CSCI 3202 Fall 2003 Quiz 3

Name: _____

*On my honor, as a University of Colorado student, I have neither given nor received unauthorized assistance on this work.*_____

1. **(5 Points)** During the sniper attacks in Washington last year the police focused considerable energy on finding a white van. This focus was based on numerous witness reports placing a white van at or near the location of many of the incidents. In the end, it was determined that the culprits were driving a blue Chevy Caprice. Give me a very short probabilistic account of where the police might have gone wrong with respect to the white van.

2. **(10 Points)** Email spam filters are designed to detect spam messages as they arrive in your mailbox and redirect them to a spam folder. Recently 10% of the mail you've been receiving is spam, so you'd like to get a filter. The best filter you can find claims that 99% of the spam messages it sees will be labeled spam. On the other hand, it incorrectly labels 4% of the non-spam messages it sees as spam.

    You install the filter. You subsequently receive 1000 mail messages. How many of those will be labeled spam? How many of those will actually be spam?

3. **(20 Points)** Consider the following facts. Smoking causes lung cancer. It also separately causes emphysema. It's also known that radon (a colorless, odorless gas prevalent in many locations) can cause lung cancer; Radon is not known to have any other negative effects.
    a. Draw a Bayesian belief net that captures these facts along with the probability tables that would have to go along with it.
    b. Which of the probability tables in your network is likely to be the most problematic with respect to acquiring the required statistics?

    c. Show how you would use this network and the probabilities associated with it to compute the probability of lung cancer given that you know that radon is present.

4.   **(15 Points)** Consider the following problem. You're not an economist and you have no access to economic information telling you anything about the state of the economy. Assume that the economy can only be in one of two states: Recession or Boom; assume further that you know vaguely that the state of the economy now is based somehow on the last state it was in. You also know that when there's a boom usually the most popular car around town is a BMW, and when there's a recession it's usually a Subaru.

    a. Formalize this problem as an HMM problem. Describe the sets of parameters you would need to create a complete model.

    b. Given this model, how would you go about deciding the most likely recent history of the economy given that the sequence of most popular cars has been BMW, Subaru?