

CS535c Fall 2004: Homework 1

Out Mon 13 Sep, Due Mon 20 Sep

1 Medical diagnosis

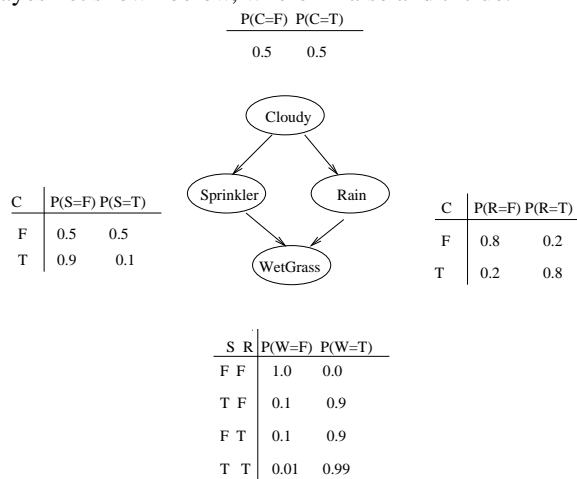
After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease, and that the test is 99% accurate (i.e., the probability of testing positive given that you have the disease is 0.99, as is the probability of testing negative given that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

2 Conditional independence

- 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?
 - $P(E_1, E_2), P(H), P(E_1|H), P(E_2|H)$
 - $P(E_1, E_2), P(H), P(E_1, E_2|H)$
 - $P(E_1|H), P(E_2|H), P(H)$
- Now suppose we know $E_1 \perp E_2|H$. Which of the above 3 sets are sufficient now?

3 Water sprinkler

Consider the water sprinkler Bayes net shown below, where f=false and t=true.



1. Compute $P(S = t|W = t)$
2. Compute $P(R = t|W = t)$
3. Compute $P(W = t)$
4. Which is the more likely cause of the wet grass: sprinkler or rain?
5. Compute $P(S = t|W = t, R = t)$. Is this lower or higher than $P(S = t|W = t)$? Given an intuitive explanation why.

Hint: you may use the following Matlab code fragment if you wish. (Remember that in Matlab, the left-most indices toggle fastest, so the elements of CPDW are stored in the following order: (1,1,1), (2,1,1), (1,2,1), (2,2,1), ...

```
C = 1; S = 2; R = 3; W = 4;
false = 1; true = 2;
CPD{C} = reshape([0.5 0.5], 2, 1);
CPD{R} = reshape([0.8 0.2 0.2 0.8], 2, 2);
CPD{S} = reshape([0.5 0.9 0.5 0.1], 2, 2);
CPD{W} = reshape([1 0.1 0.1 0.01 0 0.9 0.9 0.99], 2, 2, 2);
```