Some optional practice questions in case you want to play with
a Bayesian Network and a Naive Bayes classifier (CS2013)

These two questions expand on examples in the CS2013 lectures

Q1. The last lecture of the course contains a sketchy example of a Bayesian Network involving Holmes and Watson (who else?), a sprinkler and rain. In the lecture, only the influence diagram was given. Now let's add information about probabilities (below), letting R=rain, S=Sprinker, W=Watson's lawn is wet, H=Holmes' lawn is wet. We model a relatively simple situation by equating a number of values to either 1 or 0:

$P(R) = 0.4$
$P(S) = 0.2$
$P(W \mid R) = P(H \mid R) = P(H \mid S) = 1$
$P(W \mid \text{not-}R) = P(H \mid \text{not-}R, \text{not-}S) = 0$

Your task is to calculate some of the remaining probabilities. (Before you calculate, it's useful to ask yourself what outcomes you expect. It might be helpful to look at the diagram.)

a. Calculate $P(R \mid W)$.
b. Calculate $P(H)$.
c. Calculate $P(R \mid H)$.
d. Calculate $P(W \mid H)$.
e. Calculate $P(H \mid W)$.

Q2. The last lecture finished with a discussion of a Bayesian classifier. For concreteness, think of this as data about mushrooms, with

A=foul tasting but harmless,
B=lethal,
C=tasty& healthy.

a. Complete the example in the lecture, calculating the probabilities of the hypotheses B and C (given the data listed) and comparing these with the probability of A. What would you do with your mushrooms?

b. You receive new information: You sent a package containing some of your mushrooms to your uncle in Calgary. Yesterday he called you to say the mushrooms were delicious. You conclude that A is false. Revise your calculations in light of the new information. Which parts of your calculations change? What do you do with the mushrooms?

c. The training data in this example was tiny. Suppose you had many more data, including a data point where x=2, y=3, z=4, classified as B. Would this change your conclusion as to what is the most probable hypothesis?

(see over for some sample answers)

Sample answers

Q1.

a. P(R | W) = (P(W | R) * P(R)) / P(W) = 0.4/ (0.4*1)+(0.6*0) = 1.

b. P(H) = (P(H | R) * P(R)) + (P(H | not-R) * P(not-R) =
(P(H | R) * P(R)) + P(S) * P(not-R) = 0.4 + 0.12 = 0.52.

c. P(R | H) = (P(H | R) * P(R)) / P(H) = 0.4 / 0.52 = approx 0.77.

d. P(W | H) = P(W,R | H) + P(W, not-R | H) = approx 0.77

e. P(H | W) = P(H,R | W) + P(H, not-R | W) = 1.

What you haven't been told of course is where all these probabilities come from.. (Past experience maybe?)

Q2.

a. On this small dataset, Hypothesis A wins hands down:

P(A | data) = 0.0417.
P(B | data) = 0.008.
P(C | data) = 0.01.

b. The new answer is calculated in the same way as before, as P(hypothesis) * (data | hypothesis).
The difference is that A no longer plays a role; moreover, P(hypothesis B) and P(hypothesis C) increase: P(B) = 4/7 (was 4/15) and P(C) = 3/7 (was 3/15).
You're now faced with a choice between two hypotheses whose posterior probability is very close. Acting on C makes you die, so I'd be a bit cautious.

c. It's tempting to just "look up" the answer from the relevant data point, but it's wrong. Acting on one data point is seldom a good idea. The classifier allows you to bring more data to bear, reaching a (somewhat) safer conclusion.