Homework 3
Statistical methods in AI/ML

Instructor: Vibhav Gogate
Vibhav.Gogate@utdallas.edu

Due date: March 18, in class

Problem 1: [20 points] **

a	$p(a)$
0	0.3
1	0.7

b	$p(b)$
0	0.6
1	0.4

e	$p(e)$
0	0.7
1	0.3

y	x	$p(x\|y)$
0	0	0.10
0	1	0.90
1	0	0.30
1	1	0.70

z	y	x	$p(x\|y,z)$
0	0	0	0.25
0	0	1	0.75
0	1	0	0.60
0	1	1	0.40
1	0	0	0.10
1	0	1	0.90
1	1	0	0.20
1	1	1	0.80

Figure 3: Conditional probability tables

The question investigates the AND/OR search space of the network given in Figure 1 assuming that each variable is binary. The CPTs are given in Figure 3.

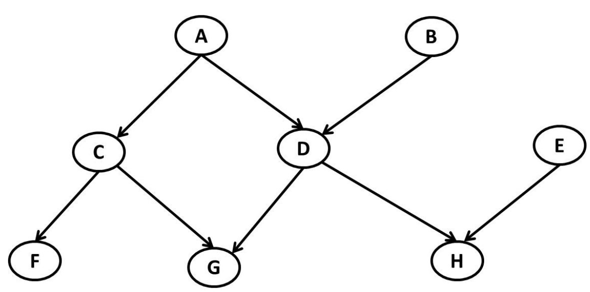


Figure 1:

The CPTs for G, H and D are identical to the 3-dimensional CPT in Figure 3 and the CPTs for H and F are identical to the 2-dimensional CPT in the same figure

- Find and present a pseudo tree of the network whose depth is minimal. Call it T_1. Do the best you can.
- Generate an AND/OR search tree driven by T_1 assuming that each variable has at most two values.
- Annotate the arcs with appropriate weights
- What is the computational cost of computing the probability of evidence $G = 0$ and $H = 1$ in such a network if you use depth-first search over the AND/OR search tree. Demonstrate your computation.
- Can the AND/OR search tree be reduced to a smaller AND/OR search graph.
- Assume that the CPT $P(x|y, z)$ is changed by making some entries deterministic as follows: the first two probabilities are changed to 1 and 0 respectively. Similarly, the last two probabilities are changed to 1 and 0 respectively. Show what would be the changes in the AND/OR search tree as a result.

Problem 2: Formula-based inference [20 points]

Read the first 3 sections of the paper "Vibhav Gogate and Pedro Domingos, Formula-Based Probabilistic Inference, In 26th Conference on Uncertainty in Artificial Intelligence (UAI), 2010." available on my publications page.

Recall that the partition function associated with a set of weighted formulas $\{(f_1, w_1), \ldots, (f_m, w_m)\}$ is given by:

$$Z = \sum_{\mathbf{x}} \prod_{i=1}^{m} \phi_{f_i}(\mathbf{x}) \tag{1}$$

where $\mathbf{x}$ is an assignment of values to all variables, $\phi_{f_i}(\mathbf{x}) = w_i$ if $\mathbf{x}$ satisfies f_i and 1 otherwise.

In class, we saw a possible way of encoding a Markov network into weighted logic. To recap, the encoding works as follows. Assume that all variables are binary. For each tuple $[L_1, \ldots, L_n, w]$ in each potential, where $L_1, \ldots, L_k$ are literals, we added a formula (F, w) where $F = L_1 \wedge \ldots \wedge L_n$. Let us call it *Encoding 1*.

Consider an alternative encoding: For each tuple m $[L_1, \ldots, L_n, w]$ in each potential, where $L_1, \ldots, L_k$ are literals, we add two formulas: $(\neg(L_1 \wedge \ldots \wedge L_k \Leftrightarrow A_m), 0)$ and (A_m, w) where A_m is a propositional variable associated with the tuple. We will refer to this encoding as *Encoding 2*.

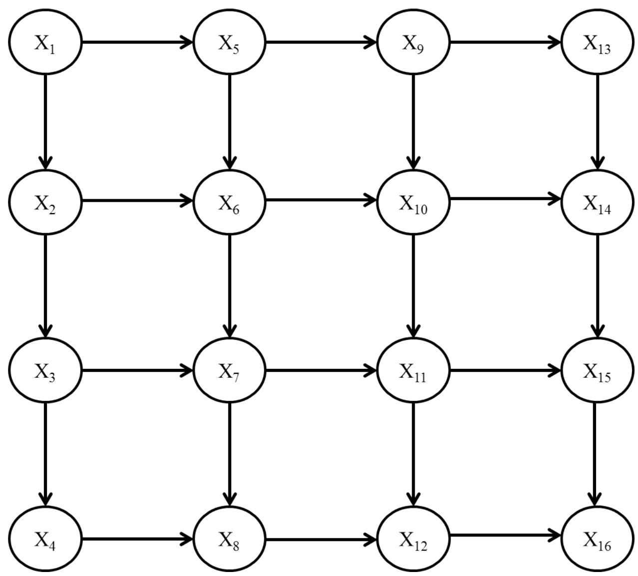
Prove that Encoding 1 and Encoding 2 are equivalent in the sense that they have the same partition function.

Assume that you are using the Logical Decomposition and Conditioning for inference (see the slides). Which encoding (Encoding 1 or Encoding 2) do you think will yield a smaller search space?

Now, assume that you are using Formula Decomposition and conditioning for inference, namely we condition on formulas instead of variables. Which encoding do you think will yield a smaller search space?

Problem 3: Inference [10 points]*** Consider a chain Markov network $X_1 - X_2 - X_3 - \ldots - X_n$. Provide an optimal algorithm which calculates $\Pr(X_i, X_j)$ for **all pairs** $i \neq j$. Prove its optimality.

Problem 4: Iterative Join Graph Propagation [20 points] Assume you are given a 4×4 directed grid (Bayesian network) shown below:



1. Construct an arc-minimal join-graph whose maximal cluster size is 4. Explicitly show the variables and functions in each cluster. Label the arcs with the appropriate separators.

2. Show the schematic messages for one iteration of IJGP on your join graph.