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

$$\mathbf{X} \perp \mathbf{Y} | \mathbf{Z} \iff p(\mathbf{y}, \mathbf{x}|\mathbf{z}) = p(\mathbf{x}|\mathbf{z})p(\mathbf{y}|\mathbf{z})$$
$$\iff p(\mathbf{x}|\mathbf{y}, \mathbf{z}) = p(\mathbf{x}|\mathbf{z})$$

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

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

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 $p(\mathbf{x})$

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

factorization $\iff$ conditional independence

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

Conditional Independence Implies Factorization

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

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 $\mathsf{de}(i) \subseteq \{i+1, \ldots, n\}$, and, as a consequence all nodes in $\{1, \ldots, i-1\}$ are *non-descendants*
2. Use the chain rule
$$p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i | \mathbf{x}_{\{1,\ldots,i-1\}})$$
3. Split into parents and other non-descendants
$$p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i | \mathbf{x}_{\mathsf{pa}(i)}, \mathbf{x}_{\{1,\ldots,i-1\}\ \mathsf{pa}(i)})$$
4. Simplify using conditional independence
$$p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i | \mathbf{x}_{\mathsf{pa}(i)})$$

Factorization Implies Conditional Independence

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

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

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!

Proof:

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

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

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

$$\begin{aligned} p(x_i | \mathbf{x}_{\mathsf{nd}(i)}) &= \frac{p(x_i, \mathbf{x}_{\mathsf{nd}(i)})}{p(\mathbf{x}_{\mathsf{nd}(i)})} \\ &= \frac{p(x_i | \mathbf{x}_{\mathsf{pa}(i)}) \cdot \prod_{j \in \mathsf{nd}(i)} p(x_j | \mathbf{x}_{\mathsf{pa}(j)})}{\prod_{j \in \mathsf{nd}(i)} p(x_j | \mathbf{x}_{\mathsf{pa}(j)})} \quad \text{Use lemma twice} \\ &= p(x_i | \mathbf{x}_{\mathsf{pa}(i)}) \end{aligned}$$

This demonstrates that $X_i \perp \mathbf{X}_{\mathsf{nd}(i)} | \mathbf{X}_{\mathsf{pa}(i)}$ for all i.