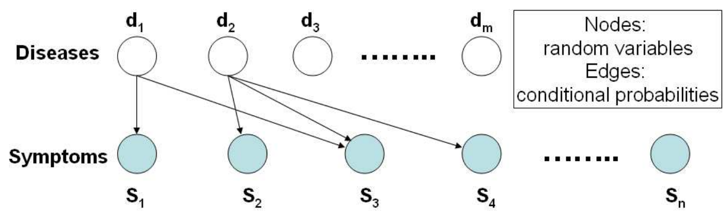
CS281A/Stat241A: Statistical Learning Theory

Directed Graphical Models (9/2/04)

Lecturer: Michael I. Jordan *Scribes: June Wang*

1 Another example of graphical models

Fault networks: $p(d_i|s)$ is very hard to calculate whereas $p(s|d_i)$ is trivial to calculate in this graphical model.



2 Directed graphical models

Notation: A graph is defined as $G = (V, E)$, where V represents vertices (nodes) and E represents directed edges (see figure 1).

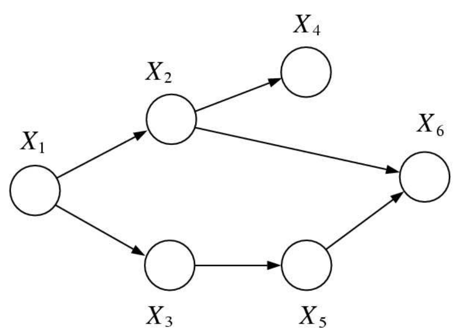


Figure 1: An example of a directed graphical model.

V indexes the set of random variables associated with nodes in the graph.
X_A represents all random variables indexed by A, e.g., $A = \{1, 2, 3, 4\}$.
X_V represents all random variables associated with G.

Definition: π_i is the set of parents of node i, e.g., in figure 1, $\pi_6 = \{2, 5\}$, $X_{\pi_6} = \{X_2, X_5\}$.

Definition: A *factor*, $f_i(x_i, x_{\pi_i})$, is a non-negative function where $\sum_{x_i} f_i(x_i, x_{\pi_i}) = 1$.
In the above example, one possible factor is $f_2(x_2, x_1)$.

Definition: There is a family, D_1, of probability distributions associated with G where

$$D_1 = \{P(x_V) = \prod_{i=1}^{n} f_i(x_i, x_{\pi_i})\} \tag{1}$$

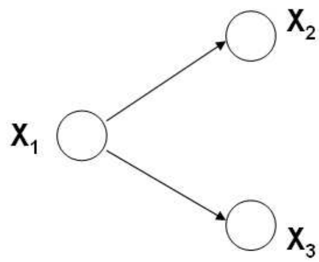


Figure 2: Small graph to illustrate factors.

An example of factors, in figure 2, the following equation holds for this graph: $p(x_1, x_2, x_3) = f_1(x_1)f_2(x_2, x_1)f_3(x_3, x_1)$

We can verify that this is indeed a joint probability as follows:

$$\begin{aligned}
\sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_1, x_2, x_3) &= \sum_{x_1}\sum_{x_2}\sum_{x_3} f_1(x_1)f_2(x_2, x_1)f_3(x_3, x_1) \\
&= \sum_{x_1}\sum_{x_2} f_1(x_1)f_2(x_2, x_1)\sum_{x_3} f_3(x_3, x_1) \\
&= \sum_{x_1}\sum_{x_2} f_1(x_1)f_2(x_2, x_1) \\
&= \sum_{x_1} f_1(x_1)\sum_{x_2} f_2(x_2, x_1) \\
&= \sum_{x_1} f_1(x_1) = 1
\end{aligned}$$

Idea: Marginalize out the leaves of the tree first, because summing over factors for leaves yields 1.

Theorem: $f_i(x_i, x_{\pi_i}) = p(x_i|x_{\pi_i})$. In other words, factors represent conditional probabilities statements (see figure 3).

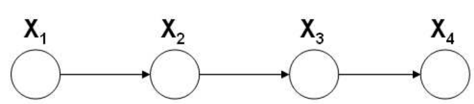


Figure 3: Small graph illustrating a conditional independence statement.

In this Markov chain, the joint probability factorizes as follows:

$$p(x_1, x_2, x_3, x_4) = f_1(x_1) f_2(x_2, x_1) f_3(x_3, x_2) f_4(x_4, x_3) \tag{2}$$

We can compute the marginal probability of $p(x_1, x_2, x_3)$ as follows:

$$\sum_{x_4} p(x_1, x_2, x_3, x_4) = p(x_1, x_2, x_3) = f_1(x_1) f_2(x_2, x_1) f_3(x_3, x_2) \tag{3}$$

Furthermore, we can compute the marginal probability of $p(x_3|x_2)$, and show that it is equivalent to the representation in terms of factors.

$$\begin{aligned} p(x_3|x_2) &= \frac{p(x_2, x_3)}{p(x_2)} = \frac{\sum_{x_1} f_1(x_1) f_2(x_2, x_1) f_3(x_3, x_2)}{\sum_{x_1} \sum_{x_3} f_1(x_1) f_2(x_2, x_1) f_3(x_3, x_2)} \\ &= \frac{f_3(x_3, x_2) \sum_{x_1} f_1(x_1) f_2(x_2, x_1)}{\sum_{x_1} f_1(x_1) f_2(x_2, x_1)} = f_3(x_3, x_2) \end{aligned}$$

Therefore $p(x_V) = \prod_{i=1}^{n} p(x_i | x_{\pi_i})$.

The chain rule of probability (makes no independence assumptions between the random variables):

$$p(x_1, x_2, x_3, x_4) = p(x_1) p(x_2|x_1) p(x_3|x_2, x_1) p(x_4|x_3, x_2, x_1) \tag{4}$$

Note that, as in the example above, factorization for a specific graph conditions on fewer variables than the chain rule requires:

$$p(x_1, x_2, x_3, x_4) = p(x_1) p(x_2|x_1) p(x_3|x_2) p(x_4|x_3).$$

Marginal Independence:

$$X_A \perp\!\!\!\perp X_B \equiv p(x_A, x_B) = p(x_A) p(x_B)$$

Conditional Independence:

$$\begin{aligned} X_A \perp\!\!\!\perp X_B | X_C &\equiv p(x_A, x_B | x_C) = p(x_A|x_C) p(x_B|x_C) \\ &\equiv p(x_A | x_B, x_B) = p(x_A|x_C) \end{aligned}$$

Associate the set of CI statements $C = \{X_i \perp\!\!\!\perp X_{\nu_i} | X_{\pi_i}\}$ with a graph G, where X_{ν_i} represents nodes that appear before X_i in some topological ordering of the node indices (see figure 4). $X_4 \perp\!\!\!\perp \{X_1, X_2, X_3\} | \{X_2, X_3\}$

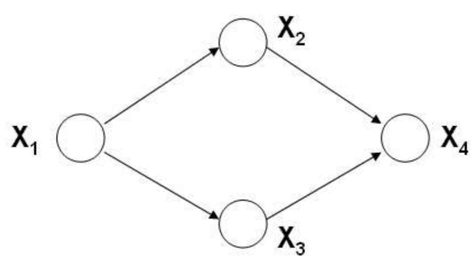


Figure 4: Another small graph illustrating a conditional independence statement.

Another example of conditional independence statements associated with a graphical model is for figure 3. Two conditional independence statements associated with this graph are: $X_3 \perp\!\!\!\perp \{X_1, X_2\}|X_2$, and $X_4 \perp\!\!\!\perp \{X_1, X_2, X_3\}|X_3$.

Definition: There is a family, D_2, of probability distributions associated with G, that includes all $p(x_\nu)$ that satisfy every CI statement associated with graph G.

Theorem: $D_1 = D_2$. We will prove this later in the course.

Given a set of conditional independence statements C, find all other CI's associated with graph G.

3 Three Canonical Graphs

Three three-node graphs will illustrate conditional independences in graphical models.

Graph 1 is a Markov chain, shown in figure 5. The important conditional independence statement associated with this graph is: $\{X \perp\!\!\!\perp Z|Y\}$
Proof:

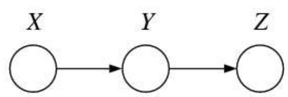


Figure 5: Small graph: *Markov chain*

$$p(z|x,y) = \frac{p(x,y,z)}{p(x,y)} = \frac{p(y)p(x|y)(p(z|y)}{p(x)p(y|x)} = p(z|y)$$

Graph 2 is the hidden cause graph; see figure 6. One conditional independence statement associated with this graph is: $\{X \perp\!\!\!\perp Z|Y\}$ (note: X is not independent of Z)
Proof:

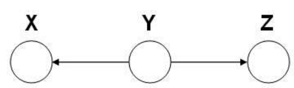


Figure 6: Small graph: *hidden cause*

$$p(z|x,y) = \frac{p(x,y,z)}{p(x,y)} = \frac{p(y)p(x|y)p(z|y)}{p(y)p(x|y)} = p(z|y).$$

Graph 3 is the explaining away graph; see figure 7. X is independent of Z (i.e., $\{X \perp\!\!\!\perp Z\}$), however X is not independent of Z conditioned on Y (i.e., $\neg\{X \perp\!\!\!\perp Z|Y\}$).

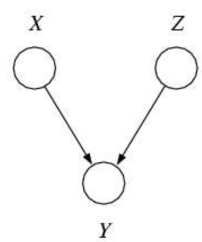


Figure 7: Small graph: *explaining away*