Markov Networks

Amir Globerson

Bayesian networks can describe a wide array of distributions. As we have seen these are almost precisely the distributions that satisfy the $I_{d-sep}(G)$ properties (*almost* because they may satisfy more CI properties, but this requires a very particular choice of parameters). It is clear that there are other set of CI properties that cannot be captured by Bayes nets. We showed one example in class.

We now present a different type of model family for distributions $p(x_1, \ldots, x_n)$ that satisfy CI properties that cannot be captured by BNs. Assume we are given an undirected graph with nodes $V = 1, \ldots, n$ and edges E where $ij \in E$ is an unordered pair. Denote the set of cliques in the graph G by $\mathcal{C}$. Give a set of non-negative functions $\phi_c(\boldsymbol{x}_c)$ where $c \in \mathcal{C}$ we have the following definition of a model distribution:

Definition 1. *The Markov Network* (G, ϕ) *is the distribution given by:*

$$p(\boldsymbol{x}) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \phi_c(\boldsymbol{x}_c) \tag{1}$$

Here Z *is a normalization constant given by* $Z = \sum_{\boldsymbol{x}} \prod_{c \in \mathcal{C}} \phi_c(\boldsymbol{x}_c)$. *It is also called the partition function, a term originating in statistical physics. We shall say that if* $p(\boldsymbol{x})$ *can be written in such a form then it factors according to the undirected graph G.*

1 Conditional Independence Properties in Markov Networks

In BNs we showed that the fact that a distribution is a BN implies it has a set of conditional independence properties (namely $I_{d-sep}(G)$). Perhaps more surprisingly we also showed that the converse holds: namely, that if it has the properties $I_{d-sep}(G)$ it must be a Bayesian network for the graph G. Here we will follow a similar path for Markov Networks (MNs in what follows). We shall see that not all the results carry over simply.

We begin with a set of CI properties that is the equivalent of $I_{d-sep}(G)$.

Definition 2. *Given an undirected graph* G *and three sets of variables* W, Y, Z *corresponding to nodes in the graph, we say that* Z *separates* W *and* Y *in* G *if every path between* W *and* Y *has a node in* Z. *We now define the following conditional independence properties based on G:*

$$I_{sep}(G) = \{W \perp Y | Z : \text{Z separates W and Y in } G\} \tag{2}$$

We then have the following result:

Theorem 1. *If p factorizes according to G then* $I(p) \supseteq I_{sep}(G)$.

The proof relies on the factorization lemma we showed in class and the decomposition lemma for CI (see book).

You may wonder if the converse holds. The short answer is no (the long answer is yes, if $p(\boldsymbol{x}) > 0$ for all $\boldsymbol{x}$).

2 Other (smaller) CI sets and relation to $I_{sep}(G)$

Recall that in BNs we discussed the Local Markov (LM) property which was a subset of $I_{d-sep}(G)$. But, in fact we showed that LM was not a weaker property than $I_{d-sep}(G)$ in the sense that any distribution that satisfies the former satisfies the latter. Below we show a related result for MNs.

Define the two following sets of CI properties:

Definition 3. *Define $I_{pair}(G)$ as the following sets of CI properties*

$$I_{pair}(G) = \{X_i \perp X_j | X_{V\setminus\{i,j\}} : ij \notin E\} \tag{3}$$

Note that there are $O(n^2)$ such properties and that $I_{pair}(G) \subseteq I_{sep}(G)$.

In words, $I_{pair}(G)$ states that two non-adjacent variables are conditionally independent given the rest of the graph.

The next property says that a variable X_i is independent of the rest of the graph given its neighbors. By neighbors of i we mean the nodes j such that $ij \in E$. We denote this set by $Nbr(i)$.

Definition 4. *Define $I_{LM}(G)$ as the following sets of CI properties*

$$I_{LM}(G) = \{X_i \perp X_{Nbr(i)} | X_{V\setminus\{Nbr(i),i\}} : i = 1, \dots, n\} \tag{4}$$

Note that there are $O(n)$ such properties and that $I_{LM}(G) \subseteq I_{sep}(G)$. However, there is no strict relation between $I_{pair}(G)$ and $I_{LM}(G)$.

What is the relation between these three properties? Clearly $I_{sep}(G)$ is stronger than the other two (since it contains them).

The next theorem shows that $I_{LM}(G)$ is stronger than $I_{pair}(G)$:

Theorem 1. *If $I(p) \supseteq I_{LM}(G)$ then $I(p) \supseteq I_{pair}(G)$ (this is sometimes denoted by $I_{LM}(G) \Rightarrow I_{pair}(G)$)*

The proof is in the book and involves use of the weak union property:

$$X \perp Y, W | Z \Rightarrow X \perp Y | Z, W \tag{5}$$

To understand the converse, we shall use a stronger result which says that $I_{pair}(G) \Rightarrow I_{sep}(G)$ when $p(\boldsymbol{x})$ is strictly positive.

Theorem 2. *If $p(\boldsymbol{x})$ is a strictly positive distribution and $I(p) \supseteq I_{pair}(G)$ then $I(p) \supseteq I_{sep}(G)$.*

The proof is in the book and involves use of the following *intersection* property which holds for strictly positive distributions:

$$U \perp Y | Z, W \quad \& \quad U \perp W | Z, Y \Rightarrow U \perp Y, W | Z \tag{6}$$