Graphical Models

Ernesto Brau

Dept. of Computer Science
University of Arizona

27 Jan 2009

Outline

1. Markov Random Fields
   - Factorization
   - Relation to directed graphs
   - Thoughts
2. Inference in Graphical Models
   - Introduction
   - Inference on a chain
   - Factor graphs

Factorization

Given a graph with nodes $x_1, \ldots, x_D$, the corresponding joint density is given by

$$p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{Z} \prod_C \psi_C(\mathbf{x}_C),$$

where $\mathbf{x} = (x_1, \ldots, x_D)^{\mathsf{T}}$ and ψ_C are functions over the maximal cliques of the graph, called *potential functions*.

Potential Functions

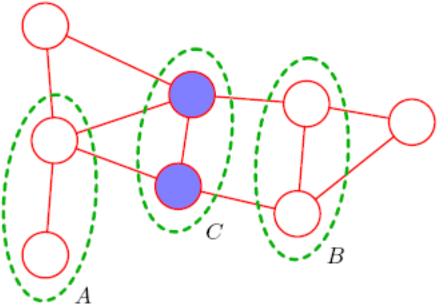
- What are pontential functions? They are functions

$$\psi_C : \mathbb{R}^{|C|} \to \mathbb{R},$$

  where $|C|$ is the number of nodes in the maximal clique C.
- They represent some relationship between adjacent nodes (random variables) in the graph.
- They need not have a probabilistic interpretation. In image de-noising example, potential functions represented correlation between nodes of (two-node) cliques.

Conditional Independence

In a MRF, conditional independence is simpler: if all paths between node a and b pass through c, then $a \perp\!\!\!\perp b \mid c$. For example, in the following graph, it is true that $A \perp\!\!\!\perp B \mid C$.



Directed $\rightarrow$ undirected

For every node x

- Add links between all parent nodes
- Drop arrows
- Set

$$\psi_{x,\mathrm{pa}_x}(x, \mathrm{pa}_x) = p(x|\mathrm{pa}_x) \prod_{y \in \mathrm{pa}_x} p(y)$$

Note that this process loses information about independence: in the directed graph, the parents of x were independent of each other and, in the undirected graph, they are not.

Thoughts on undirected graphs

It seems (to me) that – because of the “fuzziness” of the potential functions – there are two ways of getting an undirected graph:

(i) generate it from a directed graph.

(ii) generate it from a model that has a natural graphical representation (e.g. the image de-noising).

In contrast, directed graphs have a very precise probabilitic interpretation and can always be generated from models.

Using independence

Let x, y and z be discrete random variables that take on five values.

- In general, $p(x, y, z)$ is a $5 \times 5 \times 5$ table of values
- If x, y and z are independent, $p(x, y, z) = p(x)p(y)p(z)$; the joint is determined by the marginals

This basic principle will be used to speed up calculations of marginals and conditional marginals (like posteriors).

Chains

A chain is a graph of the form

and its joint distribution is given by

$$p(\mathbf{x}) = \psi_{1,2}(x_1, x_2) \cdots \psi_{N-1,N}(x_{N-1}, x_N),$$

where $\psi_{1,2}(x_1, x_2) = p(x_1)p(x_2|x_1)$ and

$$\psi_{k-1,k}(x_{k-1}, x_k) = p(x_k|x_{k-1}), \quad \text{for } k = 3, \ldots, N.$$

Inference on chains

Given a chain of discrete random variables (with K possible values for each), compute $p(x_n)$, given by

$$p(x_n) = \sum_{x_1} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_N} p(\mathbf{x})$$
$$= \sum_{x_1} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_N} \psi(x_1, x_2) \cdots \psi(x_{N-1}, x_N).$$

Each x_i can take on K possible values; sum over K^{N-1} values!

We can take advantage of independence properties ($x_{k+1} \perp\!\!\!\perp x_{k-1} | x_k$) to make this computation more efficient.

Inference on chains

Consider the following:

$$
\begin{aligned}
p(x_n) &= \sum_{x_1} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_N} p(\mathbf{x}) \\
&= \sum_{x_1} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_{N-1}} \Big[\psi(x_1, x_2) \cdots \psi(x_{N-1}, x_N^1) \\
&\qquad + \psi(x_1, x_2) \cdots \psi(x_{N-1}, x_N^2) \\
&\qquad \vdots \\
&\qquad + \psi(x_1, x_2) \cdots \psi(x_{N-1}, x_N^K) \Big] \\
&= \sum_{x_1} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_{N-1}} \Big[\psi(x_1, x_2) \cdots \psi(x_{N-2}, x_{N-1}) \sum_{x_N} \psi(x_{N-1}, x_N) \Big]
\end{aligned}
$$

Inference on chains

Applying this “trick” repeatedly, and realizing that we can separate into two parts, we get that

$$\begin{aligned} p(x_n) = &\left[\sum_{x_{n-1}} \psi(x_{n-1}, x_n) \cdots \left[\sum_{x_1} \psi(x_1, x_2)\right] \cdots \right] \\ &\left[\sum_{x_{n+1}} \psi(x_n, x_{n+1}) \cdots \left[\sum_{x_N} \psi(x_{N-1}, x_N)\right] \cdots \right]. \end{aligned}$$

Computation reduced to $O(NK^2)$.

Factor Graphs

Express joint density using “factors”:

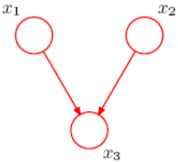
$$p(\mathbf{x}) = \prod_s f_s(\mathbf{x}_s)$$

Factor graphs:

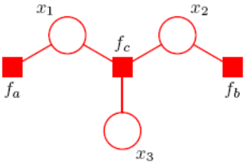
- Square nodes for “factors”
- Links between factor nodes and variable nodes used in factor
- Can convert between directed/undirected graphs to factor graphs

Example

Consider $p(\mathbf{x}) = p(x_1)p(x_2)p(x_3|x_1, x_2)$, with directed graph



Let $f_a(x_1) = p(x_1)$, $f_b(x_2) = p(x_2)$, $f_c(x_1, x_2, x_3) = p(x_3|x_1, x_2)$; the corresponding factor graph is



Summary

- Undirected graphs give a lot of flexibility; but not always applicable.
- Conditional independence is important! Taking advantage of it speeds up computation.
- Discuss!!