COMPSCI 688: Probabilistic Graphical Models

Lecture 6: Undirected Graphical Models

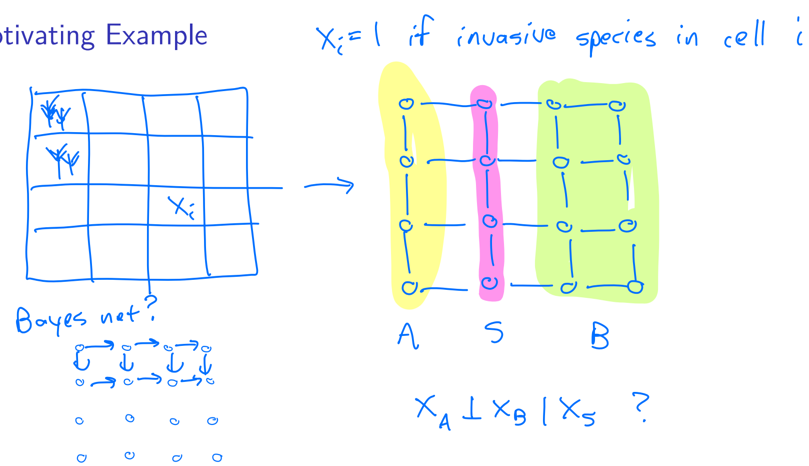
Dan Sheldon

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)

Motivation

Motivating Example

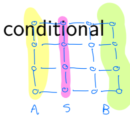


Motivating Example

Markov Properties for Undirected Graphical Model

Undirected graphical models are probability distributions that satisfy a set of conditional independence properties with respect to a *dependence graph* $\mathcal{G}$. Formally:

undirected



- Let $\mathcal{G} = (V, E)$ be a graph with nodes $V = \{1, \ldots, n\}$
- For $A, B, S \subseteq V$, say that S separates A from B if all paths from A to B in $\mathcal{G}$ go through S, written $\text{sep}_{\mathcal{G}}(A, B|S)$.

The joint distribution of random variables $X_1, \ldots, X_n$ satsifes the **global Markov property** with respect to $\mathcal{G}$ if

$$\text{sep}_{\mathcal{G}}(A, B|S) \implies \mathbf{X}_A \perp \mathbf{X}_B \mid \mathbf{X}_S \qquad \text{(G)}$$

What form of distribution $p(x_1, \ldots, x_n)$ has this property?

HW due Mon
Quiz due Fri

Undirected graphical models

Bayes nets
- directed
- factorization
- $X_i \perp X_{nd(i)} \mid X_{pa(i)}$
- d-separation
- learning (easy)

Markov Random Fields
- undirected
- factorization ("potentials")
- Markov properties based on graph separation
- inference
- learning (harder)

Warmup: Characterization of Conditional Independence

Recall the definition of conditional independence

$$p(x,y,z) = p(z)p(x|z)p(y|z) \quad [\phi_1(x,z),\ \phi_2(y,z)]$$

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

Today we'll use two other properties of conditional independence:

1. $\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z} \iff p(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \phi_1(\mathbf{x}, \mathbf{z})\phi_2(\mathbf{y}, \mathbf{z})$ for some ϕ_1, ϕ_2 ≥ 0
2. $\mathbf{X} \perp (\mathbf{Y}, \mathbf{W}) \mid \mathbf{Z} \implies \mathbf{X} \perp \mathbf{Y} | \mathbf{Z}$

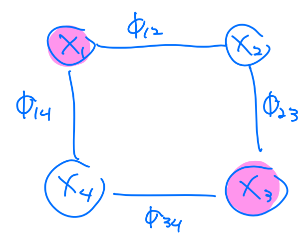
Proofs: exercise

Note: (1) says that conditional independence holds iff the joint distribution factorizes in a certain way, which is very important.

Markov Random Field Example

Example: $p(x_1, x_2, x_3, x_4) = \phi_{12}(x_1, x_2)\phi_{23}(x_2, x_3)\phi_{34}(x_3, x_4)\phi_{14}(x_1, x_4) \cdot \frac{1}{Z}$

$= f(x_1, x_2, x_3) \cdot g(x_1, x_3, x_4)$



$X_2 \perp X_4 \mid X_1$? NO

$\Rightarrow$ $X_2 \perp X_4 \mid X_1, X_3$

$X_1 \perp X_3 \mid X_2, X_4$ (symmetry)

no Bayes net satisfies these CIs

Markov Random Fields

$p(x_1,x_2,x_3,x_4) = \frac{1}{Z}\phi_{12}(x_1,x_2)\phi_{123}(x_1,x_2,x_3)\phi_{34}(x_3,x_4)$

$c = \{1,2,3\}$ $\quad \mathcal{C} = \{\{1,2\},\{1,2,3\},\{3,4\}\}$

$x_c = (x_1,x_2,x_3)$

A Markov random field is a probability distribution that factorizes over a set of "cliques" $\mathcal{C}$:

$$p(\mathbf{x}) = \frac{1}{Z}\prod_{c\in\mathcal{C}}\phi_c(\mathbf{x}_c), \quad Z = \sum_{\mathbf{x}}\prod_{c\in\mathcal{C}}\phi_c(\mathbf{x}_c)$$

(unnormalized prob.) $\quad f(x_1,x_4,x_5)$ $\quad \mathrm{Scope}(f) = \{1,4,5\}$

- Each $c \subseteq V = \{1,\ldots,n\}$ is a set of indices, or "clique"
- The function ϕ_c is a non-negative *factor* or *potential*. It only depends on x_i for $i \in c$. We say it has *scope* c and define $\mathrm{Scope}(\phi_c) := c$
- Z is the normalizing constant or "partition function"

Concrete Example

$\mathcal{C} = \{\{1,2\},\{2,3\}\}$ $\quad x_i \in \{0,1\}$

$p(x_1,x_2,x_3) = \frac{1}{Z}\phi_{12}(x_1,x_2)\phi_{23}(x_2,x_3)$

x_1	x_2	ϕ_{12}
0	0	1
0	1	2
1	0	2
1	1	2

x_2	x_3	ϕ_{23}
0	0	1
0	1	1
1	0	1
1	1	2

x_1	x_2	x_3	$\phi_{12}\cdot\phi_{23}$	$p(x_1,x_2,x_3)$
0	0	0	$1\cdot1=1$	1/18
0	0	1	$1\cdot1=1$	1/18
0	1	0	$2\cdot1=2$	2/18
0	1	1	$2\cdot2=4$	⋮
1	0	0	$2\cdot1=2$	⋮
1	0	1	$2\cdot1=2$	⋮
1	1	0	$2\cdot1=2$	
1	1	1	$2\cdot2=4$	4/18
			18	

Dependence Graph

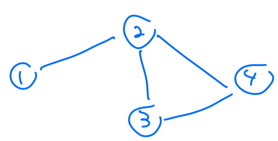
The *dependence graph* $\mathcal{G} = (V,E)$ of the MRF $p(\mathbf{x}) = \frac{1}{Z}\prod_{c\in\mathcal{C}}\phi_c(\mathbf{x}_c)$ is the graph where nodes i and j are connected by an edge if they appear together in some factor:

$$V = \{1,\ldots,n\}, \quad E = \{(i,j) : i \in c \text{ and } j \in c \text{ for some } c \in \mathcal{C}\}$$

With this definition, every $c \in \mathcal{C}$ is a clique (fully connected set) in $\mathcal{G}$.

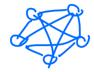
$p(x_1,x_2,x_3,x_4) = \phi_{12}(x_1,x_2)\,\phi_{234}(x_2,x_3,x_4)\,\phi_{34}(x_3,x_4)\,\frac{1}{Z}$

$\phi_{234}(x_2,x_3,x_4)$



Factorization and Markov Properties

Factorization



Let $\mathcal{G}$ be a graph. A distribution $p(\mathbf{x})$ factorizes with respect to $\mathcal{G}$ if

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \phi_c(\mathbf{x}_c), \quad \mathcal{C} = \text{cliques}(\mathcal{G}) \qquad \text{(F)}$$

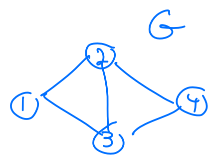
subset of nodes that is fully connected

In other words, it is an MRF with dependence graph $\mathcal{G}$.

As in Bayes nets, there is a close relationship between factorization and Markov properties obtained from graph separation.

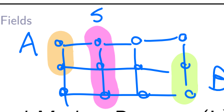
$p(x_1, x_2, x_3, x_4) = \Phi_{123}(x_1, x_2, x_3)\Phi_{234}(x_2, x_3, x_4) \cdot \frac{1}{Z}$

$\mathcal{C} = \{\{1,2,3\}, \{2,3,4\}\}$



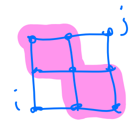
Markov Properties

$p(x_1, \ldots, x_n)$



The *global Markov property* (G), the *local Markov Property* (L) and *pairwise Markov property* (P) are three different properties of a distribution that hold relative to a graph $\mathcal{G}$.

every A→B path hits S



$$\text{sep}_{\mathcal{G}}(A, B|S) \implies \mathbf{X}_A \perp \mathbf{X}_B \mid \mathbf{X}_S \qquad \text{(G)}$$

$$i \in V \implies X_i \perp \mathbf{X}_{V \setminus (\text{nb}(i) \cup \{i\})} \mid \mathbf{X}_{\text{nb}(i)} \qquad \text{(L)}$$

$$(i,j) \notin E \implies X_i \perp X_j \mid \mathbf{X}_{V \setminus \{i,j\}} \qquad \text{(P)}$$

Above, $\text{nb}(i)$ is the set of neighbors of node i in $\mathcal{G}$.

Claim: (G) $\Rightarrow$ (L) $\Rightarrow$ (P)

It's easy to see (G) $\Rightarrow$ (L) and (G) $\Rightarrow$ (P) by taking the appropriate choices of A, B, S. We leave (L) $\Rightarrow$ (P) as an exercise.

Markov Property Examples

Markov Property Examples

Factorization Implies Markov

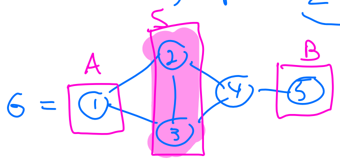
$\phi(x,z)\phi(y,z)$ $\quad x \perp y \mid z$

Like in Bayes nets, factorization implies conditional independencies (Markov properties).

Claim: (F) ⇒ (G) ⇒ (L) ⇒ (P)

Proof ("by example"): We only need to show (F) ⇒ (G).

Assume (F). $p(x) = \frac{1}{Z}\phi_{123}(x_1,x_2,x_3)\,\phi_{234}(x_2,x_3,x_4)\,\phi_{45}(x_4,x_5)$



$= f(x_1,x_2,x_3)\; g(x_2,x_3,x_4,x_5)$

$\Rightarrow x_1 \perp (x_4,x_5) \mid x_2,x_3$

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

Want: $sep_G(A,B|S) \Rightarrow x_A \perp x_B \mid x_S$

Factorization Implies Markov Proof

Factorization Implies Markov Proof

Suppose $p(\mathbf{x}) = \prod_{c\in\mathcal{C}} \phi_c(\mathbf{x}_c)$ (assume $1/Z$ is included in one of the factors) and sep$_\mathcal{G}(A,B;S)$. We'll show that $\mathbf{X}_A \perp \mathbf{X}_B \mid \mathbf{X}_S$.

First, remove S from $\mathcal{G}$. The resulting graph is disconnected and has no paths from A to B

- Let $\tilde{A}$ be the union of all connected components containing a node from A
- Let $\tilde{B} = V \setminus \tilde{A}$

Then each $c \in \mathcal{C}$ is a subset of either $\tilde{A} \cup S$ or $\tilde{B} \cup S$

- Let $\mathcal{C}_A$ be the cliques contained in $\tilde{A} \cup S$
- Let $\mathcal{C}_B$ be the cliques contained in $\tilde{B} \cup S$

Then

$$p(\mathbf{x}) = \prod_{c \in \mathcal{C}_A} \phi_c(\mathbf{x}_c) \prod_{c \in \mathcal{C}_B} \phi_c(\mathbf{x}_c) = h(\mathbf{x}_{\tilde{A}}, \mathbf{x}_S) k(\mathbf{x}_{\tilde{B}}, \mathbf{x}_S)$$

$$\implies \mathbf{X}_{\tilde{A}} \perp \mathbf{X}_{\tilde{B}} \mid \mathbf{X}_S$$

$$\iff (\mathbf{X}_A, \mathbf{X}_{\tilde{A} \setminus A}) \perp (\mathbf{X}_B, \mathbf{X}_{\tilde{B} \setminus B}) \mid \mathbf{X}_S$$

$$\implies \mathbf{X}_A \perp \mathbf{X}_B \mid \mathbf{X}_S$$

Markov Implies Factorization: Hammersley-Clifford Theorem

p(x) > 0 ∀ x

There is a famous partial converse. For a *positive* distribution, (P) ⇒ (F), which implies all the conditions are equivalent:

Theorem (Hammersley-Clifford). If $p(\mathbf{x}) > 0$ for all $\mathbf{x}$, then

$$\text{(F)} \iff \text{(G)} \iff \text{(L)} \iff \text{(P)}.$$

The theorem holds for a very general class of distributions, e.g., ones with continuous, discrete, or both types of random variables.