Graphical Models: Homework 1

vishvAs vAsuki

February 9, 2010

Notation. V represent the set of vertices. E represents the set of edges.

1 Topological Numbering

Lemma 1.0.1. *If a directed graph G is acyclic, it has a topological numbering.*

Proof. Take a DAG G = (V, E). We show that G has a topological numbering by constructing one.

For every node $v \in V$, define level l(v) to be the length of the longest directed path ending in v. If there are no directed paths terminating in v, define l(v) = 0.

Since there are a finite number of directed paths of length n or less in a DAG, and because there are no cycles in G, one can find $l(v) \forall v \in V$. Observe that, for any $u \in V$, there can only be an incoming path from v if $l(v) < l(u)$: otherwise, l(u) would be atleast l(v)+1.

Algorithm 1 Topological Numbering Generator

Input: Graph G.
Find $l(v) \forall v \in V$, set $i := 0$, $n := 1$.
repeat
  Take $S = \{v : v \in V, l(v) = i\}$. Assign topological numbers from $[n, n + |S| - 1] \cap N$ arbitrarily to nodes in S.
  $i := i + 1, n := n + |S| - 1$.
until i = n

Algorithm 1 produces a valid topological ordering, because, given a node v, no ancestor u of v has a gigher topological number assigned to it. This is inturn because u would have a lower level, and Algorithm 1 would have assigned it a smaller topological number due to its definition.

□

Lemma 1.0.2. *If a directed graph G is cyclic, it does not have a topological numbering.*

Proof. Proof by contradiction.

Suppose that there were a topological ordering $f : V \to N$. Take two nodes u, v, paricipating in a cycle in G. Because of our assumption that f is a topological ordering, $f(u) < f(v)$ as there is a path from u to v. By the same assumption, $f(v) < f(u)$ as there is a path from u to v. Since both of these cannot be true, our supposition that there is a topological ordering f must have been wrong. □

Theorem 1.0.3. *A directed graph has a topological numbering if and only if it is acyclic.*

Proof. This follows from the lemmata proved. □

2 Consistency

2.1 Problem setup

Take a DAG G. Let π(i) :=parents of i. For each i, let $f_i(x_i, x_{\pi(i)})$ be such that $\sum_{x_i} f_i(x_i, x_{\pi(i)}) = 1, f_i(x_i, x_{\pi(i)}) \geq 0$. p(x) :=$\prod_i f_i(x_i, x_{\pi(i)})$.

Theorem 2.1.1. *p(x) is a probability distribution, ie:* $\sum_x p(x) = 1$.

Proof. Consider the following expression.

$$\sum_x p(x) \quad = \quad \sum_x \prod_i f_i(x_i, x_{\pi(i)})$$

We prove that this sums to 1 using induction on decreasing topological numbers. Take a topological numbering $t : V \to N$ for G. Let $x_{[i]}$ represent $x_s : t(x_s) = i$.

Let $g(s) := \sum_{x_{[s]}..x_{[n]}} \prod_i f_i(x_i, x_{\pi(i)})$. We know that g(n) = 1 from the definition of f_i. Make the inductive assumption that g(s) = 1. If this is the case, we see that g(s-1) = 1, as follows.

$g(s-1) = \sum_{x_{[s-1]}} f_{j=t^{-1}(s)}(x_j, x_{\pi(j)}) \sum_{x_{[s]}..x_{[n]}} \prod_{i=t^{-1}(u):\forall u \geq s} f_i(x_i, x_{\pi(i)}) = \sum_{x_{[s-1]}} f_{j=t^{-1}(s)}(x_j, x_{\pi(j)}) = 1$. Hence, by the principle of mathematical induction, we see that $g(1) = \sum_x p(x) = 1$. □

Notation. As we have proved p() to be a probability distribution, we will begin using Pr().

Lemma 2.1.2. $Pr(x_i|x_{\pi_i}) = f_i(x_i, x_{\pi_i}) \forall i$.

Proof. Take a topological numbering $t : V \to N$ for G. For any i, we can sum out all $x_{j:t(j)>t(i)}$, as in the previous proof. We use this below.

$$
\begin{aligned}
Pr(x_{j:t(j)\le t(i)}) &= \prod_{j:t(j)\le t(i)} f_j(x_j, x_{\pi_j}) \\
Pr(x_{j:t(j)<t(i)}) &= \prod_{j:t(j)<t(i)} f_j(x_j, x_{\pi_j}) \\
\therefore Pr(x_i|x_{j:t(j)<t(i)}) &= \frac{Pr(x_{j:t(j)\le t(i)})}{Pr(x_{j:t(j)<t(i)})} \\
&= f_i(x_i, x_{\pi_i})
\end{aligned}
$$

But, observe that, irrespective of what values $x_{j:t(j)<t(i)\wedge j\notin\pi_i}$, $Pr(x_i|x_{j:t(j)<t(i)}) = f_i(x_i, x_{\pi_i})$, so $x_i \perp x_{j:t(j)<t(i)\wedge j\notin\pi_i}$. So, $Pr(x_i|x_{\pi_i}) = Pr(x_i|x_{j:t(j)<t(i)}) = f_i(x_i, x_{\pi_i})$ □

3 Tree Model Factorization

3.1 Problem setup

Undirected graphical model T is a tree. $Pr(x)$ associated with T. $Pr(x_i) :=$ marginal probability for random variable i.

Theorem 3.1.1.

$$
Pr(x) = \prod_i Pr(x_i) \prod_{i,j\in T} \frac{Pr(x_i, x_j)}{Pr(x_i)Pr(x_j)}
$$

.

Proof. We prove this by induction on the size of the tree. The statement is obviously true for a tree of size 1. Assume that it is true for any T with $|V| = n$.

Consider T' with $|V| = n + 1$. Pick any (leaf) node u with just 1 edge (u, v). Decompose T' into $T = T - u$ with n nodes, and u. Let x_T represent values of variables represented in T.

We have proved in solution to another problem that, $X \perp Y|Z \equiv Pr(x, y, z) = \frac{Pr(x,z)Pr(y,z)}{Pr(z)}$. We use this here. As $u \perp (T - v)|v$, we have

$$
\begin{aligned}
Pr(X_{T'}) &= \frac{Pr(X_T)Pr(u, v)}{Pr(v)} \\
&= \frac{Pr(u, v)}{Pr(v)} \prod_{i\in T} Pr(x_i) \prod_{i,j\in T} \frac{Pr(x_i, x_j)}{Pr(x_i)Pr(x_j)} \\
&= \prod_{i\in T'} Pr(x_i) \prod_{i,j\in T'} \frac{Pr(x_i, x_j)}{Pr(x_i)Pr(x_j)}
\end{aligned}
$$

□

4 Conditional Independence

4.1 a

Claim 4.1.1. *$X \perp Y \implies X \perp Y|Z$ is false.*

Proof. Take Z = X + Y with $Y, X \sim u[0,1]$. □

4.2 b

Claim 4.2.1. *$X \perp Y|W \wedge X \perp Z|W \implies X \perp (Y,Z)|W$ is false.*

Proof. Take X = (Y+Z)W with $Y, Z, W \sim u[0,1]$. □

4.3 c

Claim 4.3.1. *$X \perp (Y,Z)|W \implies X \perp Y|W$ is true.*

Proof.

$$\begin{aligned}
X \perp (Y,Z)|W &\implies \\
Pr(X,Y,Z|W) &= Pr(X|W)Pr(Y,Z|W) \\
\therefore \sum_Z Pr(X,Y,Z|W) &= Pr(X|W)\sum_Z Pr(Y,Z|W) \\
\therefore Pr(X,Y|W) &= Pr(X|W)Pr(Y|W) \\
&\implies X \perp Y|W
\end{aligned}$$

□

4.4 d

Claim 4.4.1. *$X \perp Y|Z \wedge W = f(X) \implies X \perp Y|Z,W$ is true.*

Proof.

$$\begin{aligned}
X \perp Y|Z &\implies \\
Pr(X,Y|Z) &= Pr(X|Z)Pr(Y|Z) \\
\frac{Pr(X,Y|Z)}{Pr(W)} &= \frac{Pr(X|Z)Pr(Y|Z)}{Pr(W)} \\
Pr(X,Y|Z,W) &= Pr(X|Z,W)Pr(Y|Z)
\end{aligned}$$

But, as W = f(X), $(Z, Y) \perp W|X$; so, $Pr(W|X,Y,Z) = Pr(W|X)$. Also $(Z) \perp W|X$; so, $Pr(W|X,Z) = Pr(W|X)$. We use these below.

$$\begin{aligned} Pr(Y,W|Z) &= \sum_X Pr(X,Y,W|Z) \\ &= \sum_X Pr(X,Y|Z)Pr(W|X,Y,Z) \\ &= \sum_X Pr(Y|Z)Pr(X|Z)Pr(W|X) \\ &= Pr(Y|Z)\sum_X Pr(X|Z)Pr(W|X,Z) \\ &= Pr(Y|Z)Pr(W|Z) \end{aligned}$$

So, $Pr(Y|Z) = Pr(Y|Z,W)$. Using this in the earlier equation:

$$\begin{aligned} Pr(X,Y|Z,W) &= Pr(X|Z,W)Pr(Y|Z) \\ &= Pr(X|Z,W)Pr(Y|Z,W) \end{aligned}$$

This is what we wanted to show. □

4.5 e

Claim 4.5.1. *$X \perp Y|Z \wedge X \perp Y|(W,Z) \implies X \perp (Y,W)|Z$ is true.*

Proof.

$$\begin{aligned} Pr(X,Y,W|Z) &= Pr(X|Z)Pr(Y|X,Z)Pr(W|X,Y,Z) \\ &= Pr(X|Z)Pr(Y|Z)Pr(W|Y,Z) \ \texttt{Using } \perp \texttt{ conditions.} \\ &= Pr(X|Z)Pr(Y,W|Z) \end{aligned}$$

This is what we wanted to show. □

5 Factorization

5.1 1

Lemma 5.1.1. *If $X \perp Y|Z$, Pr(x, y, z) = Pr(x, z)Pr(y,z)/ Pr(z).*

Proof. If $X \perp Y|Z$, $Pr(y|x,z) = Pr(y|z)$. We use this below.

$$
\begin{aligned}
Pr(x,y,z) &= Pr(x,z)Pr(y|x,z) \\
&= Pr(x,z)Pr(y|x,z) \\
&= Pr(x,z)Pr(y|z) \\
&= Pr(x,z)Pr(y,z)/Pr(z)
\end{aligned}
$$

□

Lemma 5.1.2. *If Pr(x, y, z) = Pr(x, z)Pr(y,z)/ Pr(z),* $X \perp Y|Z$.

Proof.

$$
\begin{aligned}
Pr(x,y,z) &= Pr(x,z)Pr(y,z)/Pr(z) \\
Pr(x,z)Pr(y|x,z) &= Pr(x,z)Pr(y,z)/Pr(z) \\
\therefore Pr(y|x,z) &= Pr(y|z) \\
\therefore X \perp Y|Z &
\end{aligned}
$$

□

5.2 2

Lemma 5.2.1. *If* $X \perp Y|Z$, *Pr(x, y, z) = f(x, z)g(y, z).*

Proof. We showed in answer to an earlier question: Pr(x, y, z) = Pr(x, z)Pr(y,z)/ Pr(z). Take f(x, z) = Pr(x, z)/ Pr(z), g(y, z) = Pr(y,z). □

Lemma 5.2.2. *If Pr(x, y, z) = f(x, z)g(y, z),* $X \perp Y|Z$.

Proof.

$$\begin{aligned}
\frac{Pr(x,y,z)}{Pr(z)} &= f(x,z)g(y,z)/Pr(z) \\
Pr(x,y|z) &= f(x,z)g(y,z)/Pr(z) \\
\therefore Pr(y|z) &= (\sum_x f(x,z))g(y,z)/Pr(z) \\
\therefore Pr(x|z) &= f(x,z)(\sum_y g(y,z))/Pr(z) \\
\therefore Pr(x|z)Pr(y|z) &= \frac{f(x,z)g(y,z)}{Pr(z)Pr(z)}(\sum_y g(y,z))(\sum_x f(x,z)) \\
&= \frac{f(x,z)g(y,z)}{Pr(z)Pr(z)}(\sum_{x,y} g(y,z)f(x,z)) \\
&= \frac{f(x,z)g(y,z)}{Pr(z)Pr(z)}(\sum_{x,y} Pr(x,y,z)) \\
&= \frac{f(x,z)g(y,z)}{Pr(z)} \\
&= Pr(x,y|z)
\end{aligned}$$

□

6 Positive Density

Theorem 6.0.3. *$Pr(x,y,z) > 0$. $X \perp Y|Z \wedge X \perp Z|Y \implies X \perp (Y,Z)$.*

Proof. [**Incomplete**] □

Remark. When the condition $Pr(x,y,z) > 0$ is relaxed, this does not necessarily hold. Consider: $X = Y + Z$, with $Y, Z \sim u[0,1]$.

7 Separation Example

7.1 Problem setup

A graph was given, which is not reproduced here.

7.2 Independent pairs of RV's

It is the union of the following sets.

$\{\{u,v\} : u \in \{1,6,4\}, v \in V - \{1,6,4\}\}$.

$\{\{u,v\} : u \in \{2,10,3\}, v \in \{7,8\}\}$.

$\{\{7,8\}\}$.

7.3 Independence from 1 given 2, 9

$\{7, 10, 3, 5, 8\}$.

7.4 Independence from 8 given 2, 9

$\{5, 6, 1\}$.

8 Bayesian Network Marginalization

Theorem 8.0.1. *Let G = (V, E) be the DAG corresponding to Pr(x). The DAG corresponding to $Pr(x_{V-A})$ is obtained as follows: Take subgraph S in G induced by (V - A). For every $(u, v) \in (V - A)^2$, add a new edge if $\exists$ a directed path (u, s, v) in G, such that s is a sequence of vertices in A.*

Proof. S corresponds to $Pr(x_{V-A})$ iff the factorization obtained by summing out x_A from factorization of Pr(x) according to G corresponds to the form derived using S. We show that this is the case.

$$
\begin{aligned}
Pr(x_{V-A}) &= \sum_{x_A} Pr(x) \\
&= \sum_{x_A} \prod_i Pr(x_i|\pi_i)
\end{aligned}
$$

Take a topological numbering t on V. All variables in A which are downstream from every node in V-A can be removed from the summation using the same logic used in proving theorem 2.1.1 earlier.

Variables in A which are not downstream from any variables in V-A can be eliminated from the summation, using the same strategy explained for the class of nodes below.

Now, only nodes in A which lie on some directed path (u, s, v) in G, so that s is a sequence of nodes in A, remain to be removed. Let par(s) be the parents of s in V-A. Now, the factor for v can be replaced with $Pr(u, s|par(s))$, summing s out, we end up with the factor $Pr(u|par(s))$. □