COMPSCI 688: Probabilistic Graphical Models

Lecture 6: Undirected Graphical Models

Dan Sheldon

Manning College of Information and Computer Sciences
University of Massachusetts Amherst

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

Motivation

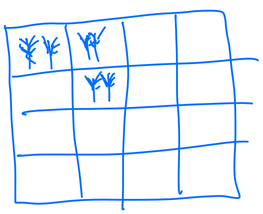
Bayes nets:

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

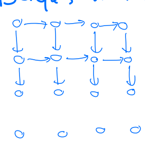
+ $\boxed{\text{inference}}$

Motivating Example

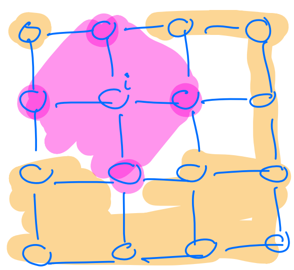
$X_i = 1$ if invasive plant present in cell i



Bayes net?



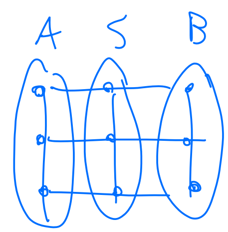
asymmetry



undirected

$X_i \perp$ non-neighbors | neighbors

Motivating Example



$X_A \perp X_B \mid X_S$

Markov Properties for Undirected Graphical Model

undirected

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

- 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 $\text{sep}_{\mathcal{G}}(A, B|S)$.

$p(x_1, \ldots, x_n)$

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?

Markov Random Fields

Warmup: Characterization of Conditional Independence

Recall the definition of conditional independence

$$p(x,y,z) = \underbrace{p(x|z)}_{\phi_1(x,z)}\, \underbrace{p(y|z)\cdot p(z)}_{\phi_2(y,z)}$$

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

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

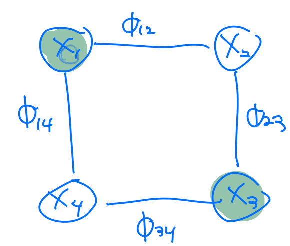
1. $\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z} \iff p(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \phi_1(\mathbf{x}, \mathbf{z})\phi_2(\mathbf{y}, \mathbf{z})$ for some ϕ_1, ϕ_2 (b)
2. $\mathbf{X} \perp (\mathbf{Y}, \mathbf{W}) \mid \mathbf{Z} \implies \mathbf{X} \perp \mathbf{Y} | \mathbf{Z}$

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



$z = (x_1, x_3)$

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

definition of CI from prev. slide

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

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

no Bayes net that satisfies these CIs

Markov Random Fields

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

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

$\phi_c(x_c)$

$c=\{1,2,3\}$ $\quad x_c = (x_1,x_2,x_3)$

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

$$p(\mathbf{x}) = \frac{1}{Z}\underbrace{\prod_{c\in\mathcal{C}}\phi_c(\mathbf{x}_c)}_{\text{unnormalized prob.}}, \quad Z = \sum_{\mathbf{x}}\prod_{c\in\mathcal{C}}\phi_c(\mathbf{x}_c)$$

normalization constant

$f(x_1,x_4)$ $\quad \text{Scope}(f)=\{1,4\}$

- 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 $\text{Scope}(\phi_c) := c$
- 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

$Z=18$

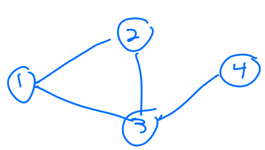
Dependence Graph

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 $\mathcal{G}$.

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



$\{1,2\}$
$\{1,2,3\}$
$\{3,4\}$

Factorization and Markov Properties

Factorization

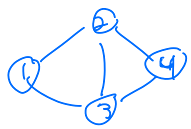
Let $\mathcal{G}$ be a graph. A distribution $p(\mathbf{x})$ factorizes with respect to $\mathcal{G}$ if

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

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

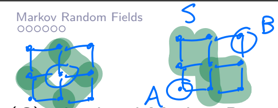
As in Bayes nets, there is a close relationship between factorization and Markov properties obtained from graph separation.

$\mathcal{C}$



$p(x_1,\dots,x_4) = \phi_{123}(x_1,x_2,x_3)\phi_{234}(x_3,x_3,x_4)$

Markov Properties



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 $\mathcal{G}$.

non-neighbors

$$\text{sep}_{\mathcal{G}}(A,B|S) \implies \mathbf{X}_A \perp \mathbf{X}_B \mid \mathbf{X}_S \tag{G}$$
$$i \in V \implies X_i \perp \mathbf{X}_{V\setminus(\text{nb}(i)\cup\{i\})} \mid \mathbf{X}_{\text{nb}(i)} \tag{L}$$
$$(i,j)\notin E \implies X_i \perp X_j \mid \mathbf{X}_{V\setminus\{i,j\}} \tag{P}$$

Above, 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

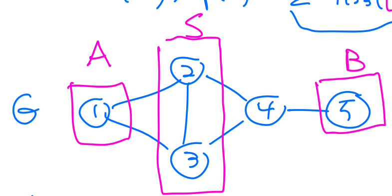
Factorization Implies Markov

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



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

$\Rightarrow X_1 \perp (X_4, X_5) \mid X_2, X_3$

$\Rightarrow X_1 \perp X_5 \mid X_2, X_3$ □

Assume $\text{sep}_G(A, B \mid S) \longrightarrow X_A \perp X_B \mid X_S$

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 $\text{sep}_{\mathcal{G}}(A, B; S)$. 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

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

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

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

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

Markov Implies Factorization: Hammersley-Clifford Theorem

$(F) \Rightarrow (G) \Rightarrow (L) \Rightarrow (P)$

$p(x) > 0$

There is a famous partial converse. For a *positive* distribution, (P) $\Rightarrow$ (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.