University of Virginia

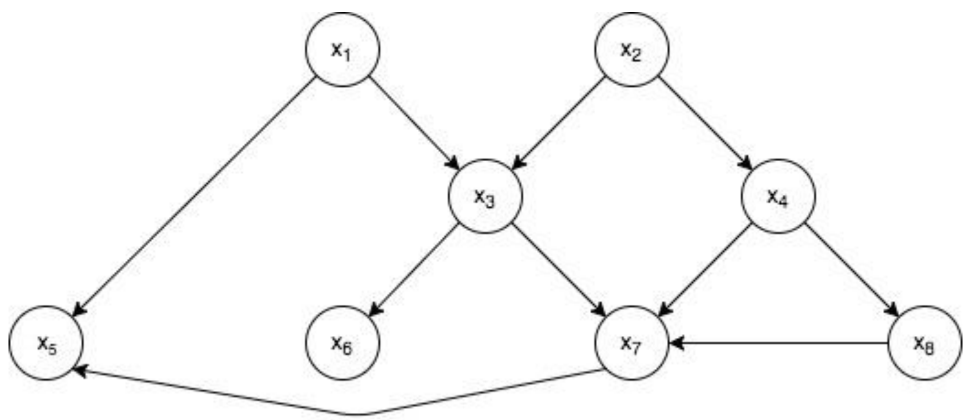
Statistical Learning and Graphical Models, Fall 2017

Homework 1 Due date: 9/7, 5:00 pm

Total points: 65 + 20* (*ed problems are usually harder, aimed primarily at graduate students)
UG grading : $max(p, 65) \times\ 100/60$
Grad grading : $max(p, 85) \times\ 100/80$
where p is the points obtained in the HW.

It is OK to discuss the problems with others but you must list the names of the people with whom you have had discussions. Also, solutions must be written and turned in individually, not in groups.

1. Consider the Bayesian network below.



A. (10 pts) Identify if each independence statement below is true or false. A statement is true if it is true for any probability distribution that factorizes with respect to the network above, that is, it must follow from d-separation.

i. $x_1 \perp\!\!\!\perp x_2 \mid x_3$
ii. $x_1, x_8 \perp\!\!\!\perp x_4 \mid x_2$
iii. $x_3 \perp\!\!\!\perp x_4 \mid x_2$
iv. $x_1 \perp\!\!\!\perp x_2 \mid x_6$
v. $x_8 \perp\!\!\!\perp x_1 \mid x_2, x_3$

B. (5 pts) List all independence properties of the form $x_i \perp\!\!\!\perp x_j$ that follow from d-separation. Note that there is no conditioning.

C. (5pts*) Continuing part B, prove no other such relations exist.

2. Bayesian Networks are defined over DAGs (Directed Acyclic Graphs), that is directed graphs with no directed cycles. In this problem, we explore the relationship between DAGs and topological orderings. A topological ordering (TO) is an ordering of the vertices of a directed graph such that there are no edges in the graph from a vertex to another vertex that appears before it in the ordering.

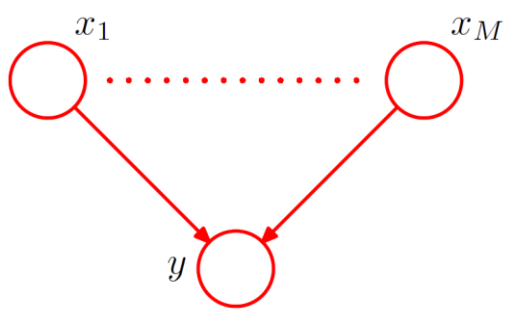
A. (5pts) Prove that if a directed graph has a TO it must be a DAG.

B. (5pts) Find a TO for the graph of problem 1.

C. (5pts) Prove that every DAG has a TO.

Hint for C: Show that every DAG has a node with no incoming edges.

3. (8.6 of Bishop) Consider the Bayesian network below.



A. (5 pts) What is the number of parameters required to specify the CPD $p(y|x_1^M)$, where $y, x_i \in \{0,1\}$.

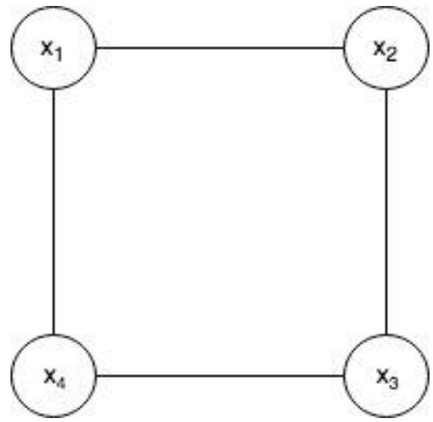
B. (5 pts) An alternative parametric representation (Pearl, 1988) is given by

$$p(y = 0|x_1^M) = \prod_{i=1}^{M} q_i^{x_i}$$

which reduces the number of parameters to M. This conditional distribution is known as the noisy-OR model. Argue that this can be interpreted as a 'soft' (probabilistic) form of the logical OR function (i.e., the function that gives y = 1 whenever at least one of the $x_i = 1$).

Hint: Describe what happens when some x_i is changed from 0 to 1. How does this differ from logical or?

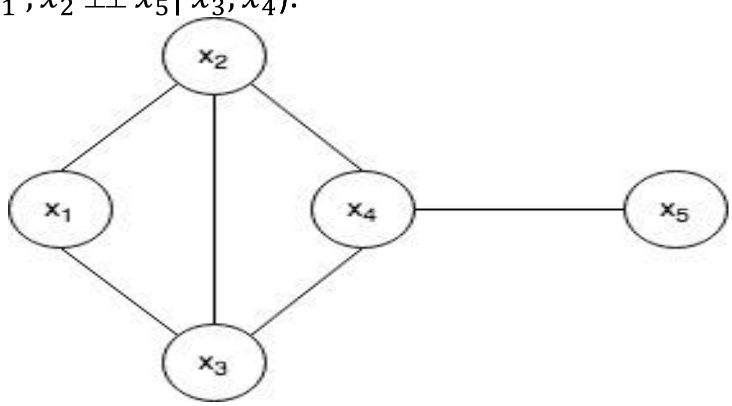
4. Consider the Markov Random Field (MRF) below with $P(x_1^4) = \frac{1}{Z}\psi(x_1, x_4)$ $\psi(x_2, x_3)\ \psi(x_3, x_4)\ \psi(x_1, x_4)$.



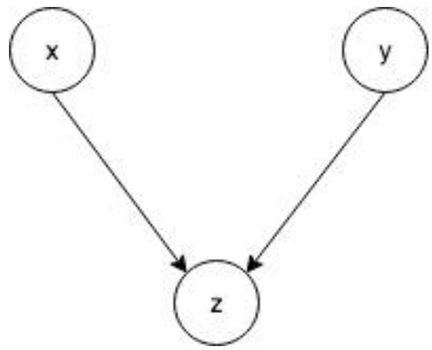
   A. (5 pts) Show that $x_1 \perp\!\!\!\perp x_3 | x_2, x_4$
   B. (10 pts) Find Z if $\psi(x_i, x_j) = \begin{cases} 1, x_i = x_j \\ 1/2, x_i \neq x_j \end{cases}$

5. Prove that
   A. (5 pts) For events A, B, C, if A, B $\perp\!\!\!\perp$ C|D then A $\perp\!\!\!\perp$ C|D
   B. (5 pts) For events A, B, C, if A $\perp\!\!\!\perp$ B|C then P(A|BC) = P(A|B), assuming P(BC)$\neq$0

6. (10 pts) For the MRF below write all the conditional independence statements in which all 5 nodes are present and there are 2 conditioning nodes (eg. $x_1, x_2 \perp\!\!\!\perp x_5 | x_3, x_4$).



7. (15 pts*) Consider the following BN, where x $\in$ { x_0 , x_1}, y $\in$ { y_0 , y_1}, z $\in$ { z_0 , z_1}.



Assume that

$$p(z_1|x_1, y) > p(z_1|x_0, y) \; for \; y \in \{y_0, y_1\} \qquad (1)$$
$$p(z_1|x, y_1) > p(z_1|x, y_0) \; for \; x \in \{x_0, x_1\} \qquad (2)$$

These conditions imply that both x and y positively influence z.
Recall that "explaining away" refers to the relation $p(x_1| z_1) > p(x_1| y_1 , z_1)$.
Although it may seem counter intuitive, this relation does not always hold, even if (1) and (2) hold. For example, consider

	$p(z_1\| xy)$	$p(z_0\| xy)$
x_0y_0	0.1	0.9
x_0y_1	0.11	0.89
x_1y_0	0.11	0.89
x_1y_1	0.9	0.1

Find conditions on the CPD of z such that "explaining away" occurs.