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)

Review

- Conditional independence
  \[ X \perp Y | Z \iff p(y, x | z) = p(x | z)p(y | z) \]
  \[ \iff p(x | y, z) = p(x | z) \]

- Directed acyclic graph (DAG) \( G \): parents, children, descendants, non-descendants
- Bayes net: distribution is factorized. Each variable \( i \) "only depends on" it's parents
  \[ p(x) = \prod_{i=1}^{N} p(x_i | x_{pa(i)}) \]
Bayesian Networks

Conditional Independence and Factorization

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

**Claim:** for a probability distribution \( p(x) \)

\[
p(x) = \prod_{i=1}^{N} p(x_i | x_{pa(i)}) \iff X_i \perp X_{nd(i)} | X_{pa(i)} \text{ for all } i
\]

▶ RHS in words: \( X_i \) is conditionally independent of its non-descendants given its parents

Conditional Independence Implies Factorization

Assume \( X_i \perp X_{nd(i)} | X_{pa(i)} \) for all \( i \)

Review of Argument

0. Assume \( X_i \perp X_{nd(i)} | X_{pa(i)} \) for all \( i \)
1. Number nodes according to a topological ordering: \( i \to j \iff i < j \). Then we also have that \( de(i) \subseteq \{i + 1, \ldots, n\} \), and, as a consequence all nodes in \( \{1, \ldots, i-1\} \) are non-descendants
2. Use the chain rule

\[
p(x) = \prod_{i=1}^{N} p(x_i | x_{\{1,\ldots,i-1\}})
\]
3. Split into parents and other non-descendants

\[
p(x) = \prod_{i=1}^{N} p(x_i | x_{pa(i)}, x_{\{1,\ldots,i-1\}} pa(i))
\]
4. Simplify using conditional independence

\[
p(x) = \prod_{i=1}^{N} p(x_i | x_{pa(i)})
\]
Bayesian Networks

Conditional Independence and Factorization

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

\[
p(x_{-j}) = \prod_{i \neq j} p(x_i | x_{pa(i)})
\]

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

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

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

\[
p(x_A) = \sum_{x_B} p(x_A, x_B) = \prod_{i \in A} p(x_i | x_{pa(i)})
\]

**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 \( p(x) = \prod_{i=1}^N p(x_i | x_{pa(i)}) \). Then for any \( i \)

\[
p(x_i | x_{nd(i)}) = \frac{p(x_i, x_{nd(i)})}{p(x_{nd(i)})} = \frac{p(x_i | x_{pa(i)}) \cdot \prod_{j \in nd(i)} p(x_j | x_{pa(j)})}{\prod_{j \in nd(i)} p(x_j | x_{pa(j)})}
\]

Use lemma twice

This demonstrates that \( X_i \perp X_{nd(i)} | X_{pa(i)} \) for all \( i \).