Wednesday, November 12, 2014 3:04 PM

Undirected Graphical Models
Inference in Graphical Models

Admin:

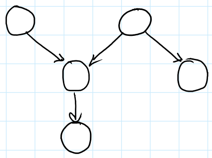
A5: Marked version due
A6: Due now
CP: out tonight (pick your model at 8pm) (due ~Dec 5th)
Auditors Too
Midterm: Monday (2 pgs front + back of notes)
A7: out monday (Remember: only top 6 assns count)

Last time:

DAG Models

$$p(x) = \prod_{j=1}^{d} p(x_j \mid x_{\pi(j)})$$

with notation $x_j = (x_i)_j$



Undirected Graphical Models

"Markov Random Fields"
"Markov Networks"

Divide $\{x_1, x_2, \cdots x_d\}$ into (possibly overlapping) subsets 'c'.

ex: $[1,2]$, $[1,5,3]$, $[5,4]$, $\cdots$

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

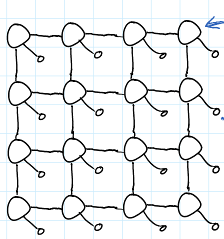
$$z = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_d} \prod_{c \in C} \phi_c(x_c)$$

$\phi_c(x_c) \geq 0$

ex: $\phi_c(x_1, x_2, x_3) \rightarrow \mathbb{R}^+$ real positive

Graph:
V: variables x_j
E: edge j-k if x_j and x_k appear in some subset c

Example: image segmentation



which are part of face?
- strong spacial correlation

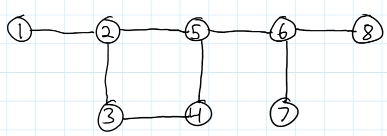
"Local Markov Property"

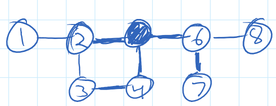
$$x_j \perp x_{(1:j) \setminus nei(j)} \mid x_{nei(j)}$$

(neighbours in graph)
- only dependent on neighbours

- determine if independence follows from using seperation

"" vs  dependent





$p(x) =$ 12, 23, 34, 25, 45, 56, 67, 68

$x_1 \not\perp x_2 \mid x_5$ dependent

$x_1 \perp x_6 \mid x_5$ independent

Pairwise UGMs

$$p(x) = \frac{\prod_{j=1}^{d} \left[\psi_j(x_j) \prod_{j,k \in E} \psi_{jk}(x_j, x_k) \right]}{Z}$$

Gaussian Graphical Models

$X \sim N(\mu, \Sigma)$

$\Sigma_{jk} = 0$ marginal independence

$\Downarrow$

x_j is not reachable from x_k (can't find any path)

$(\Sigma^{-1})_{jk} = 0$

$\Downarrow$

edge x_j to x_k is missing

Ising Models

$x_j = \{-1, 1\}$

$\psi_j(x_j) = \exp(-w_j x_j)$

$\psi_{jk}(x_j, x_k) = \exp(-w_{jk} x_j x_k)$

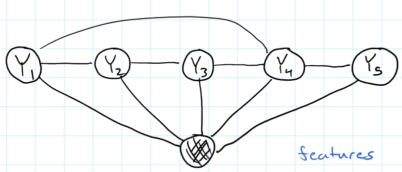
$$p(x) = \frac{\exp\left(-\sum_j \left[w_j x_j - \sum_{j,k \in E} w_{jk} x_j x_k \right]\right)}{Z}$$

If $x_j \in \{1, 2, \cdots S\}$,

"Potts model"

Special case of "log-linear model"

Conditional Random Field



features

train: $\{X, y\}$
test: $\hat{x} \to \hat{y}$

Logistic is a special case.

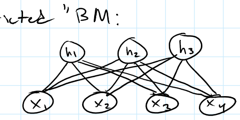
$\psi_1(x_1) = \exp(-w^T x)$

generalization

Boltzmann Machine

Ising model w/ hidden variables

"Restricted" BM:



Computation
(discrete)

compute: $p(x) \longrightarrow$ easy if you know Z,
else,
#P - hard   dag was easy

$p(x_j) \longrightarrow$ #P - hard

$p(x_j \mid x_k) \longrightarrow$ #P - hard

$\max_x p(x) \longrightarrow$ NP hard   same as dag

We can solve these efficiently if 'G' is nice.

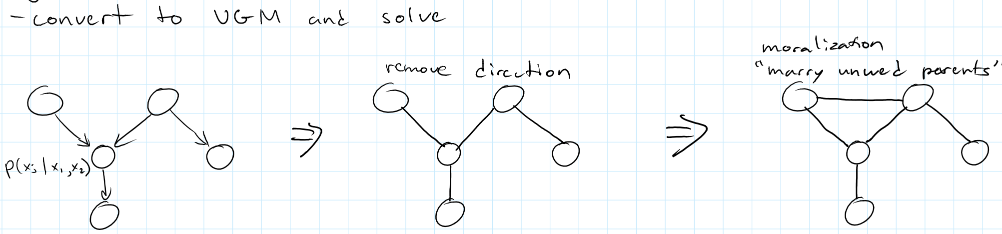
DAGs vs UGMs

Easy ops: $p(x)$

$p(x_i \mid x_{1:j})$

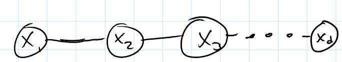
parameter esimation

everything else is as hard as UGM
- convert to UGM and solve



Viterbi Decoding

Consider chain-structured UGM



$$p(x) = \frac{\psi_1(x_1) \prod_{j=1}^{d-1} \psi_j(x_j, x_{j+1})}{Z}$$

"Decoding" $\max_x p(x)$

"most likely sequence"

If $x_j \in \{1, 2, \cdots S\}$, there are S^d configurations

but, can be solved in $O(dS^2)$ by "dynamic programming"

Dynamic Programing:

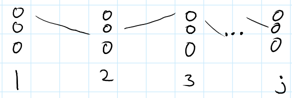
Strategy for solving discrete optimization, where solution is defined recursively.

Ex:

Let V_{jk} be the potential of the most probable sequence based on

$$\psi_1(x_1),\ \psi_1(x_1, x_2),\ \psi_2(x_2, x_3)\ \cdots\ \psi_{j-1}(x_{j-1}, x_j)$$

that ends with $x_j = k$.

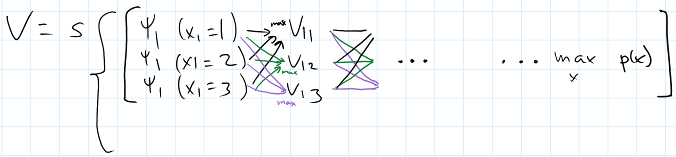


Optimal solution: $\max_k V_{dk}$

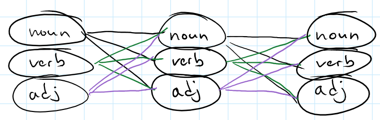
$$V_{1k} = \psi_1(x_1 = k)$$

$$V_{jk} = \max_{k'} \left\{ \psi_{j-1}(x_{j-1} = k', x_j = k)\ V_{(j-1)k'} \right\}$$

"max product" belief propagation



Ex: the ball is red



- backtrack through the argmax values to get $\underset{x}{\operatorname{argmax}}\ p(x)$

Forward - Backward

Forward

we can compute Z with a similar algorithm

$$V_{1k} = \psi_1(x_1 = k)$$

$$V_{jk} = \sum_{k'} \psi_{j-1}(x_{j-1} = k', x_j = k)\ V_{(j-1)k'}$$

its the same, but instead of max, you sum.

"sum product" belief propagation

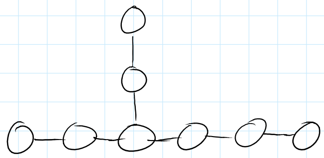
Backward

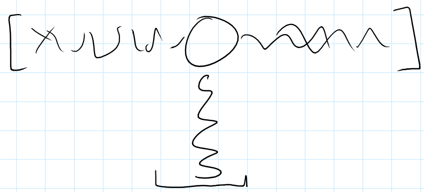
$$B_{dk} = 1$$

$$B_{jk} = \sum_{k'} \psi_j(k', k)\ B_{(j+1)k'}$$

$$p(x_j = k) \propto V_{jk} B_{jk}$$

What if we don't have a chain?





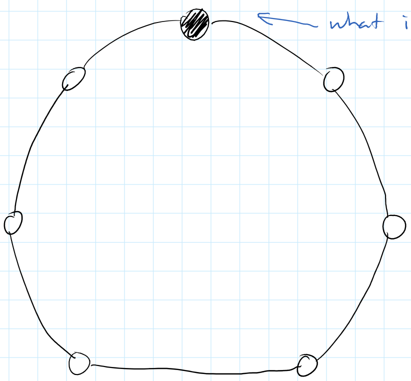
combining two tables

"belief propagation"

$O(ds^2)$

any tree structure
- no cycles!

What if we have cycles?



what if we observe this?

$\Rightarrow$