summation from right to left.

- Evaluate once and store the values to be used later.

- Normalize.
Variable elimination algorithm (2)

- **Time and space:**
  - linear in terms of the size of Bayes net for singly connected networks.
  - Exponential for multiply connected networks.

- **Singly-connected networks vs Multiply-connected networks**
  - In singly-connected networks, also called polytrees, there is at most one undirected path between any two nodes.
  - In multiply-connected networks, there could be 2 or more undirected paths between 2 nodes.
Naïve Bayes

- A special case of Bayes net: one parent node and the rest are its children.
- Random variables: One cause and multiple effects.
- Assume that all effects are conditionally independent given the cause.
- Very tractable.
- Can be used for classification: Naïve Bayes classifier.
Review questions: true/false

1. Given the full joint probability distribution, we can answer most, but not all, inference queries.

2. A Bayes net completely and implicitly defines the full joint probability distribution of all random variables in a probabilistic model.

3. In a Bayes net, a node is conditionally independent of all other nodes given its parents.

4. In a Bayes net, a node is conditionally independent of all other nodes given its Markov blanket.

5. The Markov blanket of a node consists of its parents and its children.
6. Variable elimination algorithm can be used to do exact inference in any Bayes net.

7. For any Bayes net, variable elimination algorithm takes linear time and space in terms of the size of the Bayes net.

8. In singly connect networks, also called polytrees, there is at most one indirect path between any two nodes.

9. Bayes rule is very useful because it provides a way to calculate the conditional probability of a hidden variable given some evidence, which is usually hard to estimate directly.
Announcement & Reminder

- W2 is due today
  --- Turn in hard copies in class

- W1 will be returned after class today

- W3 has been released and is due on Tuesday Nov. 10th
  --- Turn in hard copies in class