Graphical Models

Ron Parr
CPS 271

Review of Bayes Nets

- BNs represent a joint distribution
- Exploit conditional independence

- Potential for exponential reduction in storage
- Potential for exponential reduction in computation for computing (e.g.) marginals

- Worst case is exponential, but we hope that doesn't happen too often

Bayes Net Construction

- Can construct a BN by adding one variable at a time, connecting each variable to parents upon which it depends
- Exponentially many ways to do this
- Some may be much more compact than others

- Every Bayes net implies a partial ordering on the variabes

"Generative Models"

- A Bayes net is an example of a generative model of a probability distribution
- Generative models allow one to generate samples from a distribution in a natural way
- Sampling algorithm:
  - While some variables are not sampled
    - Pick variable x with no unsampled parents
    - Assign this variable a value from p(x|parents(x))
- Our first approximate inference algorithm for Bayes nets!

Additional Comments on Sampling

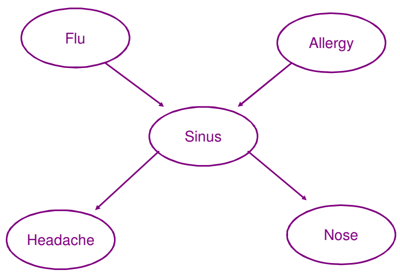
- Sampling is the easiest algorithm to implement
- Can compute marginal or conditional distributions by counting

- Problem: How do we handle observed values?

Continuous Variables

- So far, we have variables with discrete domains
- What if our variables are continuous, e.g., jointly Gaussian?
- We can still construct a Bayesian network in the same way, but what are the benefits?
  - Exponential reduction in storage?
  - Exponential reduction in time?
  - Saliency?

Recall our Sinus Model



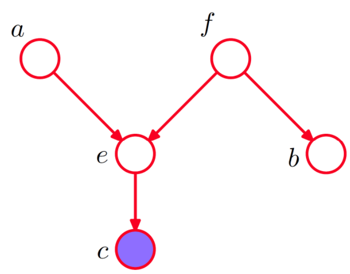
Determining Dependencies

- Recall that BN variables are conditionally independent of their non-descendents given their parents
  - Flu and Allergy are independent
  - Runny nose and headache are independent given sinus inflammation
- Are flu and allergy independent given sinus inflammation?

D-Separation

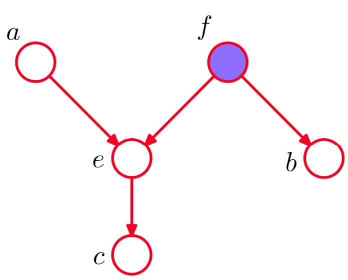
- For a given BN structure, how do we determine if some variable A is conditionally independent of B given C??? (defined for sets, not just individual variables)
- Some terminology: A path through the network is *blocked* if it includes a node where
  - The arrows meet head-to-tail or tail-to-tail and the node is in C, **or**
  - The arrows meet head to head, and neither the node nor its descendents are in C
- If all paths between A and B are blocked then A is conditionally independent of B given C.
- Note symmetry between A and B

D-Separation Examples



Are a and f conditionally independent given c???

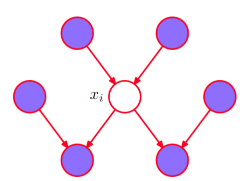
D-Separation Examples



Are a and b conditionally independent given f???

Markov Blanket

- Q: What variables completely isolate one variable from the rest of the network?
- A: Its parents, its children, and its spouses (Yes, BN variables are polygamous)

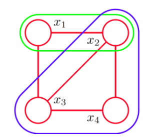


Undirected Models

- BNs are great, but…
  - Determining conditional independence is tricky
  - Markov blanket is a bit unintuitive
- Markov Random Fields (MRFs), also known as Markov Networks
  - Have all undirected arcs
  - Have more salient dependency relationships
  - Are (in some ways) easier to specify

MRF Specifications

- An MRF is a graph w/one node per random variable
- An MRF defines a function called a potential f(C) for each maximal clique in the graph
- What's a maximal clique:
  - A clique is a fully connected set of nodes
  - A maximal clique is not part of any larger clique



Probabilities for MRFs

$$p(\mathbf{x}) = \frac{1}{Z}\prod_C \varphi_C(\mathbf{x}_C)$$

This is very compact notation; parse it carefully!

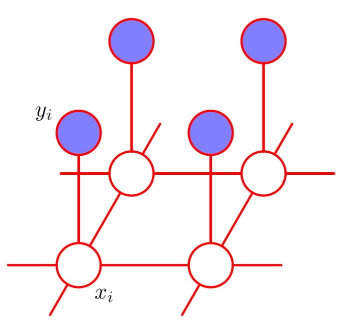
$$z = \sum_{\mathbf{x}}\prod_C \varphi_C(\mathbf{x}_C)$$

- **x** is an assignment to all RVs in the network
- $\mathbf{x}_C$ is the subset of the variables in clique C
- z is is normalizing constant
- Good news: Potentials don't need to be distributions
- Bad news: Hard to compute z

Image De-Noising Example

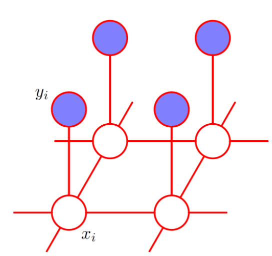
- Suppose we observe a 2-color digital image with pixels values y_i that are corrupted by some noise
- Suppose that the true image has pixel values x_i
- Assumptions:
  - Y_i and x_i should be strongly correlated
  - x_i and x_j should be strongly correlated if i and j are neighbors

MRF Structure



Note: Simple structure means that cliques are defined just over pairs

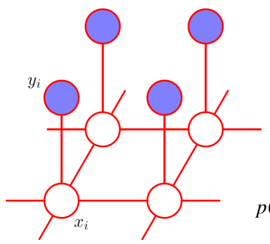
Potentials



$$\varphi(x_i) = e^{hx_i}$$
$$\varphi(x_i, x_j) = e^{-\beta x_i x_j}$$
$$\varphi(x_i, y_i) = e^{-\eta x_i y_i}$$

Note: x_i is not a maximal clique, but this is just a mater of convenience since there is an equivalent formulation in terms of maximal cliques.

Probability of an Assignment



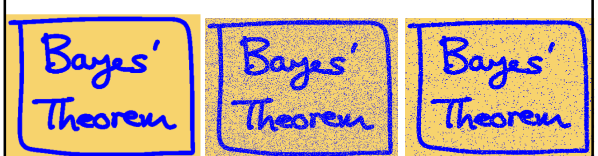
$$p(x,y)=\frac{1}{z}\prod_i e^{hx_i}\prod_{i,j}e^{-\beta x_i x_j}\prod_i e^{-\eta x_i y_i}$$

How do we deal with this?

Iterated Conditional Modes

- Initialize y=x
- Repeat until convergence (or few changes)
  - Pick an x
  - Freeze all other variables
  - Update x in the direction that increases p(x,y)
- Coordinate ascent to find argmax_x p(x,y)
- Note that we ignore z

ICM Example



Original Corrupted De-noised

Inference in Graphical Models

- Variable elimination applies to Markov nets as well as Bayes nets
- Can also find the highest probability assignment (e.g. denoising)
- Basic idea: Replace sum with max
- Yes, it's really that easy…
- For trees, if we use log probabilities, this has a natural interpretation as dynamic programming to find the lowest cost path

Graphical Model Conclusions

- Bayes nets are an instance of the general class that we call graphical models
- Graphical models may be:
  - Directed or undirected
  - Discrete or continuous
- Cost of storage and inference depends upon
  - Structure of graphical model
  - Form of potentials/CPTs