CS 6360: Advanced AI
Assignment 4

Due: 04/13/2017, at midnight

General Instructions:

If anything is ambiguous or unclear:

1. **Discuss possible interpretations with other students, your TA, and instructor**
2. **Make assumptions, state them explicitly, and then use them to work out the problem**
3. **Use Piazza for discussions among yourselves, and also for questions, also for questions and clarifications you need from the instructor and the TA. Piazza levels the playing field because the responses to questions asked are of interest to all, and shared by all. If you have very specific questions, you may email the TA**

Remember that after general discussions with others, you are required to work out the problems by yourself. All submitted work must be your own work. Please refer to the Honor code for clarifications.

Solving Problems using Bayesian Networks

Assignment (100 points)

For this assignment you will solve a set of three uncertain reasoning problems using Bayesian Networks. You will download the JavaBayes package from the website: http://www.cs.cmu.edu/~javabayes/Home/. This is a package developed by researchers at CMU and the package is well-documented (for documentation see the website) and easy to use. You will need to have Java installed before on your desktop/lap top to run JavaBayes. (For downloading and installation of Java, visit https://www.java.com/). A paper describing the variable elimination method used in JayaBayes is available on Blackboard.

You will construct Bayesian networks for the three problems that appear below. For each problem, you will:

1. Construct the appropriate Bayesian network in JavaBeans.
2. Answer the given questions using queries in JavaBayes.
3. Save the Bayes net and submit the required file files. (For example, for problem 1, save the Bayes net as File→Save as P1_BN.) The TA will execute your files using JavaBayes to make sure your model is correct, and it produces the right answer.

Submission: You will submit your code (see bullet 3 above), along with your report that includes the results of your experiments, and explanations for why those results make sense as zipped file on Blackboard.

Problem 1:

You will model a small Bayesian network that represents the relationship between yellow fingers, smoking, cancer, radiation, solar flares, and using a microwave.

In this model, **smoking** can cause **yellow fingers** and **cancer**. **Solar flares** and **making microwave popcorn** can cause **radiation**, and **radiation** can cause **cancer** as well.

The prior probability of **smoking** $P(S)$ is 0.3. The prior probability of **solar flares** $P(F)$ is 0.8. The prior probability of using the **microwave** $is\ P(M)$ is 0.9.

The conditional probability table for radiation is

Solar flares (F)	Making microwave popcorn (M)	P(Radiation) $P(R)$
0	0	0.1
0	1	0.2
1	0	0.2
1	1	0.9

The conditional probability table for cancer is

Smoking S	Radiation R	P(Cancer) $P(C)$
0	0	0.1
0	1	0.6
1	0	0.3
1	1	0.9

The conditional probability table for yellow fingers is

Smoking S	P(Yellow fingers) $P(YF)$
0	0.11
1	0.8

Answer the following questions using JavaBayes:

1. What is the prior probability of cancer?
2. What is the probability of smoking given cancer?
3. What is the probability of smoking given cancer and radiation?
4. Explain why knowing that radiation was present almost cancels the effect of knowing that cancer is present on the probability of smoking.
5. What is the markov blanket of yellow fingers?
6. Are solar flares and using the microwave independent given cancer? Why?
7. What is the probability of cancer if you never use a microwave?
8. Is this a realistic model for predicting the probability of cancer? (if you think so, look at the American Cancer Society web site for actual statistics). Explain how the model could be improved.

Problem 2:

Four generation of laboratory mice have been bred as follows:

Initial generation: Alice, Bob, Cindy, Dave, Ellen

Second generation:

- Fred has parents Alice and Bob
- Gwen has parents Cindy and Bob
- Henry has parents Cindy and Dave
- Iona has parents Ellen and Bob

Third generation:

- John has parents Gwen and Fred
- Katherine has parents Henry and Iona

Fourth generation:

- Louis has parents Katherine and John

The mouse Louis suffers from a life-threatening disease and dies soon after birth. The disease is hereditary and carried by a recessive gene, meaning that two parents have to carry the gene for the offspring to be diseased. Since the disease is recessive, the parents themselves need only be carriers of the disease, but not diseased themselves. A completely healthy animal has genotype *pp* (pure), a carrier is *pd* or *dp* (but hereafter written only as *pd*), and a diseased animal is *dd*.

We solve the following problem using the Bayesian reasoning inference: Which members of the ancestry of Louis are likely to be carriers of the disease?

According to the laws of genetics (and basic probability theory), the possible combinations of parent genotypes as probability triples $(P(dd), P(pd), P(pp))$ for the genotype of the descendent is as follows:

	dd	***pd***	***pp***
dd	(1, 0, 0)	(0.5, 0.5, 0)	(0, 1, 0)
pd	(0.5, 0.5, 0)	(0.25, 0.5, 0.25)	(0, 0.5, 0.5)
pp	(0, 1, 0)	(0, 0.5, 0.5)	(0, 0, 1)

You can make the following assumptions: None of the mice other than Louis are actually diseased, otherwise they would not have lived long enough to procreate. This means that you can reduce the probability above by eliminating the rows and columns marked with dd. Therefore, the conditional probability table $P(x|parents(x))$ for Louis is:

	pd	***pp***
pd	(0.25, 0.5, 0.25)	(0, 0.5, 0.5)
pp	(0, 0.5, 0.5)	(0, 0, 1)

For all the other mice, you can make an additional simplifying assumption: As we know that all the other mice aren't diseased, you will need to calculate only probability pairs $(P(dp), P(pp))$ as the entries of the conditional probability tables $P(x|parents(x))$. Therefore, the CPT's for all the mice in the second and third generations are:

	pd	***pp***
pd	(0.67, 0.33)	(0.5, 0.5)
pp	(0.5, 0.5)	(0, 1)

Finally, you will need prior probability information mice being disease carriers to determine the prior probabilities that the first-generation mice are disease carries. Assume that $P(pd) = 0.01$ and $P(pp) = 0.99$.

Model the above genealogy in a Bayesian network in JavaBayes to calculate the disease carrier probabilities for all the mice conditioned on the fact that Louis has the disease.

Problem 3:

Consider the following domain:

- There are three soccer teams, A, B, and C. Each has a fixed, unknown quality of 0, 1, 2, or 3 (larger is better).
- Each team plays every other team once, resulting in a *win*, *loss*, or *draw*.
- When two teams play, the result is a noisy, probabilistic function of the difference between their qualities:
  - If the difference is 0, a team wins 30%, loses 30%, or draws 40% of the time.
  - If the difference is 1, the better team wins 50%, loses 20%, or draws 30%.
  - If the difference is 2, the better team wins 70%, loses 10%, or draws 20%.
  - If the difference is 3, the better team wins 90% or draws 10%.

Use JavaBayes to build the Bayes net and execute the queries to answer questions. *Parts (c) through (h) should take very little time with the help of the applet.*

1. Draw a reasonable Bayes' net for this domain, briefly justifying what arcs you do and do not include.
2. For each node, specify a reasonable local conditional probability table. If multiple nodes in your network have the same local distributions, you only need to write those distributions once.
3. What is the posterior distribution over the game between B and C, given no evidence? Does this answer make sense?
4. If A beats B, what is the updated posterior distribution over the outcome of the game between B and C? Does this make sense?
5. If A beats B but draws against C, what is the updated posterior distribution over the outcome of the game between B and C? Does this change make sense?
6. If A beats B but draws against C, what are the posterior distributions over the unobserved qualities of each team? Do these distributions make sense?
7. If A beats B (and the result of the game against C is unknown), what is the posterior distribution over C's quality? Why can you answer this question without any numerical calculations?
8. If A is known to have quality 1, what is the posterior distribution over the outcome of the game between A and B? Does this distribution make sense?