Bayesian Networks: Quick Review

- Recall: Conditional Independence (CI) Assumptions
- Bayesian Networks: Digraph Model
  - Vertices (nodes): denote events (each a random variable)
  - Edges (arcs, links): denote conditional dependencies
- Chain Rule for (Exact) Inference in BBNs
  - Arbitary Bayesian networks: bipartite
  - Polytrees: linear time
- Example (“Sprinkler” BBN)

Learning Distributions in BBNs: Quick Review

- Learning Distributions
  - Shortcomings of Naive Bayes
  - Making judicious CI assumptions
  - Scaling up to BBNs: need to learn a CPT for all parent sets
  - Goal: generalization
  - Given D (e.g., (101, 001, 0100))
    - Would like to know P(schema); e.g., P(11*) = P(x1 = 1, x2 = 1)
- Variants
  - Known or unknown structure
  - Training examples may have missing values
- Gradient Learning Algorithm
  - Weight update rule
  \[ W_p = W_p - \frac{\partial}{\partial W_p} \log P(D) \]
  - Learns CPTs given data points D

Learning Structure: Constraints Versus Scores

- Constraint-Based
  - Perform tests of conditional independence
  - Search for network consistent with observed dependencies (or lack thereof)
  - Intuitive; closely follows definition of BBNs
  - Separates construction from form of CI tests
  - Sensitive to errors in individual tests
- Score-Based
  - Define scoring function (aka score) that evaluates how well (in)dependencies in a structure match observations
  - Search for structure that maximizes score
  - Statistically and information theoretically motivated
  - Can make compromises
- Common Properties
  - Soundness: with sufficient data and computation, both learn correct structure
  - Both learn structure from observations and can incorporate knowledge

After Spring Break

Next Class: Last BBN Presentation (Yue Jiao: Causality)
Learning Structure: Maximum Weight Spanning Tree (Chow-Liu)

- Algorithm Learn-Tree-Structure-I (D)
  - Estimate P(x) and P(x|y) for all singly RVs; pairs: I(X; Y) = (P(Y|X, Y) - P(Y))
  - Build complete undirected graph: variables as vertices, I(X; Y) as edge weights
  - Set direction on T and place the CPTs on its edges (gradient descent)
  - RETURN: tree-structured BBNs with CPT values

- Algorithm Build-MVST-Kruskal (E * V → V, Weights: E → R+)
  - H: Build-Heap (E, Weights) // aka priority queue, O(E)
  - E → D; Forest ← { (v, v) | v ∈ V }, I( X ; Y ) as edge weights
  - WHILE Forest.Size × 1 DO
    - e = H.Delete-Min() // e ∈ new edge from H, Olg(E)
      - IF (T, E) → Forest.Find(e.Start) → (T*, E: Forest.Find(e.End)) THEN
        - Olg(E)
      - E' = Union(e), Forest.Size−(1)
    - RETURN E'
  - Running Time: O(E lg |E|) = O(\sqrt{|V|^3 lg |V|}) = O(C^2 lg C)

Scores for Learning Structure: The Role of Inference

- General-Case BBN Structure Learning: Use Inference to Compute Scores
- Recall: Bayesian Inference aka Bayesian Reasoning
  - Assumption: H → H are mutually exclusive and exhaustive
  - Optimal strategy: coreline predictions of hypotheses h given observed data D
  - Compute conditional probability of hypothesis h given observed data D
  - i.e., compute expectation over unknown H for unseen cases
  - Let h: structure, parameters \theta: CPTs

P(X_{1:n}|D) = P(X_{1:k}|D) / P(X_{k+1:n}|D)

\sum_{h \in H} P(h) P(D|h) P(h)

Prior over Structures

Likelihood

Prior over Parameters

In-Class Exercise: Hugin Demo

- Hugin
  - Commercial product for BBN inference: http://www.hugin.com
  - First developed at University of Aalborg, Denmark
  - Commercial product for BBN inference: http://www.hugin.com
  - Customer support automation: http://www.cs.auc.dk/research/DSS/SACSO/
  - Diagnosis and control in unmanned subs: http://advocate.e-motive.com
  - Safety and risk evaluation: http://www.hugin.com/serene/
  - K2 found BBN different in only 1 edge from gold standard (selected from expert)

Learning Structure: K2 Algorithm and ALARM

- Algorithm Learn-BBN-Structure-K2 (D, Max-Parents)
  - FOR i ← 1 to n DO
    - arbitrary ordering of variables [x_1, x_2, …, x_n]
  - WHILE (Parent(x_i), Size ≤ Max-Parents): DO
    - find best candidate parent
    - Best = argmax_k P(D|X_i = Parent(x_i))
    - max Dirichlet score
    - IF (Parent(x_i) = Best): Score THEN Parent(x_i) = Best
    - RETURN (Parent(x_i) = 1, 2, …, n)
  - A Logical Alarm Reduction Mechanism [Beinlich et al., 1989]
    - BBN model for patient monitoring in surgical anesthesia
    - Vertices (37): findings (e.g., esophageal intubation), intermediates, observables
    - K2 found BBN different in only 1 edge from gold standard (selected from expert)

Learning Structure: State Space Search and Causal Discovery

- Learning Structure: Beyond Trees
  - Problem not as easy for more complex networks
    - Example: allow two parents (even singly-connected case, aka polytree)
      - Greedy algorithms no longer guaranteed to find optimal network
      - In fact, no efficient algorithm exists
    - Theorem: finding network structure with maximal score, where H restricted to BBNs with at most k parents for each variable, is NP-hard for k > 1
  - Heuristic (Score-Based) Search of Hypothesis Space H
    - Define: H elements denote possible structures, adjacency relation denotes transformation (e.g., arc addition, deletion, reversal)
    - Traverse this space looking for high-scoring structures
    - Algorithms: greedy hill-climbing, best-first search, simulated annealing
  - Causal Discovery: Inferring Existence, Direction of Causal Relationships
    - Want: “no unexplained correlations; no accidental independencies” (cause ≠ 0)
    - Can discover causality from observational data alone?
    - What is causality anyway?
CIS 830: Advanced Topics in Artificial Intelligence

In-Class Exercise: Hugin and CPT Elicitation

- Hugin Tutorials
  - Introduction: causal reasoning for diagnosis in decision support (toy problem)
  - Example domain: explaining low yield (drought versus disease)
  - Tutorial 1: constructing a simple BBN in Hugin
  - Eliciting utilities (or collecting from data) and entering them
  - Tutorial 2: constructing a simple influence diagram (decision network) in Hugin
  - Eliciting utilities (or collecting from data) and entering them

- Other Important BBN Resources
  - Microsoft Bayesian Networks: http://research.microsoft.com/dtas/mbn
  - BML BN (Interchange Format): http://research.microsoft.com/dtas/bnformat/
  - BBN Repository (more data sets)

CIS 830: Advanced Topics in Artificial Intelligence

Learning Structure: Conclusions

- Key Issues
  - Finding a criterion for inclusion or exclusion of an edge in the BBN
  - Each edge
    - “Slice” (axis) of a CPT or a commitment to acquire one
  - Positive statement of conditional dependency

- Other Techniques
  - Focus today: constructing (score-based) view of BBN structure learning
  - Other score-based algorithms
  - Heuristic search over space of addition, deletion, reversal operations
  - Other criteria (information theoretic, coding theoretic)
  - Constraint-based algorithms: incorporating knowledge into causal discovery

- Augmented Techniques
  - Model averaging: optimal Bayesian inference (integrate over structures)
  - Hybrid BBN/DT models: use a decision tree to record P(x | Parents(x))

- Other Structures: e.g., Belief Propagation with Cycles

CIS 830: Advanced Topics in Artificial Intelligence

In-Class Exercise: Bayesian Knowledge Discoverer (BKD) Demo

- Bayesian Knowledge Discoverer (BKD)
  - Research product for BBN structure learning: http://kmi.open.ac.uk/projects/bkd/
  - Bayesian Knowledge Discovery Project [Ramoni and Sebastiani, 1997]
  - Knowledge Media Institute (KMI), Open University, United Kingdom
  - Closed source, beta freely available for educational use

- Other Important BBN Resources
  - Sister Product: Robust Bayesian Classifier (RoC)
    - Research product for BBN classification with missing data
      - Uses Efficient Bayesian Estimator, a deterministic approximation algorithm

CIS 830: Advanced Topics in Artificial Intelligence

Continuing Research and Discussion Issues

- Advanced Topics (Suggested Projects)
  - Continuous variables and hybrid (discrete/continuous) BBNs
  - Induction of hidden variables
  - Local structure: localized constraints and assumptions, e.g., Noisy-OR BBNs
  - Online learning and incrementality (aka lifelong learning): ability to change network structure during inferential process
  - Hybrid quantitative and qualitative inference (“simulation”)

- Other Topics (Beyond Scope of CIS 830 / 864)
  - Structural EM
  - Polytree structure learning (tree decomposition): alternatives to Chow-Liu MWST
  - Complexity of learning, inference in restricted classes of BBNs
  - BBN structure learning tools: combining elicitation and learning from data

- Turn to A Partner Exercise
  - How might the Lumiere methodology be incorporated into a web search agent?
  - Discuss briefly (3 minutes)

CIS 830: Advanced Topics in Artificial Intelligence

Terminology

- Bayesian Networks: Quick Review on Learning, Inference
  - Structure learning: determining the best topology for a graphical model from data
  - Constraint-based methods
  - Score-based methods: statistical or information-theoretic degree of match
  - Both can be global or local, exact or approximate
  - Elicitation of subjective probabilities

- Causal Modeling
  - Causality: “direction” from cause to effect among events (observable or not)
  - Causal discovery: learning causality from observations

- Incomplete Data: Learning and Inference
  - Missing values: to be filled in given partial observations
  - Expectation-Maximization (EM): iterative refinement clustering algorithm
  - Estimation step: use current parameters to estimate missing (N)
  - Maximization (re-estimation) step: update parameters to maximize P(x | E_I)

CIS 830: Advanced Topics in Artificial Intelligence

Summary Points

- Bayesian Networks: Quick Review on Learning, Inference
  - Learning, eliciting, applying CPTs
  - In-class exercise: Hugin demo; CPT elicitation, application

- Learning BBN structure: constraint-based versus score-based approaches
  - K2, other scores and search algorithms

- Causal Modeling and Discovery: Learning Causality from Observations
  - Incomplete Data: Learning and Inference (Expectation-Maximization)

- Tutorials on Bayesian Networks
  - Friedman and Goldszmidt (AAAI ’98, Learning BBNs from Data):
  - Heckerman (various UAI/CAI/ICML, 1998-1999, Learning BBNs from Data)
    - http://www.research.microsoft.com/heckerman

- Next Class: BBNs and Causality
  - Later: UAI Concluded; KDD, Web Mining; GAs, Optimization

CIS 830: Advanced Topics in Artificial Intelligence