Bayesian Belief Networks

• Learning about a situation not to perform a task.
• Combination of probabilistic modeling and DAGs
• Nodes on graph are propositional variables.
  – Lifting to first order is an active research area
• Links represent apriori known causal dependencies.
• Reasoning by merging semantic models and evidence.
• Efficient representation of joint distribution
Direct World Representations

• Can compute any subset of propositions given another subset.
• Perform different types of reasoning
  – Prediction
  – Abduction
  – Explaining away
• Global semantics
• Local semantics exploit conditional independence
Reasoning with a Bayesian Net

- Reasoning without evidence
- Reasoning with evidence
- Bayesian network reasoning NP-Hard
  - Instance of propositional logic satisfiability problem
- Use Monte Carlo techniques to simulate draws from the joint distribution
Learning Networks

• Four situations
  – Structure known, All variables observed
    • Simple counting exercise!
  – Structure known, some variables unobserved
    • EM
  – Structure unknown, All variables observed
    • ???
  – Structure unknown, some variables unobserved
    • Structural EM

• Currently focus on finding best model, but will later focus on finding multiple models.
  – How? Why?
Structure Known

- **Full Observability**
  - Count to work out every conditional probability table stored at a node. Maximum likelihood est.
  - Use Laplace correction to stop zero probabilities

- **Partial Observability**
  - Postulate that a variable contains hidden/missing values
  - E step: calculate expectation of hidden values
    - How?
  - M step: Maximize likelihood like above.
Structure Unknown

• How complex should the graph be?
• Full Observability
  – How many links to postulate?
  – What graph would be the maximum likelihood?
  – Penalize complex models. BIC/MDL
  – Number of DAGS super-exponential, use relative scores
  – Can use MCMC
• Partial Observability
  – How many nodes and how many links to postulate?
  – Use BIC/MDL
  – Local search within the M Step = Structural EM
Learning Using EM

EM algorithm can also be used. Repeatedly:

1. Calculate probabilities of unobserved variables, assuming $h$

2. Calculate new $w_{ijk}$ to maximize $E[\ln P(D|h)]$ where $D$ now includes both observed and (calculated probabilities of) unobserved variables

When structure unknown...

- Algorithms use greedy search to add/substruct edges and nodes
- Active research topic
Example

- TTF, FFT, TFF, FFT, FFF, FFF, FFF
- TF?, FT?, FF?, FT?, FF?, FF?, FF?
- TF?, FT?, FF?, FTF, FF?, FFF, FF?
### BBN Learning Example

#### Complete Data

<table>
<thead>
<tr>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X1</th>
<th>X2</th>
<th>W</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>W</td>
<td>0.666667</td>
<td>F</td>
<td>0.4</td>
</tr>
<tr>
<td>W</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>D</td>
<td>0.333333</td>
<td>T</td>
<td>0.6</td>
</tr>
<tr>
<td>D</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>D</td>
<td>0.333333</td>
<td>T</td>
<td>0.6</td>
</tr>
</tbody>
</table>

#### Incomplete Data

<table>
<thead>
<tr>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X1</th>
<th>X2</th>
<th>W</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>?</td>
<td>T</td>
<td>T</td>
<td>W</td>
<td>0.666667</td>
<td>F</td>
<td>0.5</td>
</tr>
<tr>
<td>W</td>
<td>?</td>
<td>F</td>
<td>T</td>
<td>D</td>
<td>0.333333</td>
<td>T</td>
<td>0.5</td>
</tr>
<tr>
<td>D</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>D</td>
<td>0.333333</td>
<td>T</td>
<td>0.6</td>
</tr>
<tr>
<td>W</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>0.666667</td>
<td>T</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Order of infering tables is now important!

Note we are calculating the maximum likelihood of X with Z marginalized out.

#### Hidden Nodes

<table>
<thead>
<tr>
<th>X1</th>
<th>X2</th>
<th>Xhidden</th>
<th>X3</th>
<th>X4</th>
<th>X1</th>
<th>X2</th>
<th>W</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>T</td>
<td>?</td>
<td>T</td>
<td>T</td>
<td>W</td>
<td>0.666667</td>
<td>F</td>
<td>0.6</td>
</tr>
<tr>
<td>W</td>
<td>T</td>
<td>?</td>
<td>F</td>
<td>T</td>
<td>D</td>
<td>0.333333</td>
<td>T</td>
<td>0.4</td>
</tr>
<tr>
<td>D</td>
<td>T</td>
<td>?</td>
<td>T</td>
<td>T</td>
<td>D</td>
<td>0.333333</td>
<td>T</td>
<td>0.4</td>
</tr>
<tr>
<td>W</td>
<td>F</td>
<td>?</td>
<td>T</td>
<td>T</td>
<td>D</td>
<td>0.333333</td>
<td>T</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Randomly assign Xhidden z'values

Why?

Form of "relationship" between X3 and X2 may be insufficient to be represented in a simple table,
Latent Variable Models

– Attempt to find unknown classes or entities to better explain commonly occurring patterns.
– Think of an example?
– Mixture Models Are A Simple and Common Example of LVM.
  • Latent variable is an implicit class that is not explicitly given.
Definition of a Mixture Model

• Assume $m$ independent Gaussian distributed attributes and $k$ generating mechanisms

• What are the parts of the model?

• $P(h|D) = ?$
Bayes Optimal Classifier

\[\text{arg max}_{v_j \in V} \sum_{h_i \in H} P(v_j|h_i)P(h_i|D)\]

Example:

\[P(h_1|D) = .4, \quad P(-|h_1) = 0, \quad P(+|h_1) = 1\]
\[P(h_2|D) = .3, \quad P(-|h_2) = 1, \quad P(+|h_2) = 0\]
\[P(h_3|D) = .3, \quad P(-|h_3) = 1, \quad P(+|h_3) = 0\]

therefore

\[\sum_{h_i \in H} P(+|h_i)P(h_i|D) = .4\]
\[\sum_{h_i \in H} P(-|h_i)P(h_i|D) = .6\]

and

\[\text{arg max}_{v_j \in V} \sum_{h_i \in H} P(v_j|h_i)P(h_i|D) = -\]
Gibbs Algorithm (Not Sampler)

Bayes optimal classifier provides best result, but can be expensive if many hypotheses. Gibbs algorithm:

1. Choose one hypothesis at random, according to \( P(h|D) \)
2. Use this to classify new instance

Surprising fact: Assume target concepts are drawn at random from \( H \) according to priors

\[
E[\text{error}_{\text{Gibbs}}] \leq 2E[\text{error}_{\text{BayesOptimal}}]
\]

Suppose correct, uniform prior distribution over \( H \), then

- Pick any hypothesis from VS, with uniform probability
- Its expected error no worse than twice Bayes optimal