CS 440: Introduction to AI

Homework 2

Due: Thursday September 24th

Your answers must be concise and clear. Explain sufficiently that we can easily determine what you understand. We will give more points for a brief interesting discussion with no answer than for a bluffing answer.

Solutions will be posted no sooner than two days after the due date. Homework will be accepted until that point with a penalty of 10% per day that it is late. No assignments will be accepted after the solutions have been posted. Late homework will only be accepted in class, during office hours, or electronically by email to the TA.

You are expected to do each homework on your own. You may discuss concepts with your classmates, but there must be no interactions about solutions. You may consult the web but the work handed in must be done on your own.

The penalty for cheating on any assignment is straightforward. On the first occurrence, you will receive a zero for the assignment and your course grade will be reduced by one full letter grade. A second occurrence will result in course failure.

1) After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease and that the test is 99% accurate (i.e., the probability of testing positive when you have the disease is 0.99 and the probability of testing negative when you don't have the disease is 0.99). The good news is that this is a rare disease, striking only 1 in 10,000 individuals. Why is it good news that the disease is rare? What are the chances that you actually have the disease?

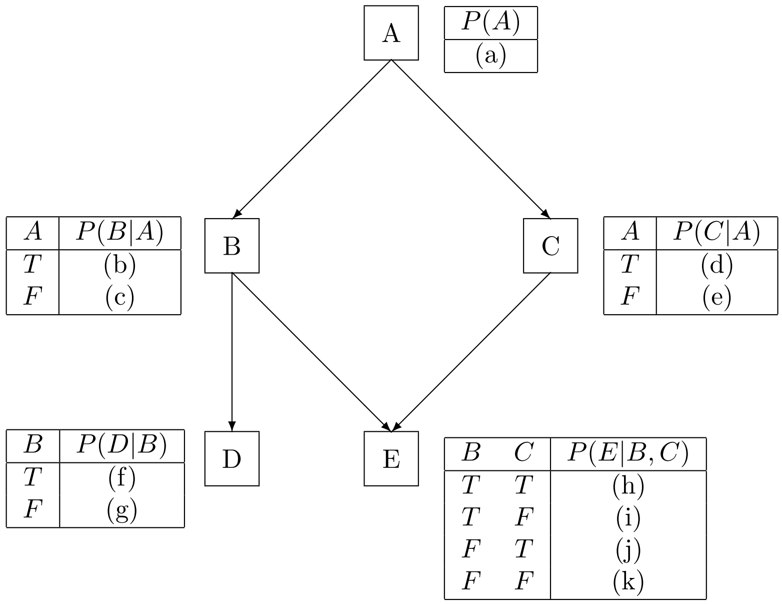
2) Our research lab is having problems with his autonomous car and wants to model the system using a Bayesian network. The objective is to avoid obstacles by reducing the speed of the car and turning to drive around the obstacle. The variable **ObstaclePresent** denotes the presence of an obstacle in the route. If an obstacle is present it is very likely that the obstacle detection system will detect this obstacle. **ObstacleDetected** determines whether and obstacle was detected or not. Besides obstacles, the autonomous driver has to pay attention in cars coming from behind. The variable **CarBehind** indicates has value true if there is a car behind our autonomous car. The variable **CarDetected** indicates that a car was detected behind our car, and it is likely to be true if there is actually a car behind.

If we detect an obstacle the driver is likely to decide to turn around, as well as it is likely to decide to reduce the speed. However, the driver is also likely to decide not to reduce the speed if it detects a car behind. **DecideToTurn** indicates that the driver decided to turn and **DecideToReduce** indicates that the driver decided to reduce the speed.

Currently, one of the big problems of our autonomous driver is the communication between the decision components and the actuators. Even if the driver decides to reduce the speed there is chance that the car will not reduce the speed and even if the driver decides to turn there is a chance that the car will not turn. There is also a chance that the car will reduce speed or turn without a decision from the driver. **ReducedSpeed** indicates that the car actually reduced its speed while **Turned** indicates whether the car actually turned or not. It is also important to remember that the car is likely to fail if it attempts to turn while driving in high speeds.

- **a)** Use the ordering **ObstacleDetected**, **ObstaclePresent**, **Turned**, **DecideToTurn**, **CarBehind**, **ReducedSpeed**, **CarDetected** and **DecideToReduce** to build a Bayesian network to represent this problem. How many parameters are necessary represent the conditional probabilities in this network?
- **b)** What is the best ordering of variables that you can find for this problem? Give the resulting graphical representation of the Bayesian network and say how many parameters would we need at each node.

3) Given the following Bayesian network of Boolean variables:



and the joint probability table

		A				$\neg A$			
		B		$\neg B$		B		$\neg B$	
		C	$\neg C$	C	$\neg C$	C	$\neg C$	C	$\neg C$
D	E	0.00084	0.00378	0.00378	0.03888	0.0096	0.0048	0.1512	0.1728
	$\neg E$	0.00336	0.03402	0.00162	0.00972	0.0384	0.0432	0.0648	0.0432
$\neg D$	E	0.00196	0.00882	0.00042	0.00432	0.0224	0.0112	0.0168	0.0192
	$\neg E$	0.00784	0.07938	0.00018	0.00108	0.0896	0.1008	0.0072	0.0048

a) Compute the probabilities *a* to *k* showing your work.

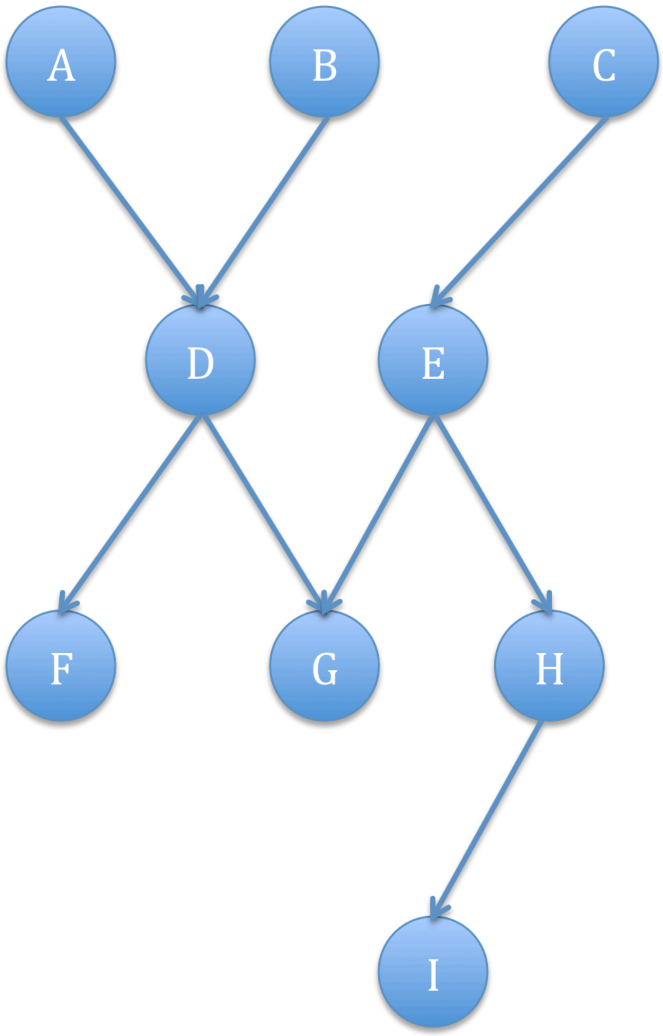
b)

i) Calculate P(E=T|A=T) (or P(e|a)) in terms of the probabilities *a* to *k*. (Use the letters *a* to *k* and not the number that they represent).

ii) Use the results from i) and a) to obtain a value for the probability.

iii) Use the values from the joint to check that i) and ii) are correct.

4) Given the Bayesian network



a) Which nodes constitute the Markov blanket of node D? To which nodes D is conditionally independent given D's Markov blanket?

b) Which minimal set of nodes should be known or unknown in order for D and E to be conditionally independent? Same thing for F and G.

c) Are the nodes B and C independent?

d) To which nodes E is conditionally independent given C? D or G?