Recitation
25/10/2007
Hetunandan
Review

- Topics covered in class this week:
  - Structure Learning In BNs
  - Expectation Maximization

- Main topic for review today: Structure Learning
BN structure learning

- Two approaches
  - Independence tests
    - Given data, determine all independencies
    - Typically using a chi square/mutual information test
    - Can be very sensitive to noise.
    - Used when number of variables are few and lot of data
  - Score-based search
Score based Structure Learning

Idea:
- Associate score with each model
- Search over all models and find one with best score
  - Hard!
  - Do greedy search and stop when score doesn’t improve
How to score?

- Wishlist for scoring function
  - Consistency
    - Given enough data, best BN must have highest score
    - I-equivalent BNs should have same score
  - Decomposability
    - $\text{Score} = \sum_{\text{fam}} \text{Score(fam)}$
    - For computational convenience: greedy search can quickly compute scores of its moves
Scoring Functions

- Maximum Likelihood score
- Score(Graph, Data) = log-likelihood(G, $\theta_{ML}$), D)
  - $\theta_{ML}$ is the ML estimate of $\theta$ for this skeleton G
  - Can rewrite as sums of single variable entropy terms and mutual information over family
  - Therefore, decomposable
Bad News: ML score overfits

- To see this, consider two node graphical model over X,Y
- Score of model with one edge always $\geq$ score of model with no edges
  - Since mutual information always $\geq 0$
- True in general: $I(X,Y \cup Z) \geq I(X,Y)$
Bayesian Score

Score(G,D) = \log P(D|G) + \log P(G)

- log P(D|G) = \int_{\theta} P(D|\theta,G) P(\theta|G) \, d\theta
- log P(G) is from structure prior

We would like to fulfill wishlist
- Good News: Bayesian score is consistent!
- Need to work on making it decomposable
Decomposability: Sufficient Conditions

- Global Parameter Independence
- Parameter modularity
Encoding Priors

- **K2 prior**
  - Choose a dirichlet prior for each value of \( P(X|Pa(X)=u) \)
  - **Pro:** Simple
  - **Con:** Lose Score Equivalence (hw3!)
BDe prior

- Prior is encoded in the form of some observed num of instances($M'$) of all variables
- $\alpha_{X|\text{Pa}(X)} = M'P'(X, \text{Pa}(X))$
- Weight of prior on variable now independent of number of parents