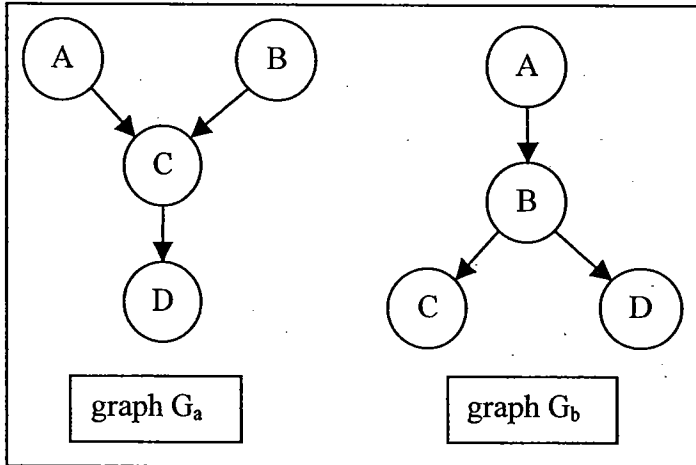


Solution

sscher@ucsc.edu
CS 290C, Winter 2006
Homework #2
Due Tu Feb 7

1. 3.15

Consider the two networks in Figure 3.24. For each of these two networks, G_a and G_b , determine whether there can be any other Bayesian network which is I-equivalent to it.



(a)

There is no graph G_{a_2} that is I-equivalent to G_a .

G_1 and G_2 are "I-Equivalent" if they imply the same set of conditional independencies. :

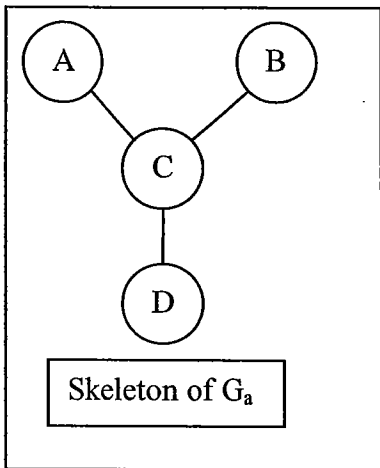
This can be broken down into two requirements:

- (1) the same skeleton (undirected graph)
- (2) the same set of immoralities

The following of conditional independencies hold for graph G_a :

- (i) $A \perp B$
- (ii) $A \perp D | C$
- (iii) $C \perp D | C$

The unique undirected graph with the same skeleton as G_a is shown below.

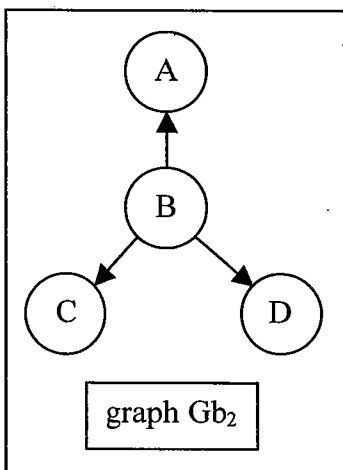


Graph G_a has only one immorality, $A \rightarrow C \leftarrow B$. If the undirected graph has this immorality, it will be G_a , as changing the direction of any edge AC or AB would remove the immorality, and changing the direction of edge CD would add a new immorality.

No graph I-equivalent to G_a exists.

(b)

G_b has no immoralities. The following graph, G_{b2} has the same skeleton as G_b & has no immoralities. It is found by reversing the direction of the edge AB. G_{b2} is I-equivalent to G_b .



2. 4.15

Consider a logistic CPD where X is a parent of Y.

20120

(1)

Assume that Y is m-valued, and consider a multinomial logit function defined as:

$$P(y^j | X) = \frac{\exp(w_{0,j} + w_{1,j}X)}{\sum_{j'=1}^m \exp(w_{0,j'} + w_{1,j'}X)}$$

This definition uses 2m parameters. Provide an alternative parameterization for this CPD that uses only (2m-2) parameters, and show how to convert from one to the other.

(2)

Now, assume that X is m -valued, and consider the parameterization of Equation 4.4:

$$P(y^j | X) = \text{sigmoid}\left(w_0 + \sum_{j=1}^m [w_{1,j} * 1\{X = x^j\}]\right)$$

where $1\{X=x^j\}$ is equal to 1 if $X=x^j$ and equal to 0 otherwise.

Once again, this definition uses $m+1$ parameters. Provide an alternative parameterization for this CDP that uses only m parameters, and show how to convert from one parameterization to the other.

(1)

We wish to subtract the final weights, $w_{0,m}$ and $w_{1,m}$, from all the other weights including the bias. The last weights will then be known to be zero, so we will have two fewer weights. To show that this does not change the distribution, multiply the original distribution by 1, or rather, multiply by $\exp(-[w_{0,m} + w_{1,m}X])\exp(-[w_{0,m} + w_{1,m}X])$.

Then distribute the denominator over each term in the sum.

$$\begin{aligned} P(y^j | X) &= \frac{\exp(w_{0,j} + w_{1,j}X)}{\sum_{j'=1}^m \exp(w_{0,j'} + w_{1,j'}X)} * 1 \\ &= \frac{\exp(w_{0,j} + w_{1,j}X)}{\sum_{j'=1}^m \exp(w_{0,j'} + w_{1,j'}X)} * \frac{\exp(-[w_{0,m} + w_{1,m}X])}{\exp(-[w_{0,m} + w_{1,m}X])} \\ &= \frac{\exp([w_{0,j} + w_{1,j}X]) * \exp(-[w_{0,m} + w_{1,m}X])}{\sum_{j'=1}^m [\exp([w_{0,j'} + w_{1,j'}X]) * \exp(-[w_{0,m} + w_{1,m}X])]} \end{aligned}$$

The two factors that are multiplied together can be combined by summing their exponents. The four terms in the exponent can be rearrange into two: a constant and linear term.

$$\begin{aligned} P(y^j | X) &= \frac{\exp([w_{0,j} + w_{1,j}X] - [w_{0,m} + w_{1,m}X])}{\sum_{j'=1}^m \exp([w_{0,j'} + w_{1,j'}X] - [w_{0,m} + w_{1,m}X])} \\ &= \frac{\exp(w_{0,j} - w_{0,m} + [w_{1,j} - w_{1,m}]X)}{\sum_{j'=1}^m \exp(w_{0,j'} - w_{0,m} + [w_{1,j'} - w_{1,m}]X)} \end{aligned}$$

The last (m^{th}) term can be pulled out of the sum in the denominator and considered separately. The exponent of this term is zero, so the term is 1.

$$\begin{aligned}
P(y^j | X) &= \frac{\exp(w_{0,j} - w_{0,m} + [w_{1,j} - w_{1,m}]X)}{\exp(w_{0,m} - w_{0,m} + [w_{1,m} - w_{1,m}]X) + \sum_{j'=1}^{m-1} \exp(w_{0,j'} - w_{0,m} + [w_{1,j'} - w_{1,m}]X)} \\
&= \frac{\exp(w_{0,j} - w_{0,m} + [w_{1,j} - w_{1,m}]X)}{\exp(0) + \sum_{j'=1}^{m-1} \exp(w_{0,j'} - w_{0,m} + [w_{1,j'} - w_{1,m}]X)} \\
P(y^j | X) &= \frac{\exp(w_{0,j} - w_{0,m} + [w_{1,j} - w_{1,m}]X)}{1 + \sum_{j'=1}^{m-1} \exp(w_{0,j'} - w_{0,m} + [w_{1,j'} - w_{1,m}]X)}
\end{aligned}$$

If we let $W_{kj} = w_{k,j} - w_{k,m}$ then this equation is of the form:

$$P(y^j | X) = \frac{\exp(W_{0,j} + W_{1,j}X)}{1 + \sum_{j'=1}^{m-1} \exp(W_{0,j'} + W_{1,j'}X)}$$

There are now only $2*(m-1)$ W_{kj} terms, with $k=[0,1]$ and $j=[1,2,\dots,m-1]$

To transform back to a parameterization with $2m$ parameters, we keep each W_{kj} as is and let $W_{k,m} = 0$. If we wish to recover the original values of $w_{k,j}$ we must know the original value of $w_{k,m}$ and let $w_{k,j} = W_{k,j} + w_{k,m}$ for all k,j .

(2)

If X is the sole parent of Y , X is m -valued, and Y is binary-valued, $P(y^1|X)$ can be parameterized by Eq. 4.4:

$$P(y^1 | X) = \text{sigmoid}\left(w_0 + \sum_{j=1}^m [w_{1,j} * 1\{X = x^j\}]\right)$$

This uses m parameters for $w_{1,j}$ and an $(m+1)^{\text{st}}$ parameter w_0 . A simple reparameterization can use only m parameters. Since the sum over m will contain only one nonzero term, we let $W_{1,j} = w_{1,j} + w_0$ for all j and we can rewrite $P(y^1|X)$ as:

$$P(y^1 | X) = \text{sigmoid}\left(\sum_{j=1}^m [W_{1,j} * 1\{X = x^j\}]\right)$$

This has only m parameters. This can be transformed back into a parameterization with $m+1$ parameters by letting $w_{1,j} = W_{1,j}$ and letting $w_{0,j} = 0$. To retrieve the original values of the $m+1$ parameter version, we must know the value of w_0 , and let $w_{1,j} = W_{1,j} - w_0$ for all j .

5.1

20/20

(Proof by contradiction) Assume that P factorizes over H :

$$P(X_1, X_2, X_3, X_4) = \frac{\pi_1(X_1, X_2)\pi_2(X_2, X_3)\pi_3(X_3, X_4)\pi_4(X_4, X_1)}{Z}$$

Proposition 1: $P(x_1^1, x_2^1, x_3^0, x_4^1) = 0$ implies that $\pi_1(x_1^1, x_2^1) = 0$ or $\pi_2(x_2^1, x_3^0) = 0$ or $\pi_3(x_3^0, x_4^1) = 0$ or $\pi_4(x_4^1, x_1^1) = 0$ this follows since P factorizes over H . On the other hand we also have that:

$$\begin{aligned} P(x_1^1, x_2^1, x_3^1, x_4^0) &= 1/8 \text{ implies that } \pi_1(x_1^1, x_2^1) \neq 0 \\ P(x_1^1, x_2^1, x_3^0, x_4^0) &= 1/8 \text{ implies that } \pi_2(x_2^1, x_3^0) \neq 0 \\ P(x_1^0, x_2^0, x_3^0, x_4^1) &= 1/8 \text{ implies that } \pi_3(x_3^0, x_4^1) \neq 0 \\ P(x_1^1, x_2^1, x_3^1, x_4^1) &= 1/8 \text{ implies that } \pi_4(x_4^1, x_1^1) \neq 0 \end{aligned}$$

which is a contradiction with the previous proposition.

4.5.2

20/20

In this exercise, you will prove that the modified energy functions $\varepsilon'_1[A, B]$ and $\varepsilon'_2[A, B]$ of Figure 5.9 result in precisely the same distribution as our original energy functions.

More generally, for any constants λ^1 and λ^0 , we can redefine:
 $\varepsilon'_1[a, b^i] := \varepsilon_1[a, b^i] + \lambda^i$