Lecture 8: Graphical Models II
Machine Learning

Andrew Rosenberg

March 5, 2010
Graphical Models
- Naive Bayes classification
- Conditional Probability Tables (CPTs)
- Inference in Graphical Models and Belief Propagation
Graphical Models

- Graphical representation of the dependency relationships between random variables.
Topological Graph

Graphical models factorize probabilities

\[ p(x_0, \ldots, x_{n-1}) = \prod_{i=0}^{n-1} p(x_i | p_{ai}) = \prod_{i=0}^{n-1} p(x_i | \pi_i) \]

Nodes are generally topologically ordered so that parents, \( \pi \) come before children.
Recall the Naive Bayes Graphical Model

There can be many variables $x_i$. Plate notation gives a compact representation of models like this:
### Naive Bayes Example

#### Data Table

<table>
<thead>
<tr>
<th>Flu</th>
<th>Fever</th>
<th>Sinus</th>
<th>Ache</th>
<th>Swell</th>
<th>Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>L</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Y</td>
<td>H</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Y</td>
<td>M</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

#### Graphical Representation

![Graphical representation of Naive Bayes example](image-url)
Naive Bayes Example

<table>
<thead>
<tr>
<th>Flu</th>
<th>Fever</th>
<th>Sinus</th>
<th>Ache</th>
<th>Swell</th>
<th>Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>L</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Y</td>
<td>H</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Y</td>
<td>M</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

\[ p(\text{flu}) = \begin{array}{c|c}
Y & N \\
.75 & .25
\end{array} \]
Naive Bayes Example

\[
p(\text{flu}) = \begin{array}{c|c|c} 
\text{Y} & \text{N} \\
.75 & .25 
\end{array}
\]

\[
p(\text{fev} | \text{flu}) = \begin{array}{c|c|c|c} 
\text{L} & \text{M} & \text{H} \\
\text{Y} & .33 & .33 & .33 \\
\text{N} & 0 & 1 & 0 
\end{array}
\]
Naive Bayes Example

### Example

<table>
<thead>
<tr>
<th>Flu</th>
<th>Fever</th>
<th>Sinus</th>
<th>Ache</th>
<th>Swell</th>
<th>Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>L</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Y</td>
<td>H</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Y</td>
<td>M</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

\[
p(\text{flu}) = \begin{pmatrix} .75 \\ .25 \end{pmatrix}, \quad p(\text{sinus} | \text{flu}) = \begin{pmatrix} Y & N \\ .67 & .33 \end{pmatrix}
\]

\[
p(\text{flu}) = \begin{pmatrix} Y & N \\ .75 & .25 \end{pmatrix}, \quad p(\text{sinus} | \text{flu}) = \begin{pmatrix} Y & .67 & .33 \\ N & 1 & 0 \end{pmatrix}
\]

---

**Flu**

**Fever**

**Sinus**

**Ache**

**Swell**

**Head**
Naive Bayes Example

<table>
<thead>
<tr>
<th>Flu</th>
<th>Fever</th>
<th>Sinus</th>
<th>Ache</th>
<th>Swell</th>
<th>Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>L</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Y</td>
<td>H</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Y</td>
<td>M</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

$p(\text{flu}) = \begin{array}{cc} Y & N \\ .75 & .25 \end{array}$

$p(\text{ache}|\text{flu}) = \begin{array}{cc} Y & N \\ .33 & .67 \end{array}$
Naive Bayes Example

\[
p(\text{flu}) = \begin{bmatrix} Y & N \\ .75 & .25 \end{bmatrix}, \quad p(\text{swell}|\text{flu}) = \begin{bmatrix} Y & N \\ .67 & .33 \\ N & 1 \end{bmatrix}
\]

<table>
<thead>
<tr>
<th>Flu</th>
<th>Fever</th>
<th>Sinus</th>
<th>Ache</th>
<th>Swell</th>
<th>Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>L</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Y</td>
<td>H</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Y</td>
<td>M</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>
Naive Bayes Example

<table>
<thead>
<tr>
<th>Flu</th>
<th>Fever</th>
<th>Sinus</th>
<th>Ache</th>
<th>Swell</th>
<th>Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>L</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Y</td>
<td>H</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Y</td>
<td>M</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

\[
p(\text{flu}) = \begin{array}{cc}
  \text{Y} & \text{N} \\
  .75 & .25 
\end{array}
\]

\[
p(\text{head}|\text{flu}) = \begin{array}{cc}
  \text{Y} & \text{N} \\
  .67 & .33 \\
  \text{N} & 1 \\
  & 0 
\end{array}
\]
Naive Bayes Example

<table>
<thead>
<tr>
<th>Flu</th>
<th>Fever</th>
<th>Sinus</th>
<th>Ache</th>
<th>Swell</th>
<th>Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>L</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Y</td>
<td>H</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Y</td>
<td>M</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>?</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Find $p(\text{flu}|\text{fever, sinus, ache, swell, head})$.

$$p(\text{flu} = Y)p(\text{fever} = M|\text{flu} = Y)p(\text{sinus} = N|\text{flu} = Y)p(\text{ache} = N|\text{flu} = Y)p(\text{swell} = N|\text{flu} = Y)p(\text{head} = N|\text{flu} = Y)$$

$$p(\text{flu}) = \begin{bmatrix} Y & N \\ .75 & .25 \end{bmatrix}$$
### Naive Bayes Example

<table>
<thead>
<tr>
<th>Flu</th>
<th>Fever</th>
<th>Sinus</th>
<th>Ache</th>
<th>Swell</th>
<th>Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>L</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Y</td>
<td>H</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Y</td>
<td>M</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>?</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Find \( p(\text{flu}|\text{fever, sinus, ache, swell, head}) \).

\[
.75 \times p(\text{fev} = M|\text{flu} = Y)p(\text{sin} = N|\text{flu} = Y)p(\text{ach} = N|\text{flu} = Y)p(\text{swe} = N|\text{flu} = Y)p(\text{head} = N|\text{flu} = Y)
\]

\[
p(\text{fev}|\text{flu}) = \begin{array}{ccc}
L & M & H \\
Y & .33 & .33 & .33 \\
N & 0 & 1 & 0
\end{array}
\]
Naive Bayes Example

<table>
<thead>
<tr>
<th>Flu</th>
<th>Fever</th>
<th>Sinus</th>
<th>Ache</th>
<th>Swell</th>
<th>Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>L</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Y</td>
<td>H</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Y</td>
<td>M</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>?</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Find \( p(flu|fever, \text{sinus}, \text{ache}, \text{swell}, \text{head}) \).

\[
.75 \times .33 \times p(sin = N|flu = Y)p(ach = N|flu = Y)p(swe = N|flu = Y)p(head = N|flu = Y)
\]

\[
p(sinus|flu) = \begin{array}{cc}
Y & .67 & .33 \\
N & 1 & 0
\end{array}
\]
Naive Bayes Example

<table>
<thead>
<tr>
<th>Flu</th>
<th>Fever</th>
<th>Sinus</th>
<th>Ache</th>
<th>Swell</th>
<th>Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>L</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Y</td>
<td>H</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Y</td>
<td>M</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>?</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Find $p(\text{flu} | \text{fever}, \text{sinus}, \text{ache}, \text{swell}, \text{head})$.

$.75 \times .33 \times .33 \times p(\text{ache} = \text{N} | \text{flu} = \text{Y})p(\text{swe} = \text{N} | \text{flu} = \text{Y})p(\text{head} = \text{N} | \text{flu} = \text{Y})$

$p(\text{ache} | \text{flu}) = \begin{array}{c|cc}
\text{Y} & \text{N} \\
\hline
.33 & .67 \\
\hline
\text{N} & 1 & 0
\end{array}$
Naive Bayes Example

<table>
<thead>
<tr>
<th>Flu</th>
<th>Fever</th>
<th>Sinus</th>
<th>Ache</th>
<th>Swell</th>
<th>Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>L</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Y</td>
<td>H</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Y</td>
<td>M</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>?</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Find \( p(\text{flu}| \text{fever}, \text{sinus}, \text{ache}, \text{swell}, \text{head}) \).

\[
.75 \times .33 \times .33 \times .67 \times p(\text{swe} = N|\text{flu} = Y)p(\text{head} = N|\text{flu} = Y)
\]

\[
p(\text{swell}|\text{flu}) = \begin{array}{cc}
Y & N \\
.67 & .33 \\
N & 1 & 0
\end{array}
\]
Naive Bayes Example

<table>
<thead>
<tr>
<th>Flu</th>
<th>Fever</th>
<th>Sinus</th>
<th>Ache</th>
<th>Swell</th>
<th>Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>L</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Y</td>
<td>H</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Y</td>
<td>M</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>?</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Find $p(\text{flu} | \text{fever}, \text{sinus}, \text{ache}, \text{swell}, \text{head})$.

$.75 \times .33 \times .33 \times .67 \times .33 \times p(\text{head} = N | \text{flu} = Y)$

<table>
<thead>
<tr>
<th></th>
<th>Y</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(\text{head}</td>
<td>\text{flu})$</td>
<td>.67</td>
</tr>
<tr>
<td>N</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Naive Bayes Example

Find $p(\text{flu} | \text{fever}, \text{sinus}, \text{ache}, \text{swell}, \text{head})$.

$.75 \times .33 \times .33 \times .67 \times .33 \times .33 = 0.0060$
Completely Observed graphical models

Suppose we have observations for every node.

<table>
<thead>
<tr>
<th></th>
<th>Flu</th>
<th>Fever</th>
<th>Sinus</th>
<th>Ache</th>
<th>Swell</th>
<th>Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>L</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Y</td>
<td>H</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Y</td>
<td>M</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

In the simplest – least general – graph, assume each independent. Train 6 separate models.

\[ (Fl) \quad (Fe) \quad (Si) \quad (Ac) \quad (Sw) \quad (He) \]

2nd simplest graph – most general – assume no independence. Build a 6-dimensional table. (Divide by total count.)
Consider this Graphical Model

- Each node has a conditional probability table $\theta_i$.
- Given the table, we have a pdf

$$p(x|\theta) = \prod_{i=0}^{M-1} p(x_i|\pi_i, \theta_i)$$

- We have $m$ variables in $x$, and $N$ data points, $X$.
- Maximum (log) Likelihood

$$\theta^* = \arg\max_{\theta} \ln p(X|\theta)$$

$$= \arg\max_{\theta} \sum_{n=0}^{N-1} \ln p(X_n|\theta)$$

$$= \arg\max_{\theta} \sum_{n=0}^{N-1} \sum_{i=0}^{M-1} \ln p(x_{in}|\theta_i)$$
First, Kronecker’s delta function.

\[ \delta(x_n, x_m) = \begin{cases} 
1 & \text{if } x_n = x_m \\
0 & \text{otherwise}
\end{cases} \]

Counts: the number of times something appears in the data

\[
m(x_i) = \sum_{n=0}^{N-1} \delta(x_i, x_{in})
\]

\[
m(X) = \sum_{n=0}^{N-1} \delta(X, X_n)
\]

\[
N = \sum_{x_1} m(x_1) = \sum_{x_1} \left( \sum_{x_2} \delta(x_1, x_2) \right) = \sum_{x_1} \left( \sum_{x_2} \left( \sum_{x_3} \delta(x_1, x_2, x_3) \right) \right) \ldots
\]
Maximum likelihood CPTs

\[ I(\theta) = \sum_{n=0}^{N-1} \ln p(X_n|\theta) \]

\[ = \sum_{n=0}^{N-1} \ln \prod_{x} p(x|\theta)^{\delta(x_n,x)} \]

\[ = \sum_{n=0}^{N-1} \sum_{x} \delta(x_n,x) \ln p(x|\theta) \]

\[ = \sum_{x_n} m(X) \ln \prod_{i=0}^{M-1} p(x_i|\pi, \theta_i) \]

\[ = \sum_{x_n} \sum_{i=0}^{M-1} m(X) \ln p(x_i|\pi, \theta_i) \]

Define a function:

\[ \theta(x_i, \pi_i) = p(x_i|\pi, \theta) \]

Constraint:

\[ \sum_{x_i} \theta(x_i, \pi_i) = 1 \]
Lagrange Multipliers

To maximize \( f(x, y) \) subject to \( g(x, y) = c \).
Maximize \( f(x, y) - \lambda(g(x, y) - c) \)

\[
I(\theta) = \sum_{i=0}^{M-1} \sum_{x_i} \sum_{\pi_i} m(x_i, \pi_i) \ln \theta(x_i, \pi_i) - \sum_{i=0}^{M-1} \sum_{\pi_i} \lambda_{\pi_i} \left( \sum_{x_i} \theta(x_i, \pi_i) - 1 \right)
\]

\[
\frac{\partial I(\theta)}{\partial \theta(x_i, \pi_i)} = \frac{m(x_i, \pi_i)}{\theta(x_i, \pi_i)} - \lambda_{\pi_i} = 0
\]

\[
\theta(x_i, \pi_i) = \frac{m(x_i, \pi_i)}{\lambda_{\pi_i}}
\]

\[
\sum_{x_i} \frac{m(x_i, \pi_i)}{\lambda_{\pi_i}} = 1 - \text{the constraint}
\]

\[
\lambda_{\pi_i} = \sum_{x_i} m(x_i, \pi_i) = m(\pi_i)
\]

\[
\theta(x_i, \pi_i) = \frac{m(x_i, \pi_i)}{m(\pi_i)} - \text{counts!}
\]
For the Bayesians, MAP leads to:

$$\theta(x_i, \pi_i) = \frac{m(x_i, \pi_i) + \epsilon}{m(\pi_i) + \epsilon|x_i|}$$
Example of maximum likelihood.

<table>
<thead>
<tr>
<th>Flu ($x_0$)</th>
<th>Fever ($x_1$)</th>
<th>Sinus ($x_2$)</th>
<th>Ache ($x_3$)</th>
<th>Swell ($x_4$)</th>
<th>Head ($x_5$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>L</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Y</td>
<td>H</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Y</td>
<td>M</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

$$\theta(x_i, \pi_i) = \frac{m(x_i, \pi_i)}{m(\pi_i)}$$
Conditional Dependence Test.

- We also want to be able to check conditional independencies in a graphical model.
- i.e. “Is achiness ($x_3$) independent of flu ($x_0$) given fever ($x_1$)?”
- i.e. “Is achiness ($x_3$) independent of sinus infection ($x_2$) given fever ($x_1$)?”

\[
p(x) = p(x_0)p(x_1|x_0)p(x_2|x_0)p(x_3|x_1)p(x_4|x_2)p(x_5|x_1, x_4)
\]

\[
p(x_3|x_0, x_1, x_2) = \frac{p(x_0, x_1, x_2, x_3)}{p(x_0, x_1, x_2)} = \frac{p(x_0)p(x_1|x_0)p(x_2|x_0)p(x_3|x_1)}{p(x_0)p(x_1|x_0)p(x_2|x_0)} = p(x_3|x_1)
\]

$x_3 \perp \perp x_0, x_2|x_1$

No problem, right?

What about $x_0 \perp \perp x_5|x_1, x_2$?
D-separation and Bayes Ball

Intuition: nodes are **separated**, or **blocked** by sets of nodes

- Example: nodes $x_1$ and $x_2$, “block” the path from $x_0$ to $x_5$, then $x_0 \perp \perp x_5 | x_2, x_3$
D-separation and Bayes Ball

Intuition: nodes are **separated**, or **blocked** by sets of nodes

- Example: nodes $x_1$ and $x_2$, “block” the path from $x_0$ to $x_5$, then $x_0 \perp \perp x_5 | x_2, x_3$
- While this is true in **undirected** graphs, it is not in directed graphs.
- We need more than simple **Separation**
- We need directed separation – **D-Separation**
- the D-separation is computed using the **Bayes Ball** algorithm.
- Allows us to prove general statements $x_a \perp \perp x_b | x_c$. 
Bayes Ball Algorithm

\[ x_a \perp x_b| x_c \]

- Shade nodes \( x_c \)
- Place a “ball” at each node in \( x_a \)
- Bounce balls around the graph according to some rules
- If no balls read \( x_b \), then \( x_a \perp x_b| x_c \), else false.
- Balls can travel along/against edges
- Pick any path
- Test to see if the ball goes through or bounces back.
Ten Rules of Bayes Ball
$x_0 \perp x_4 | x_2$?
\[ x_0 \perp x_5 | x_1, x_2? \]
What if we allow undirected graphs?
What do they correspond to?
It’s not cause/effect, or trigger/response, rather, general dependence.
Example: Image pixels, where each pixel is a bernouli.
Can have a probability over all pixels $p(x_{11}, x_{1M}, x_{M1}, x_{MM})$
Bright pixels have bright neighbors.
No parents, just probabilities.
Grid models are called **Markov Random Fields**.
Undirected separation is easy.

To check $x_a \perp x_b|x_c$, check Graph reachability of $x_a$ and $x_b$ without going through nodes in $x_c$. 
Next
- Representing probabilities in Undirected Graphs.