10708 Graphical Models: Homework 3

Due October 27th, beginning of class

October 13, 2006

Instructions: There are six questions on this assignment. Each question has the name of one of the TAs beside it, to whom you should direct any inquiries regarding the question. The last problem involves coding, which should be done in MATLAB. Do *not* attach your code to the writeup. Instead, copy your implementation to

`/afs/andrew.cmu.edu/course/10/708/your_andrew_id/HW3`

Refer to the web page for policies regarding collaboration, due dates, and extensions.

Note: Please put your name and Andrew ID on the first page of your writeup.

1 Triangulation [10 pts] [Khalid]

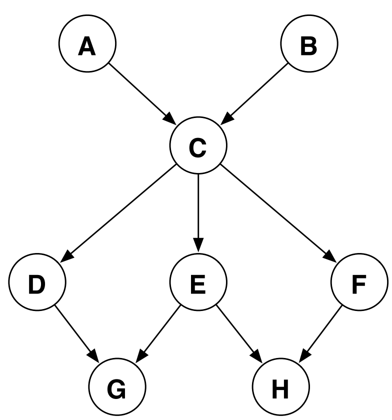


Figure 1: Bayes net for question 1

1. Moralize the Bayes net in figure 1.
2. Supply a perfect elimination ordering (*i.e.*, one that yields no fill edges).

3. Supply an elimination ordering that yields a triangulated graph with at least 5 nodes in one or more cliques
4. Draw clique trees for the elimination orderings in parts 2 and 3.

2 Clique Tree Factorization [10 pts] [Khalid]

2.1

Prove that the clique beliefs $\pi(\mathbf{C}_i) = P(\mathbf{C}_i)$ and edge beliefs $\mu_{ij}(\mathbf{S}_{ij}) = P(\mathbf{S}_{ij})$ form a fixed point for the belief propagation algorithm for a clique tree, i.e., if we start BP with these beliefs, no messages will change them.

2.2

Using the independencies we ask you to show in Question 4, prove that in a clique tree for a BN we can represent the joint distribution by:

$$P(\mathbf{X}) = \frac{\prod_i P(\mathbf{C}_i)}{\prod_{ij} P(\mathbf{S}_{ij})}.$$

You should not "prove" by corollary from the correctness of BP in clique trees.
(Hint: combine the chain rule of probabilities with the definition of conditional probabilities.)

3 Variable Elimination in Clique Trees [15 pts] [Khalid]

Consider a chain graphical model with the structure $X_1 - X_2 - \cdots - X_n$, where each X_i takes on one of d possible assignments. You can form the following clique tree for this GM: $\mathbf{C}_1 - \mathbf{C}_2 - \cdots - \mathbf{C}_{n-1}$, where $Scope[\mathbf{C}_i] = \{X_i, X_i + 1\}$. You can assume that this clique tree has already been calibrated. Using this clique tree, we can directly obtain $P(X_i, X_i + 1)$. As promised in class, your goal in this question is to compute $P(X_i, X_j)$, for any $j > i$.

3.1

Briefly, describe how variable elimination can be used to compute $P(X_i, X_j)$, for some $j > i$, in linear time, given the calibrated clique tree.

3.2

What is the running time of the algorithm in part 3.1 if you wanted to compute $P(X_i, X_j)$ for all n choose 2 choices of i and j?

3.3

Consider a particular chain $X_1 - X_2 - X_3 - X_4$. Show that by caching $P(X_1, X_3)$, you can compute $P(X_1, X_4)$ more efficiently than directly applying variable elimination as described in part 3.1.

3.4

Using the intuition in part 3.3, design a dynamic programming algorithm (caching partial results) which computes $P(X_i, X_j)$ for all n choose 2 choices of i and j in time asymptotically much lower than the complexity you described in part 3.2. What is the asymptotic running time of your algorithm?

4 Clique Tree I-maps [20 pts] [Ajit]

In order to formalize the relationship between clique trees and Bayesian Networks, in this question you will prove that if P factorizes according to a Bayesian Network, then any clique tree $\mathcal{T}$ for this BN is an I-map for P. Specifically;
In a clique tree, consider a separator $\mathbf{S}_{ij}$ between two cliques $\mathbf{C}_i$ and $\mathbf{C}_j$. Let $\mathbf{X}$ be any set of variables in the $\mathbf{C}_i$ side of the tree, and $\mathbf{Y}$ be any set of variables in the $\mathbf{C}_j$ side of the tree. Prove that $P \models (\mathbf{X} \perp \mathbf{Y} \mid \mathbf{S}_{ij})$.

5 Clique Sizes in Clique Trees [20 pts] [Khalid]

Assume that we have constructed a clique tree $\mathcal{T}$ for a given Bayesian network graph $\mathcal{G}$, and that each of the cliques in $\mathcal{T}$ contains at most k nodes. Now, the user decides to add a single edge to the Bayesian network, resulting in a network $\mathcal{G}'$. (The edge can be added between any pair of nodes in the network, so long as it maintains acyclicity.) What is the tightest bound you can provide on the maximum clique size in a clique tree $\mathcal{T}'$ for $\mathcal{G}'$? Justify your response by explaining how to construct such a clique tree.
(Note: You do not need to provide the optimal clique tree $\mathcal{T}'$. The question asks for the tightest clique tree that you can construct, using only the fact that $\mathcal{T}$ is a clique tree for $\mathcal{G}$.)

6 Variable Elimination [25 pts] [Ajit]

6.1 [15 pts]

Implement the variable elimination algorithm from class. You shall not implement pruning for inactive variables. You can reuse any code you wrote for hw1 or hw2, and you are free to use the posted solution code for hw1.

Submit your implementation to your AFS code directory. Answer the following questions in your writeup, reporting all probabilities to four significant digits.

1 Using the network in figure 2 and parameters from the first homework[1] answer the following questions using your variable elimination code:

(a) What is the ordering produced using the min-fill heuristic ?

(b) Compute $P(r_4)$. Note that this is a marginal distribution, not just a marginal probability.

(c) Compute $P(r_4 | r_1 = \text{Dislikes}, r_2 = \text{Likes}, r_3 = \text{Likes})$. Note that this is a posterior marginal distribution, not just a single posterior marginal probability.

2 Using the Alarm network in `alarm.net` compute the value of the following queries.

(a) P(StrokeVolume = High | Hypovolemia = True, ErrCauter = True, PVSat = Normal, Disconnect = True, MinVolSet = Low)

(b) P(HRBP = Normal | LVEDVolume = Normal, Anaphylaxis = False, Press = Zero, VentTube = Zero, BP = High)

(c) P(LVFailure = False | Hypovolemia = True, MinVolSet = Low, VentLung = Normal, BP = Normal)

6.2 [10 pts]

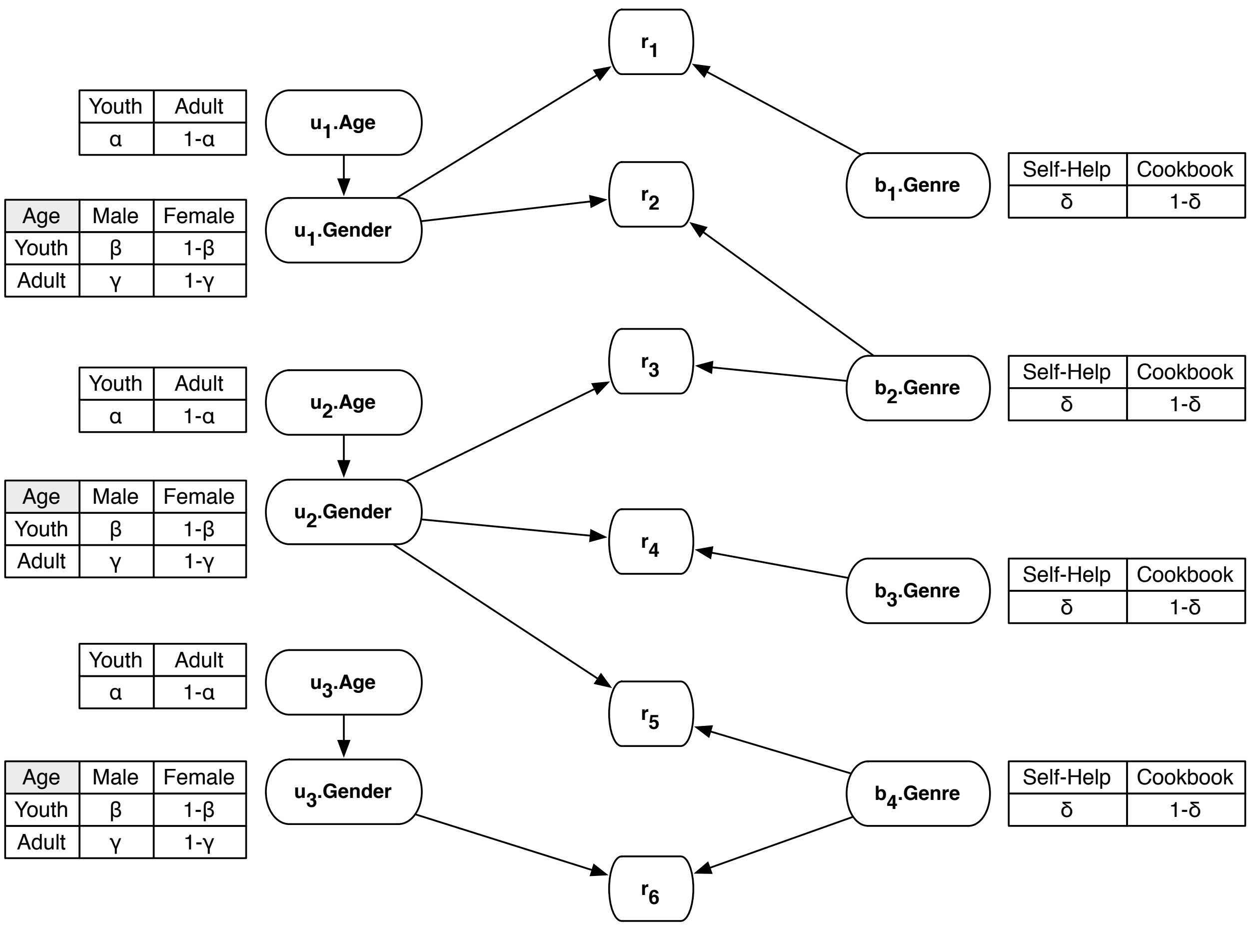
Create a naïve Bayes network on binary variables $\mathcal{V} = \{C, X_1, \ldots, X_k\}$ where C is the class and $X_1, X_2, \ldots, X_k$ are the features. Choose some parameterization of the network such that each parameter $0 < \theta_{x_i|\mathbf{u}_i} < 1$. In the context of variable elimination

1 What ordering on $\mathcal{V}$ has minimum induced treewidth (call it $\prec_o$) ?

2 What ordering on $\mathcal{V}$ has maximum induced treewidth (call it $\prec_w$) ?

Using your implementation of variable elimination

[1] $\alpha = 0.3, \beta = 0.6, \gamma = 0.46, \delta = 0.25, \epsilon = 0.35, \zeta = 0.21, \eta = 0.25, \theta = 0.15, \iota = 0.06, \kappa = 0.18, \lambda = 0.04, \mu = 0.11$



For each r_k node

Gender	Genre	Dislikes	Neutral	Likes
Male	Self-Help	ε	ζ	1-ε-ζ
Male	Cookbook	η	θ	1-η-θ
Female	Self-Help	ι	κ	1-ι-κ
Female	Cookbook	λ	μ	1-λ-μ

Figure 2: Parameters for the elaborate model.

3 For $k = 1 \ldots 10$ compute $\sum_{\mathcal{V}} P(C, X_1, \ldots, X_k)$ using $\prec_o$ and $\prec_w$. Plot the running time of each method vs. k.

(*Note*: Yes, we know that the quantity you are computing is 1.0. The point of the question is to compare running times.)