Bayes Nets Quick Intro

- Topic of much current research
- Models dependence/independence in probability distributions
- Graph based - aka “graphical models” “belief nets”
- Two kinds - directed and undirected

Basic idea:
- Model a joint distribution as a DAG with attributes (features) at the nodes
- Sample distribution from sources

At each node is a conditional distribution for the attribute depending only on incoming arcs

<table>
<thead>
<tr>
<th>Party</th>
<th>Party yes</th>
<th>Party No</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(yes) = 0.1</td>
<td>P(yes)=0.9</td>
<td>P(no) = 0.1</td>
</tr>
<tr>
<td>P(no)=0.9</td>
<td>P(no) = 0.1</td>
<td></td>
</tr>
</tbody>
</table>

Diagram:
- Party
- Sleep
- Pass Exam
- Aptitude
- Study

P(yes) = 0.6
P(no) = 0.4
Causes and Bayes' Rule

Diagnostic inference: Knowing that the grass is wet, what is the probability that rain is the cause?

Causal vs Diagnostic Inference

Causal inference: If the sprinkler is on, what is the probability that the grass is wet?

\[
\]

Diagnostic inference: If the grass is wet, what is the probability that the sprinkler is on?

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

Explaining away: Knowing that it has rained decreases the probability that the sprinkler is on.

Bayesian Networks: Causes

Causal inference:

\[
\]

and use the fact that

\[
P(R, S | C) = P(R | C) P(S | C)
\]

Diagnostic: \(P(C | W) = ?\)

Bayesian Nets: Local structure

P(F | C) = ?
Bayesian Networks: Inference

\[
P(C,S,R,W,F) = P(C) P(S|C) P(R|C) P(W|R,S) P(F|R)
\]

\[
P(C,F) = \Sigma_S \Sigma_R \Sigma_W P(C,S,R,W,F)
\]

\[
P(F|C) = P(C,F) / P(C) \quad \text{Not efficient!}
\]

Belief propagation (Pearl, 1988)

Junction trees (Lauritzen and Spiegelhalter, 1988)

Conditional independence

- Know values of some nodes, which others are independent?
  - Nodes depend on ancestors through parents
  - Node conditionally indep of rest of graph if its parents, children, and children’s parents known
  - Notion of d-path - two nodes dependent if there is a d-path (or coincidence), independent otherwise

D-paths go through

- Linear non-evidence nodes (both ways)
- Diverging non-evid. Nodes
- Converging evidence nodes:
  - Conving non-evidence w. evidence desc.

In general, working with Bayes nets hard, (easy for trees, also approximate inference - loopy propagation, sampling)

- Concise representation of some distributions
- Very flexible - can fix or learn structure, can have unobserved nodes (latent variables)
- Can capture conditional dependencies
- Generalizes HMMs and other models
Bayes nets problems:

<table>
<thead>
<tr>
<th></th>
<th>Full Observability</th>
<th>Partial Observability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Graph Known</strong></td>
<td>Maximum likelihood</td>
<td>Expectation-Maximization</td>
</tr>
<tr>
<td><strong>Graph Unknown</strong></td>
<td>Local search over models</td>
<td>Hard</td>
</tr>
</tbody>
</table>

Bayes Nets Applications

- Add probabilities to expert systems
- Vista system (Horvitz) for space shuttle problem detection/remedies
- Used in Microsoft products: Office ‘95 Answer Wizard, Office ‘97 Office Assistant, etc.
- Computational Biology
- Special cases applied to speech recognition, turbo coding, and many other places.