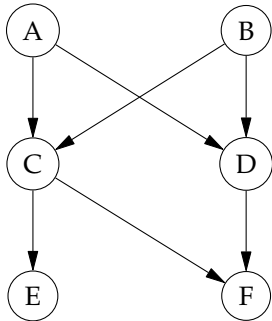


Review Problems

CS 242

April 30, 2024

1. Consider the following belief net:



If node C is observed in the above belief net, which nodes are dependent

on A:

on B:

on D:

on E:

on F:

If nodes C and D are observed, which nodes are dependent

on A:

on B:

on E:

on F:

Show the dependency graph (moralized graph) corresponding to the belief network:

Show a valid tree decomposition of the dependency graph:

2. Show that the three forms of independence below are equivalent:

$$P(X, Y) = P(X)P(Y)$$

$$P(X|Y) = P(X)$$

$$P(Y|X) = P(Y)$$

3. It is quite often useful to consider the effect of some specific propositions in the context of some general background evidence that remains fixed, rather than in the complete absence of information. The following questions ask you to prove more general versions of the product rule and Bayes' rule, with respect to some background evidence \mathbf{e} :

- (a) Prove the conditionalized version of the general product rule:

$$\mathbf{P}(X, Y | \mathbf{e}) = \mathbf{P}(X | Y, \mathbf{e})\mathbf{P}(Y | \mathbf{e}) .$$

(b) Prove the conditionalized version of Bayes' rule:

$$\mathbf{P}(Y | X, \mathbf{e}) = \frac{\mathbf{P}(X | Y, \mathbf{e})\mathbf{P}(Y | \mathbf{e})}{\mathbf{P}(X | \mathbf{e})}$$

4. Consider two medical tests, A and B, for a virus. Test A is 95% effective at recognizing the virus when it is present, but has a 10% false positive rate (indicating that the virus is present, when it is not). Test B is 90% effective at recognizing the virus, but has a 5% false positive rate. The two tests use independent methods of identifying the virus. The virus is carried by 1% of all people. Say that a person is tested for the virus using only one of the tests, and that test comes back positive for carrying the virus. Which test returning positive is more indicative of someone really carrying the virus? Justify your answer mathematically.
5. This exercise investigates the way in which conditional independence relationships affect the amount of information needed for probabilistic calculations.
 - (a) Suppose we wish to calculate $P(h | e_1, e_2)$ and we have no conditional independence information. Which of the following sets of numbers are sufficient for the calculation?
 - i. $\mathbf{P}(E_1, E_2), \mathbf{P}(H), \mathbf{P}(E_1|H), \mathbf{P}(E_2|H)$
 - ii. $\mathbf{P}(E_1, E_2), \mathbf{P}(H), \mathbf{P}(E_1, E_2|H)$
 - iii. $\mathbf{P}(H), \mathbf{P}(E_1|H), \mathbf{P}(E_2|H)$
 - (b) Suppose we know that $\mathbf{P}(E_1|H, E_2) = \mathbf{P}(E_1|H)$ for all values of H, E_1, E_2 . Now which of the three sets are sufficient?
6. Consider the variable elimination algorithm in Sec. 13.3.2.
 - (a) Sec. 13.3.2 applies variable elimination to the query

$$\mathbf{P}(\text{Burglary} | \text{JohnCalls} = \text{true}, \text{MaryCalls} = \text{true}) .$$

Perform the calculations indicated and check that the answer is correct.

- (b) Count the number of arithmetic operations performed, and compare it with the number performed by the enumeration algorithm.
 - (c) Suppose a network has the form of a *chain*: a sequence of Boolean variables X_1, \dots, X_n where $\text{Parents}(X_i) = \{X_{i-1}\}$ for $i = 2, \dots, n$. What is the complexity of computing $\mathbf{P}(X_1 | X_n = \text{true})$ using enumeration? Using variable elimination?
 - (d) Prove that the complexity of running variable elimination on a polytree network is linear in the size of the tree for any variable ordering consistent with the network structure.
7. Investigate the complexity of exact inference in general Bayesian networks:
 - (a) Prove that any 3-SAT problem can be reduced to exact inference in a Bayesian network constructed to represent the particular problem and hence that exact inference is NP-hard. (*Hint*: Consider a network with one variable for each proposition symbol, one for each clause, and one for the conjunction of clauses.)
 - (b) The problem of counting the number of satisfying assignments for a 3-SAT problem is #P-complete. Show that exact inference is at least as hard as this.
 8. Show that any second-order Markov process can be rewritten as a first-order Markov process with an augmented set of state variables. Can this always be done *parsimoniously*, i.e., without increasing the number of parameters needed to specify the transition model?
 9. Suppose we generate a training set from a decision tree and then apply decision-tree learning to that training set. Is it the case that the learning algorithm will eventually return the correct tree as the training-set size goes to infinity? Why or why not?

10. Construct by hand a neural network that computes the XOR function of two inputs. Make sure to specify what sort of units you are using.
11. A simple perceptron cannot represent XOR (or, generally, the parity function of its inputs). Describe what happens to the weights of a four-input, hard-threshold perceptron, beginning with all weights set to 0.1, as examples of the parity function arrive.