Learning Bayes nets

Data
\[ x^{(1)} \]
\[ \ldots \]
\[ x^{(m)} \]

\{ structure \} +

CPTs – \[ P(X_i \mid Pa_{xi}) \]

\[ \text{Max likelihood Data MLE} \]
\[ P(D \mid \Theta, \Theta_c) \]
Chow-Liu tree learning algorithm 1

- For each pair of variables $X_i, X_j$
  - Compute empirical distribution:
    $$P(x_i, x_j) = \frac{\text{Count}(x_i, x_j)}{m}$$
  - Compute mutual information:
    $$I(X_i, X_j) = \sum_{x_i,x_j} P(x_i, x_j) \log \frac{P(x_i, x_j)P(x_i)P(x_j)}{P(x_i, x_j)}$$
- Define a graph
  - Nodes $X_1, ..., X_n$
  - Edge $(i,j)$ gets weight $I(X_i, X_j)$

Run max spanning tree algorithm; complexity is $O(E \log n)$, $O(n^2\log n)$

Chow-Liu tree learning algorithm 2

- $\log \hat{P}(D | \theta, G) = m \sum_i I(X_i, \text{Pa}_{X_i, G}) - m \sum_i \hat{H}(X_i)$
- 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) \( d \)\(^*\)

- Most structure learning approaches use heuristics
  - (Quickly) Describe the two simplest heuristics

---

Learn BN structure using local search

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

- Score using BIC

> **penalties for dense graphs**
Learn Graphical Model Structure using LASSO

- Graph structure is about selecting parents:
  \[ P(x_i | P_{x_i}) \]
  - If no independence assumptions, then CPTs depend on all parents:
  \[ P(x_i | P_{x_i}) \subseteq \{ x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n \} \]
  - With independence assumptions, depend on key variables:
  \[ P(H|I|A|S|N) = \text{like roots of other vars are } \emptyset \]
  - One approach for structure learning, sparse logistic regression!
    Logistic regression with L1 penalty to eschew parents of each \( x_i \)

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