Moralization

More generally this conversion requires "marrying the parents". In this chain example the moral graph is complete.

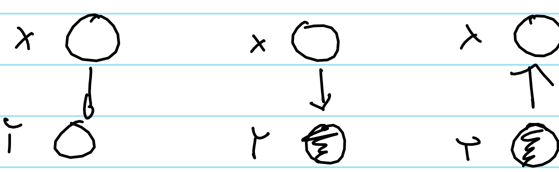
Recipe: directed G.M → undirected G.M.

1) Add links between all pairs of parents for all nodes in graph
2) Drop arrows
3) Initial clique potentials to 1
   a) Multiply in all conditional dist's associated with each clique.
4) $Z = 1$

* Inference in Graphical Models * ← KEY

Idea: exploit graphical structure in algorithms for inference.

First graphical Bayes Theorem



Joint $p(x, Y) = p(x) \cdot p(Y \mid x)$

If we obs. Y then $p(x)$ can be seen as a prior on x, and inferring the post. dist. of x can be the goal. To do this note

$$p(Y) = \sum_{x'} p(x', Y) = \sum_{x'} p(x')\, p(Y \mid x')$$

and $p(x \mid Y) = \dfrac{p(Y \mid x)\, p(x)}{p(Y)}$, reversing the arrow

Inference on a chain

- Develop intuition
- Derive algorithms for later inference techniques

G.M.:

$$x_1 \longrightarrow x_2 \longrightarrow x_3 \longrightarrow x_4 \longrightarrow x_5 \;\cdots\; x_{n-1} \longrightarrow x_n$$

- Directed and undirected versions of graph expressions are the same in terms of conditional independences

i.e.

$$p(\vec{x}) = \frac{1}{Z}\, \psi_{1,2}(x_1, x_2)\, \psi_{2,3}(x_2, x_3) \cdots \psi_{N-1}(x_{N-1}, x_N)$$

Case: Discrete K state variables (N of them)

- examples - prices on gambling exchange over time
  - number of people or packets in a queue,
- Each potential function has $K \times K$ vars (nat.)
- Joint dist has $(N-1)K^2$ params.

Problem: find marginal distribution $p(x_n)$

- e.g. given no obs's, how many people will be standing in line at time n

- Required calculation

$$p(x_n) = \sum_{x_1} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_N} p(\vec{x})$$

Cost? $O(K^N)$ - exponential in N!

Key Idea : Exploit Conditional Independence

this is why we have focused on cond. indep.

$$p(x_n) = \sum_{x_1} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_N} \frac{1}{Z} \psi_{1,2}(x_1, x_2)\, \psi_{2,3}(x_2, x_3) \cdots \psi_{N-1,N}(x_{N-1}, x_N)$$

but, note, summation over x_N can move this sum

$$= \sum_{x_1} \cdots \sum_{x_{n-1}} \psi_{1,2}(x_1, x_2)\, \psi_{2,3}(x_2, x_3) \cdots \psi_{x_{n-1}, x_n}(x_{n-1}, x_n)$$

$$\times \sum_{x_{n+1}} \psi(x_n, x_{n+1}) \left[\sum_{x_{n+2}} \psi(x_{n+1}, x_{n+2}) \left[\cdots \sum \psi(x_{N-2}, x_{N-1}) \left[\sum \psi(x_{N-1}, x_N) \right] \right] \right]$$

ugly but clear

can do this first yielding some function (in discrete case a vector) of x_{N-1} only.

Front half can be done this way too

$$\left[\sum_{x_{n-1}} \cdot \right] = \underbrace{\left[\sum_{x_2} \left[\sum_{x_1} \psi_{12}(x_1, x_2) \right] \psi_{23}(x_2, x_3) \right] \cdots \psi(x_{n-1}, x_n)}_{\mu_\alpha(x_n)} \times$$

$$\underbrace{\left[\sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1}) \cdots \left[\sum_{x_N} \psi_{N-1,N}(x_{N-1}, x_N) \right] \cdots \right]}_{\mu_\beta(x_n)}$$

This is important!

Computational trick $\overset{3 \text{ ops}}{ab + ac} = \overbrace{a(b+c)}^{2 \text{ ops}}$

Computational Cost of Shortcut

- $N-1$ summations over a $K \times K$ table
- $N-1$ multiplies of a K vector into a $K \times K$ table

overall $O(NK^2) \ll O(K^N)$

thank you conditional independence!

Messages

Can write

$$p(x_n) = \frac{1}{Z} \mu_\alpha(x_n) \mu_\beta(x_n)$$

Think of $\mu_\alpha(x_n)$ as message from x_{n-1} to x_n
and $\mu_\beta(x_n)$ " " " x_{n+1} to x_n

- Each message is a K-dim vector
- Product is pointwise
- Note Z can be determined by "inspection"

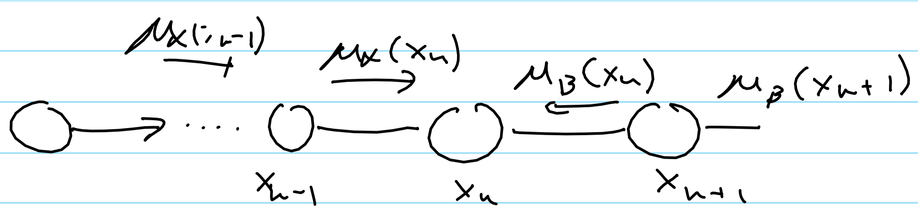
Recursive Computation

Note

$$\mu_\alpha(x_n) = \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \underbrace{\left[\sum_{x_{n-2}} \cdots \right]}_{\mu_\alpha(x_{n-1})}$$

$$= \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n)\, \mu_\alpha(x_{n-1})$$

e.g.



PRINTABLE NOTEPAD 2

Same holds for $\mu_\beta(x_n)$

$$\mu_\beta(x_n) = \sum_{x_{n+1}} \psi_{n+1,n}(x_{n+1}, x_n) \underbrace{\left[\sum_{x_{n+2}} \cdots \right]}_{\updownarrow}$$

$$= \sum_{x_{n+1}} \psi_{n+1,n}(x_{n+1}, x_n) \overbrace{\mu_\beta(x_{n+1})}$$

- All Marginals $p(x_1) \cdots p(x_n) \cdots p(x_N)$ at once?

  Naive approach - repeat message passing N times costs $O(N^2K^2)$

  - Why repeat computation?
    - to compute $p(x_n)$ and $p(x_3)$ we need $\mu_\alpha(x_2)$

  - Approach
    - compute all messages in both directions
    - twice as expensive is all ... !

- Observing variables

  - Introduce indicator functions $I(x_n, \hat{x}_n)$ for all observed $\hat{x}_n$
  - Note these indicators can be absorbed into clique potential functions yielding a 1 in only 1 entry
  - Messages can be passed as usual

- Learning potential function parameters
  - left till later