10708 Graphical Models: Homework 1

Due Monday, February 25, beginning of class

Feburary 8, 2013

Instructions: There are five questions on this assignment. There is a problem involves coding. You can program in whatever language you like, although we suggest MATLAB. Do *not* attach your code to the writeup. Instead, put your code in a directory called "andrewid-HW1" and tar it into a tgz named "andrewid-HW1". For example, epxing-HW1.tgz. Email your tgz file ONLY to gunhee@cs.cmu.edu, seunghak@cs.cmu.edu and kpuniyan@cs.cmu.edu. Refer to the web page for the policies regarding collaboration, due dates, extensions, and late days.

1 Conditional Independencies

1.1 Independence Properties [10 points]

Prove or disprove (by providing a counter-example) each of the following properties of independence:

1. $(X \perp Y, W | Z)$ implies $(X \perp Y | Z)$.
2. $(X \perp Y, W | Z)$ implies $(X, W \perp Y | Z)$.
3. $(X \perp Y, W | Z)$ and $(Y \perp W | Z)$ imply $(X, W \perp Y | Z)$.

1.2 Conditional Probability Distribution [5 points]

Provide an example of a distribution $P(X_1, X_2, X_3)$ where for each $i \neq j$, we have that $(X_i \perp X_j) \in I(P)$, but we also have that $(\{X_1, X_2\} \perp X_3) \notin I(P)$.

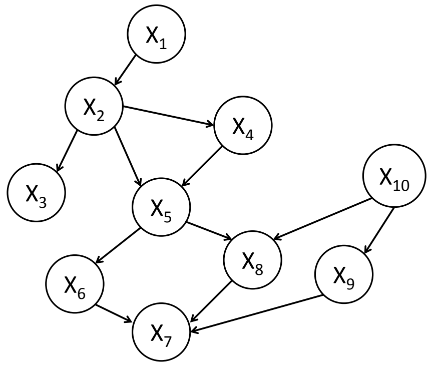

Figure 1: Bayes Net for Q1

1.3 Bayes Nets [5 points]

With reference to figure 1, for each of the following assertions of (conditional) independence, state if they are True or False with justication.

1. $X_1 \perp X_9$
2. $X_{10} \perp X_7 | X_9$
3. $X_2 \perp X_4 | X_5$
4. $X_2 \perp X_8 | X_5$
5. $X_5 \perp X_{10} | X_7$
6. $X_5 \perp X_{10} | X_8$
7. $X_1 \perp X_9 | X_7$

2 I-equivalence [20 points]

Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two graphs over $\mathcal{X}$. In this question we will explore when $\mathcal{G}_1$ and $\mathcal{G}_2$ are I-equivalent.

1. Prove that two network structures $\mathcal{G}_1$ and $\mathcal{G}_2$ are I-equivalent if the following two conditions hold:
   (a) The two graphs have the same set of trails, and
   (b) A trail is active in $\mathcal{G}_1$ iff it is active in $\mathcal{G}_2$.
   (*Hint:* Use the notion of d-separation.)

2. Prove that if $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same skeleton and the same set of v-structures then they are I-equivalent.(*Hint:* use the result from part 1)

3. Can part 2 be extended to an if and only if statement? If so, prove the other direction. If not, provide an example of two I-equivalent graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ that have the same skeleton, but different v-structures.

Your answers to the above questions should convince you that same v-structures, although sufficient, are not necessary for I-equivalence. In the following parts, you will provide a condition that precisely relates I-equivalence and similarity of network structures. We begin with a few definitions you will need:

Definition 1 (Minimal Active Trail) *Consider an active trail $T = X_1, X_2, \ldots, X_m$. We call this active trail* minimal *if no subset of the nodes in T forms an active trail between X_1 and X_m. In other words, T is minimal if no other active trail between X_1 and X_m "shortcuts" any of the nodes in T.*

Definition 2 (Triangle) *Consider a trail $T = X_1, X_2, \ldots, X_m$. We call any three consecutive nodes in the trail a* triangle *if their undirected skeleton is fully connected (i.e., forms a 3-clique). In other words, X_{i-1}, X_i, X_{i+1} form a triangle if we have $X_{i-1} \rightleftharpoons X_i \rightleftharpoons X_{i+1}$ and $X_{i-1} \rightleftharpoons X_{i+1}$.*

4. Prove that the only possible triangle in a minimal active trail is one where $X_{i-1} \leftarrow X_i \rightarrow X_{i+1}$, with an edge between X_{i-1} and X_{i+1}, and where either X_{i-1} or X_{i+1} is the center of a v-structure in the trail. (*Hint:* prove by cases.)

5. Consider two networks $\mathcal{G}_1$ and $\mathcal{G}_2$ that have the same skeleton and same immoralities. Prove, using the notion of minimal active trail, that $\mathcal{G}_1$ and $\mathcal{G}_2$ imply precisely the same conditional independence assumptions, i.e., that if X and Y are d-separated given $\mathbf{Z}$ in $\mathcal{G}_1$, then X and Y are also d-separated given $\mathbf{Z}$ in $\mathcal{G}_2$. (*Hint:* prove by contradiction.)

6. Finally, prove that two networks $\mathcal{G}_1$ and $\mathcal{G}_2$ that induce the same conditional independence assumptions must have the same skeleton and the same immoralities. (*Hint:* prove by contradiction.)

3 Variable Elimination [30 points]

3.1 Example [14 pts]

Once upon a time, your friendly neighbourhood TA was intent on world domination, but now just wants to graduate. Her first step, obviously, was to build a graphical model, as shown in Figure 2. The variables being: Graduate (G), which depends on papers (P) and proposal (R). As we know, papers and theorems (T) are both generated by drinking lots of coffee (C). And all grad students are dependent on free coffee provided by the university (U),

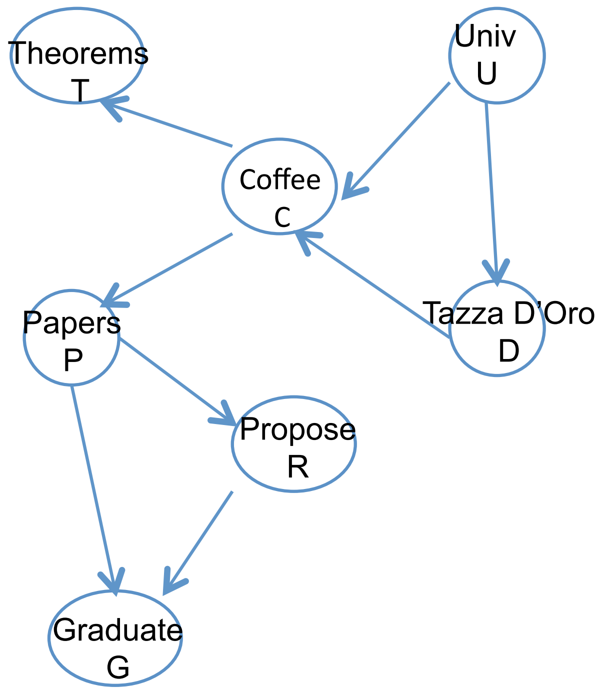

Figure 2: Plans to graduate network

which also allows our favorite coffee shop (D). All the variables are binary valued $\{T, F\}$. The CPT parameters are:

$$P(U = T) = 0.1 \tag{1}$$

$$P(D = T|U = T) = 0.6, \quad P(D = T|U = F) = 0.5 \tag{2}$$

$$\begin{aligned} P(C = T|U = T, D = T) = 0.7, \quad P(C = T|U = T, D = F) = 0.5 \\ P(C = T|U = F, D = T) = 0.6, \quad P(C = T|U = F, D = F) = 0.05 \end{aligned} \tag{3}$$

$$P(P = T|C = T) = 0.8, \quad P(P = T|C = F) = 0.6 \tag{4}$$

$$P(T = T|C = T) = 0.7, \quad P(T = T|C = F) = 0.6 \tag{5}$$

$$P(R = T|P = T) = 0.7, \quad P(R = T|P = F) = 0.1 \tag{6}$$

$$\begin{aligned} P(G = T|P = T, R = T) = 0.9, \quad P(G = T|P = T, R = F) = 0.3 \\ P(G = T|P = F, R = T) = 0.5, \quad P(G = T|P = F, R = F) = 0.1 \end{aligned} \tag{7}$$

Help your friendly TA make some urgent inferences about her graduation plans; but make sure you're not baited by her nemesis: Exponential Computational Complexity.

1. How likely is the TA to graduate if her coffee supply runs out? $P(G = T|C = F) = ?$
2. How likely is the TA to graduate if Tazzo is open all the time? $P(G = T|D = T) = ?$
3. Should we even be worried about her coffee supply running out? $P(C = T) = ?$
4. If she is writing theorems, does this mean she will also publish? $P(P = T|T = T) = ?$

Additionally, report the ordering used and the factors produced after eliminating each variable for the first query [$P(G = T|C = F)$].

3.2 Variable Elimination in Clique Trees [12 pts]

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 $P(X_i, X_i + 1)$. Your goal in this question is to compute $P(X_i, X_j)$, for any $j > i$.

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.
2. What is the running time of the algorithm in part one ? if you wanted to compute $P(X_i, X_j)$ for all n choose 2 choices of i and j?
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 3.2.1.
4. Using the intuition in part three, 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 3.2.2. What is the asymptotic running time of your algorithm?

3.3 Chains or trees [4 pts]

Discuss whether true or false: the complexity of variable elimination is the same in graphical models that are chains or trees.

4 Belief Propagation [20 points]

Two graduate students in CMU have gotten into an argument over the weather. One thinks summer is over and Autumn has already come, while the other thinks it is still summer.

In Pittsburgh, there are four seasons – Spring (S), Summer (M), Autumn (A), and Winter (W). Given the season the previous day, the season on a day is conditionally independent of the season on all previous days. The weather is either Hot (H), Rainy (R) or Freezing (F). Given the season on any given day, the weather that day is independent of all other variables.

More formally, if we let C^i denote the season on the i-th day (taking values S, M, A, W) and O^i denote the observed weather pattern (one of H, R, F). We have $\forall j < i-1, C^i \perp C^j | C^{i-1}$ and $\forall X, O^i \perp X | C^i$ where X is any random variable other than C^i, O^i.

1. (3 pts) Draw a graphical model over $C^1 \dots C^N, O^1 \dots O^N$ that satisfies the conditional independencies listed above.
2. (7 pts) Implement sum-product and max-product algorithms in R or MATLAB for this graphical model.
3. (10 pts) We has made 20 observation of the weather over the last few months (*i.e.*, $O^1 \dots O^N$): $\{R, F, F, H, F, H, H, H, H, H, H, H, H, R, H, H, H, R, H, H\}$.

   Some of the values for the conditional probability table (CPT) are as follows.

   $P(C^1)$:

S	M	A	W
0.15	0.6	0.2	0.05

   $P(C^{t+1} = j | C^t = i)$ for all $t \geq 1$ (i:row, j:column)

	S	M	A	W
S	0.8	0.17	0.02	0.01
M	0.1	0.7	0.19	0.01
A	0.02	0.05	0.7	0.23
W	0.2	0.01	0.04	0.7

   $P(O^t = j | C^t = i)$

	H	R	F
S	0.4	0.3	0.3
M	0.5	0.45	0.05
A	0.3	0.4	0.3
W	0.0001	0.2499	0.75

   For inference, apply both sum-product and max-product algorithms to the following problems. Submit all of your codes (zipped as 'hw1_bp.zip') and report the results.

   (a) Compute the probability of (S, M, A, W) for each of all 20 observations (*e.g.*, $\forall t,\ P(C^t = M | O^1 \dots O^N)$). Save the result of (4×20) probability matrix as 'gamma.txt, and draw it into a figure as 'gamma.png' (x-axis: 20 time steps, y-axis: probability). Submit the 'gamma.txt' and ' 'gamma.png'.

   (b) Determine the most likely sequence of $C^1 \dots C^N$ that generated this observed sequence.

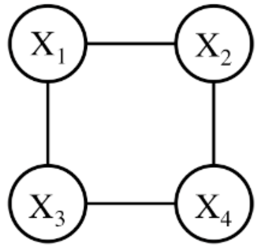

Figure 3: Correctness of Max-product

5 Correctness of Max-Product [10 points]

Show that max-product algorithm is not correct for a simple loopy graph in Fig.3. (*hint*: Use the similar technique discussed in class).