COMPSCI 688: Probabilistic Graphical Models
Lecture 6: Undirected Graphical Models

Dan Sheldon
Manning College of Information and Computer Sciences
University of Massachusetts Amherst

Motivating Example

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 $G$. Formally:

- Let $G = (V, E)$ 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 $G$ go through $S$, written sep$_G(A, B | S)$.

The joint distribution of random variables $X_1, \ldots, X_n$ satisfies the global Markov property with respect to $G$ if

$$\text{sep}_G(A, B | S) \implies X_A \perp X_B | X_S \quad \text{(G)}$$

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

Warmup: Characterization of Conditional Independence

Recall the definition of conditional independence

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

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

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

Proofs: exercise

Note: (1) says that conditional independence holds iff the joint distribution factorizes in a certain way, which is very important.
A Markov random field is a probability distribution that factorizes over a set of "cliques" $C$:  

$$p(x) = \frac{1}{Z} \prod_{c \in C} \phi_c(x_c)$$  

$$Z = \sum_x \prod_{c \in C} \phi_c(x_c)$$

- Each $c \subseteq V = \{1, \ldots, n\}$ is a set of indices, or "clique".
- The function $\phi_c$ 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".

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

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

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

$$p(x_1, x_2, x_3, x_4) = \frac{1}{Z} \prod_{c \in C} \phi_{c_1}(x_{c_1}) \cdot \frac{1}{2}$$

Concrete Example:  

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

$$p(x_1, x_3, x_4) = \frac{1}{2} \phi_{c_1}(x_{c_1}) \phi_{c_2}(x_{c_2}) \phi_{c_3}(x_{c_3})$$

With this factorization, we can compute $p(x_1, x_3, x_4)$ for specific values:

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
<th>$p(x_1, x_3, x_4)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1/4</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1/4</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1/4</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1/4</td>
</tr>
</tbody>
</table>

$Z = 16$
Motivation
Markov Random Fields
Factorization and Markov Properties

**Factorization**

Let $G$ be a graph. A distribution $p(x)$ factorizes with respect to $G$ if

$$p(x) = \frac{1}{Z} \prod_{c \in C} \phi_c(x_c), \quad C = \text{cliques}(G)$$

(F)

In other words, it is an MRF with dependence graph $G$.

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

**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 $G$.

$$\text{sep}_G(A, B | S) \Rightarrow X_A \perp X_B | X_S$$

$$i \in V \Rightarrow X_i \perp X_{V \setminus \{\text{nb}(i)\}\cup\{i\}} | X_{\text{nb}(i)}$$

(L)

$$\forall (i, j) \notin E \Rightarrow X_i \perp X_j | X_{V \setminus \{i,j\}}$$

(P)

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

**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**
Motivation

Markov Random Fields

Factorization and Markov Properties

Factorization Implies Markov

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

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

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

Suppose \(p(x) = \prod_{c \in C} \phi_c(x_c)\) \((\text{assume } 1/Z \text{ is included in one of the factors})\) and \(\text{sep}(A, B; S)\).

Let \(\tilde{A}\) be the union of all connected components containing a node from \(A\) to \(B\).

Let \(\tilde{B} = V \setminus \tilde{A}\).

Then each \(c \in C\) is a subset of either \(\tilde{A} \cup S\) or \(\tilde{B} \cup S\).

Let \(C_A\) be the cliques contained in \(\tilde{A} \cup S\).

Let \(C_B\) be the cliques contained in \(\tilde{B} \cup S\).
Then

\[ p(x) = \prod_{c \in C_A} \phi_c(x_c) \prod_{c \in C_B} \phi_c(x_c) = h(x_A, x_S)k(x_B, x_S) \]

\[ X_A \perp X_B \mid X_S \]

\[ (X_A, X_{\tilde{A} \backslash A}) \perp (X_B, X_{\tilde{B} \backslash B}) \mid X_S \]

\[ X_A \perp X_B \mid X_S \]

---

**Markov Implies Factorization: Hammersley-Clifford Theorem**

\[
(F) \Rightarrow (G) \Rightarrow (L) \Rightarrow (P)
\]

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

**Theorem (Hammersley-Clifford).** If \( p(x) > 0 \) for all \( x \), then

\[ (F) \iff (G) \iff (L) \iff (P) \]

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