15-381 Spring 2007
Assignment 5: Probability and Uncertainty, Bayes Nets and MDPs

Questions to Einat (einat@cs.cmu.edu)

Spring 2007
Out: April 3
Due: April 17, 1:30pm Tuesday

The written portion of this assignment must be turned in at the beginning of class at 1:30pm on April 17th. Type or write neatly; illegible submissions will not receive credit. Write your name and andrew id clearly at the top of your assignment. If you do not write your andrew id on your assignment, you will lose 5 points.

The code portion of this assignment must be submitted electronically by 1:30pm on April 17th. To submit your code, please copy all of the necessary files to the following directory:

/afs/andrew.cmu.edu/course/15/381/hw5_submit_directory/yourandrewid

replacing yourandrewid with your Andrew ID.

Late Policy. Both your written work and code are due at 1:30pm on 4/3. Submitting your work late will affect its score as follows:

- If you submit it after 1:30pm on 4/17 but before 1:30pm on 4/18, it will receive 90% of its score.
- If you submit it after 1:30pm on 4/18 but before 1:30pm on 4/19, it will receive 50% of its score.
- If you submit it after 1:30pm on 4/19, it will receive no score.

Collaboration Policy.

You are to complete this assignment individually. However, you are encouraged to discuss the general algorithms and ideas in the class in order to help each other answer homework questions. You are also welcome to give each other examples that are not on the assignment in order to demonstrate how to solve problems. But we require you to:

- not explicitly tell each other the answers
- not to copy answers
- not to allow your answers to be copied

In those cases where you work with one or more other people on the general discussion of the assignment and surrounding topics, we ask that you specifically record on the assignment the names of the people you were in discussion with (or "none" if you did not talk with anyone else). This is worth five points: for each problem, you solution should either contain the names of people you talked to about it, or "none." If you do not give references for each problem, you will lose five points. This will help resolve the situation where a mistake in general discussion led to a replicated weird error among multiple solutions. This policy has been established in order to be fair to everyone in the class. We have a grading policy of watching for cheating and we will follow up if it is detected.

Problem 1 - Bayes Rule (15 points)

You have just moved to a new town. You were told that there are two types of busses in town: some busses run every 10 minutes, and the others run every 30 minutes (depending on the bus line). Waiting for a bus for the first time in your nearest bus station, you don't yet know which type of bus it serves.

You've been waiting for 20 minutes at the bus station, and the bus hasn't arrived yet.

Assuming that a waiting time interval is modeled by the exponential distribution function:

1. Use Bayes rule to derive your posterior belief about which type of bus this bus stop serves. What is the probability that you are waiting for a bus of the first type (give a number)? of the second type? Show your work, and results.

   Do you find these results intuitive?

2. How would your results be, given that you've been waiting for **5** minutes at the bus stop, and the bus hasn't arrived yet? Give the numerical results.

3. After how many minutes of waiting, would you believe that the bus you're waiting for is more probable to be of the second type?

About the exponential distribution: An exponential distribution is often used to model the time between independent events that happen at a constant average rate.

The probability density function of an exponential distribution has the form:

$$f(x = X \mid \beta) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$$

The cumulative distribution function is given by:

$$F(x <= X \mid \beta) = 1 - e^{-\frac{x}{\beta}}$$

where β is the mean interval length (waiting period, here), $x >= 0$ and $\beta > 0$.

Hint: Since the bus hasn't yet arrived, then what you should be estimating here is $f(\beta \mid x > X)$.

Problem 2 (15 points)

Suppose we have two sensors – x_1, x_2 – taking samples of the random variable y. For example, consider two optical sensors, that are intended at measuring an object's location, y. Each sensor reports a noisy estimate of the true sampled signal. The variability of each sensor is known (it is determined by the specific environment conditions for each sensor). Your job is to determine the best estimate of the object's location.

We will assume the following Gaussian probability distributions: $f_{(y,a^2)}(x_1)$, $f_{(y,b^2)}(x_2)$.

The notation $f_{(m,\sigma^2)}(x)$ denotes a Gaussian density function of varilable x, for which the mean is m and the variance is σ^2:

$$f_{(m,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(x - m)^2)$$

That is, the mean of the measurements given by each sensor is the true sampled location value y in both cases. But, the variance of x_1 is a^2 and the variance of x_2 is b^2.

The location y is also a Gaussian, distributed $f_{(0,1)}(y)$.

1. Describe a Bayes net model for this situation.
2. Derive the posterior distribution $f(y \mid x_1, x_2)$.

   Note: following is a recipe for the product of gaussian densities, which you should find helpful:

   The product of two Gaussians is:

   $$f_{(m_1,\sigma_1^2)}(x) \times f_{(m_2,\sigma_2^2)}(x) = C_c f_{(m_c,\sigma_c^2)}(x)$$

   where

   $$\sigma_c^2 = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}},$$

   $$m_c = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \left(\frac{m_1}{\sigma_1^2} + \frac{m_2}{\sigma_2^2} \right),$$

   and

   $$C_c = f_{m_1,\sigma_1^2+\sigma_2^2}(m_2).$$

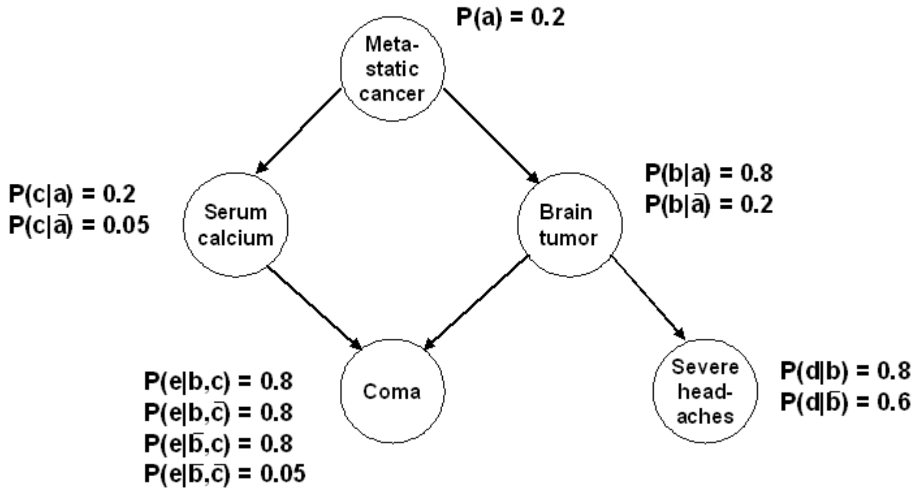
   A hint and advice:

   - Due to the symmetry of the Gaussian density function, $f_{(y,\sigma^2)}(x) = f_{(x,\sigma^2)}(y)$
   - Calculation can be quite neat if you keep C_c as a "black box" constant (if you have things right, the constants would cancel out :-)).
3. What is the maximum likelihood estimate (MLE) for the object's location given both sensor's measurements (that is, what is $E(P(y \mid x_1, x_2))$? (Hint: rather than derive the expression, you may reason about this considering the shape of the posterior distribution function).
4. Interpret the MLE expression, and describe how it weights the information.

Problem 3 - Bayes Nets (35 points)

Given is a simplified version of a network that could be used to diagnose patients arriving at a clinic. Each node in the network corresponds to some condition of the patient. This network demonstrates some cuasality links. For example, both brain tumor and serum calcium increase the chances of a coma. A brain tumor can cause severe headaches and a comma, and so on.

1. (2 points) What is the joint distribution $P(a, b, c, d, e)$? give a factorized expression, according to the network's structure.
2. (3 points) Give an example of 'explaining away' in this Bayes net.
3. (5 points) One of your patients experiences severe headaches, had a comma and serum calcium. What is the probability of him having cancer ? show full dereivation of this probability, as well as the numerical result.
4. (5 points) What is the probability of a positive serum calcium given severe headaches? Derive this expression. Specifically, start from the joint distribution, factorize it and use variable elimination, so as to lower calculation cost. (Note: a numerical result is *not* required here.)



5. (15 points) Write code for sampling joint and conditional probabilities in Bayes nets. Use the standard name "sample" for your code. The command line arguments should be: FILE-NAME (VAR-NAME=value,VAR-NAME=..) (VAR-NAME=value,VAR-NAME=..). (In case you are to use a non-standard language, be sure to included a README file with runtime instructions.)

   The first argument specifies the file describing the Bayes net. The second group of arguments gives the required variables' values, whose probability we'd like to evaluate. Several such values can be specified, using a comma separator within the same round brackets. The last group of arguments gives the variables' values that the requested probability is conditined on (again, multiple values can be specified, using a comma separator and bounding brackets). If the second group is empty (unfilled brackets), this means we are looking for a joint probability expression, that is conditioned on nothing.

   For example: "cancer.text (A=true) (D=true,E=true,C=true)" would give you the probability for question 4.3 .

   A text file that includes this net's information is handed out with this homework, on the class' website. Using the program, evaluate the probabilities specified in questions 4.3 and 4.4, generating n samples (where a single sample assigns values to all the nodes in the network). What are the estimated probabilities for $n = 500$? $n = 1000$? $n = 20,000$?

6. (5 points) Consider the following Bayes net, called "Asia", which can be used to diagnose respiratory diseases (this is a fictitious network).

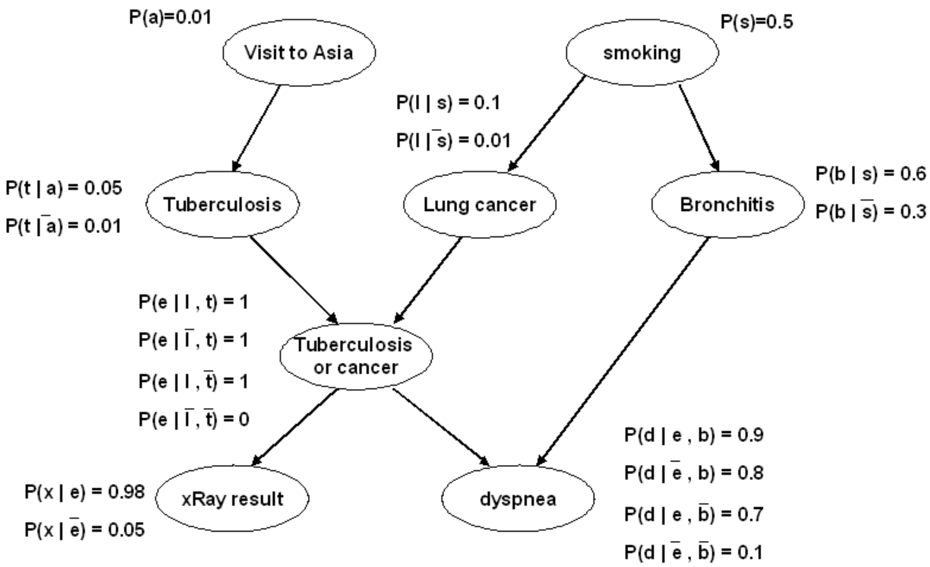
   Use your code to evaluate the probabilities that:

   - a patient who smokes, has visited Asia, and got positive xRay results, has lung cancer (give a numerical result).
   - a patient who smokes and has dyspnea, has got lung cancer (give a numerical result).

Problem 4 - HMM Applications (10 points)

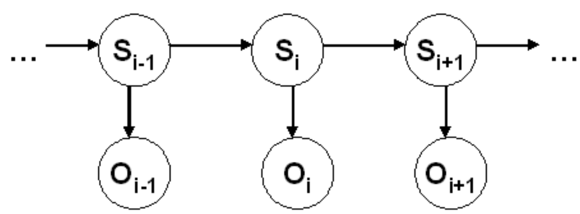
Hidden Markov Models are used for a variety of sequential data processing. As shown in the figure, the model includes a *hidden layer*, which described the data as a sequence of pre-defined states. The output layer maps to the *observed* signals, that are emmitted from the unknown states.

For example, in OCR systems, the hidden nodes represent the sequence of the underlying true characters. The observed decoded characters are not necessarily identical to the hand-written ones, due to the high variability of hand writing. For example, the hand-written "i" character is decoded as either "i", "j", or "l"



with high probabilities. The advantage of using HMM for this problem is that it allows modeling context. For example, if the previous (hidden) state was assumed to be "q", then the current character is more likely to be a vowel.

Given the observed OCR output sequence, the final output to the user is the sequence of states which maximizes the joint probability of the HMM model.



1. (3 points) In the described OCR scenario, what is the size of the CPT table between two states of the hidden layer? What is the size of the emission probabilities table for every state?

2. (3 points) Suggest a way for obtaining the transition and omission probabilities for the model.

3. (4 points) Suggest an HMM representation for the problem of speech recognition, or automatic Part-of-Speech labelling. (POS labelling is task of assigning every word its grammatical label, e.g., noun, verb, preposition etc.) You may suggest an HMM modeling for other problems as well.

   Your suggestion should include the definition of states, the definition of the omitted signals, and a description of the values the hidden and observed nodes can take.

   Explain why you think that the addressed problem would benefit using an HMM model.

Problem 5 - Markov Decision Processes(25 points)

Jane is a student. Every Sunday, she needs to make a decision - either go out with her friends, in which case she will go to sleep late and start the next week tired. Or, she may choose not go out, that is, go to sleep early. Suppose that Jane understood the material in class the week before. Then, if she chooses to sleep early, she is most likely (80%) to understand next week's lectures. If she starts the week tired, though, the chance that she will understand the material in class next week is only 50%. On the other hand, if Jane didn't understand the material in class last week, then the chance that she understands this week's lectures if she goes out is only 25%. Or, if she chooses to got to sleep early and start the week fresh, her chance to understand the classes next week is 50%.

Jane is interested in understanding the material in class, of course. Especially, since in case she doesn't understand the material on a particular week, then she has to put a lot of time into homeworks, which costs her 4 satisfaction points for that week. If she understands the material, it only costs her 1 satisfaction point.

1. (4 points) Describe this situation as a Markov Decision Process diagram, where nodes denote the states, and edges are labeled with the transition probabilities. Be sure to include actions in the diagram's edges.

2. (4 points) Enumerate all the possible policies that Jane can take. For each policy, give its transition matrix.

   In addition to the above, if Jane decides not to go out with her friends on Sunday, then her friends are disappointed. Even worse, they call her a wimp, which costs her 2 satisfaction points. Consider this information in the following questions:

3. (5 points) On the first week of school, Jane is equally likely to understand classes, or not. What is her cost value then for every policy, assuming that the term is short, including three weeks? Assume that future satisfaction is discounted by $\alpha = 0.9$, per week. Show your work.

4. (2 points) What is the optimal policy that Jane should take then ?

5. (10 points) What is Jane's optimal policy for the whole school year (40 weeks). Explain how you calculate this. You may use code or a spreadsheet, for computation purporses. Your answer for this question should include a table of the relevant values per week, until convergence. Also, specify what the optimal policy is.