Bayesian Networks – (Structure) Learning

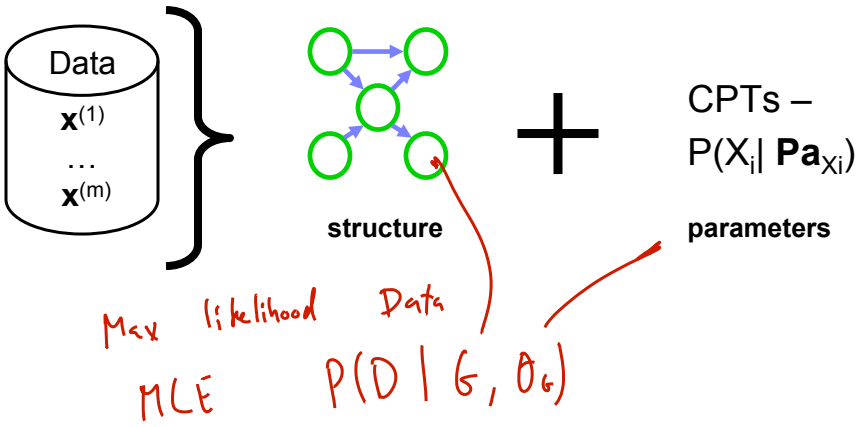
Machine Learning – CSE546

Carlos Guestrin

University of Washington

November 20, 2014

Learning Bayes nets



Chow-Liu tree learning algorithm 1

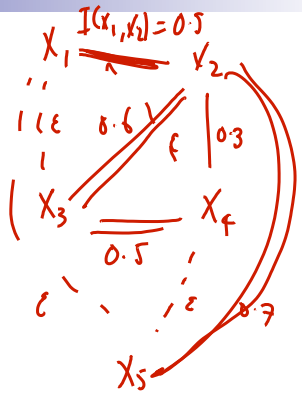
- For each pair of variables X_i,X_j
  - Compute empirical distribution:

$$\hat{P}(x_i, x_j) = \frac{\text{Count}(x_i, x_j)}{m}$$

  - Compute mutual information:

$$\hat{I}(X_i, X_j) = \sum_{x_i, x_j} \hat{P}(x_i, x_j) \log \frac{\hat{P}(x_i, x_j)}{\hat{P}(x_i)\hat{P}(x_j)}$$

- Define a graph
  - Nodes $X_1,\ldots,X_n$
  - Edge (i,j) gets weight $\hat{I}(X_i, X_j)$



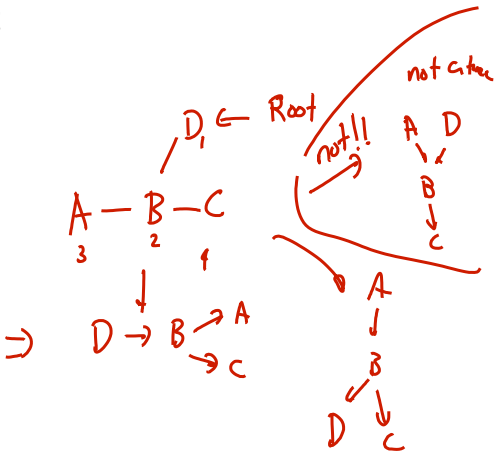
1) run max spanning tree ← complexity is about $O(E \log E)$

$O(n^2 \log n)$

Chow-Liu tree learning algorithm 2

$$\log \hat{P}(\mathcal{D} \mid \theta, \mathcal{G}) = m \sum_i \hat{I}(X_i, \mathbf{Pa}_{X_i, \mathcal{G}}) - m \sum_i \hat{H}(X_i)$$

- Optimal tree BN
  - Compute maximum weight spanning tree
  - Directions in BN: pick any node as root, breadth-first-search defines directions



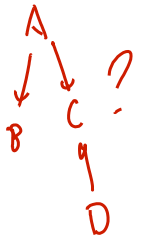
Structure learning for general graphs

- In a tree, a node only has one parent
- **Theorem**:
  - The problem of learning a BN structure with at most *d* parents is NP-hard for any (fixed) *d*>1
- Most structure learning approaches use heuristics
  - (Quickly) Describe the two simplest heuristic

Learn BN structure using local search

Starting from Chow-Liu tree

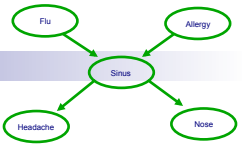
Local search, possible moves:
- Add edge
- Delete edge
- Invert edge



Score using BIC

penalizes for dense graphs

Learn Graphical Model Structure using LASSO



- Graph structure is about selecting parents:
  $P(X_i | Pa_{X_i,G})$ ← "like logistic regression"
  $Pa_{X_i,G} \in \{X_1 \dots X_{i-1}, X_{i+1}, \dots, X_n\}$
- If no independence assumptions, then CPTs depend on all parents:
  P(H|FASN)
- With independence assumptions, depend on key variables:
  equal to P(H|S) ⇒ like coeffs of other vars are ∅
- One approach for structure learning, sparse logistic regression!
  Logistic regression with L1 penalty to estimate parents of each X_i

Serious caveat: approach will create cycles | BN assumes graph has no directed cycles ⇒ acyclicity of LR | edges in graph

What you need to know about learning BN structures

- Decomposable scores
  - Maximum likelihood
  - Information theoretic interpretation
- Best tree (Chow-Liu)
- Beyond tree-like models is NP-hard
- Use heuristics, such as:
  - Local search
  - LASSO