COMPSCI 688: Probabilistic Graphical Models

Lecture 3: Directed Graphical Models

Dan Sheldon

Manning College of Information and Computer Sciences
University of Massachusetts Amherst

Partially based on materials by Benjamin M. Marlin (marlin@cs.umass.edu) and Justin Domke (domke@cs.umass.edu)

Bayesian Networks

Review

- Conditional independence

$$\begin{aligned} \mathbf{X} \perp \mathbf{Y} | \mathbf{Z} &\iff p(\mathbf{y}, \mathbf{x} | \mathbf{z}) = p(\mathbf{x}|\mathbf{z}) p(\mathbf{y}|\mathbf{z}) \\ &\iff p(\mathbf{x} | \cancel{\mathbf{y}}, \mathbf{z}) = p(\mathbf{x}|\mathbf{z}) \end{aligned}$$

- Directed acyclic graph (DAG) G: parents, children, descendants, non-descendants
- Bayes net: distribution is factorized. Each variable i "only depends on" it's parents

$$p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i \mid \mathbf{x}_{\mathsf{pa}(i)})$$

Conditional Independence and Factorization

Conditional Independence and Factorization

We assumed factorization in a Bayes net: $p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i \mid \mathbf{x}_{\mathsf{pa}(i)})$. What does this have to do with conditional independence?

Claim: for a probability distribution $p(\mathbf{x})$

Fix G

$$p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i \mid \mathbf{x}_{\mathsf{pa}(i)}) \quad \Longleftrightarrow \quad X_i \perp \mathbf{X}_{\mathsf{nd}(i)} \mid \mathbf{X}_{\mathsf{pa}(i)} \text{ for all } i$$

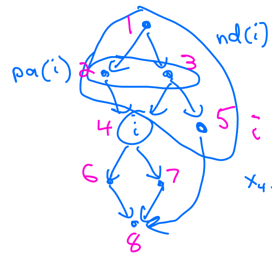
factorization $\Longleftrightarrow$ conditional independence

- RHS in words: X_i **is conditionally independent of its non-descendants given its parents**

$p(x)$, G

Conditional Independence Implies Factorization

Assume $X_i \perp \mathbf{X}_{\mathsf{nd}(i)} \mid \mathbf{X}_{\mathsf{pa}(i)}$ for all i



$x_4 \perp x_1, x_5 \mid x_2, x_3$

$p(x_5 \mid x_1, x_2, x_3, x_4)$

1. Number nodes in G according to topo ordering

$i \to j \Rightarrow i < j$

$\Rightarrow de(i) \subseteq \{i+1, \dots, n\}$

$\Rightarrow \{1, 2, \dots, i-1\} \subseteq nd(i)$

2. Chain rule

$p(x_1, \dots, x_n) = p(x_1) p(x_2 \mid x_1) \cdots p(x_n \mid x_1, \dots, x_{n-1})$

$= \prod_i p(x_i \mid x_1, \dots, x_{i-1})$

$= \prod_i p(x_i \mid x_{pa(i)}, \cancel{x_{\{1,\dots,i-1\} \setminus pa(i)}})$ by CI

3. Split into parents + non-descendants

$= \prod_i p(x_i \mid x_{pa(i)})$

4. Used CI assumptions to simplify

$A \perp (B, C) \mid D \Rightarrow A \perp B \mid D$
$A \perp C \mid D$

Review of Argument

0. Assume $X_i \perp \mathbf{X}_{\mathsf{nd}(i)} \mid \mathbf{X}_{\mathsf{pa}(i)}$ for all i
1. Number nodes according to a topological ordering: $i \to j \implies i < j$. Then we also have that $\mathsf{de}(i) \subseteq \{i+1, \dots, n\}$, and, as a consequence all nodes in $\{1, \dots, i-1\}$ are *non-descendants*
2. Use the chain rule
$$p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i \mid \mathbf{x}_{\{1,\dots,i-1\}})$$
3. Split into parents and other non-descendants
$$p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i \mid \mathbf{x}_{\mathsf{pa}(i)}, \mathbf{x}_{\{1,\dots,i-1\} \setminus \mathsf{pa}(i)})$$
4. Simplify using conditional independence
$$p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i \mid \mathbf{x}_{\mathsf{pa}(i)})$$

Factorization Implies Conditional Independence

$a \to b \to c$, $a \to c$ $\quad p(a,b,c) = p(a)\,p(b|a)\,p(c|b,a)$

$a \to b$ $\quad p(a,b) = p(a)\,p(b|a)$

To show this, first we'll argue that marginalizing *descendants* in a Bayes net is easy:

Warmup: suppose j is a node with no children in a Bayes net (a "leaf"). Then

Lemma

$$p(\mathbf{x}_{-j}) = \prod_{i \neq j} p(x_i | \mathbf{x}_{\mathsf{pa}(i)}) = \sum_{x_j} p(x_{-j}, x_j)$$

In words: can marginalize x_j by dropping factor $p(x_j | x_{\mathsf{pa}(j)})$ to get a Bayes net with one less node.

This is *only* true for leaf nodes. Marginalizing non-leaf nodes may be very hard!

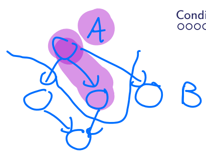
$p(a)\,p(b|a)\,p(c|b,a)$

Proof:

$$\begin{aligned} p(\mathbf{x}_{-j}) &= \sum_{x_j} p(\mathbf{x}_{-j}, x_j) \\ &= \sum_{x_j} p(x_j | \mathbf{x}_{\mathsf{pa}(j)}) \boxed{\prod_{i \neq j} p(x_i | \mathbf{x}_{\mathsf{pa}(i)})} \\ &= \prod_{i \neq j} p(x_i | \mathbf{x}_{\mathsf{pa}(i)}) \cdot \cancel{\underbrace{\sum_{x_j} p(x_j | \mathbf{x}_{\mathsf{pa}(j)})}_{1}} \end{aligned}$$

Pushing the sum inside in the last line is possible because j is a leaf.

Marginalizing a *Set* of Descendants



Lemma: suppose A and B partition the nodes of a Bayes net and there is no path from B to A. Then

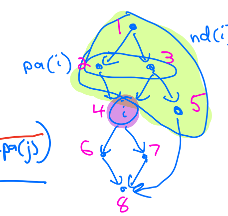
$$p(\mathbf{x}_A) = \sum_{\mathbf{x}_B} p(\mathbf{x}_A, \mathbf{x}_B) = \prod_{i \in A} p(x_i | \mathbf{x}_{\mathsf{pa}(i)})$$

Proof idea: at least one node in B is a leaf. Eliminate it using the warmup lemma and then repeat.

Factorization Implies Conditional Independence

$X_i \perp X_{nd(i)} \mid X_{pa(i)}$

Assume $p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i \mid \mathbf{x}_{\mathsf{pa}(i)})$. Then for any i



$$p(x_i | x_{nd(i)}) = \frac{p(x_i, x_{nd(i)})}{p(x_{nd(i)})}$$

$$= \frac{p(x_i | x_{pa(i)}) \cdot \cancel{\prod_{j \in nd(i)} p(x_j | x_{pa(j)})}}{\cancel{\prod_{j \in nd(i)} p(x_j | x_{pa(j)})}}$$

$$= p(x_i | x_{pa(i)})$$

$$\Rightarrow X_i \perp X_{nd(i)} \mid X_{pa(i)}$$

$p(a,b,c)$ $\xrightarrow{\text{marginalize } c}$ $p(a,b)$

$$\begin{tikzcd} a \arrow[r] \arrow[rr, bend left] & b \arrow[r] & c \end{tikzcd}$$

$p(a)$ $p(b|a)$ $p(c|b,a)$

$$\begin{tikzcd} a \arrow[r] & b \end{tikzcd}$$

$p(a)$ $p(b|a)$

drop $p(a)$

$$p(b,c) \stackrel{?}{=} p(b|a)\,p(c|b,a)$$

$$p(b,c) = \sum_a p(a)\,p(b|a)\,p(c|b,a)$$