CSE 515: Statistical Methods in Computer Science

Homework #2

Due **in class** on April 29, 2009

Guidelines: You can brainstorm with others, but please solve the problems and write up the answers by yourself. You may use textbooks (Koller & Friedman, Russell & Norvig, Wikipedia, etc.), and lecture notes. Please do NOT use any other resources or references (e.g., example code, online problem solutions, etc.) without asking.

Submission instructions: Submit a **hard copy** of this assignment in class to Daniel Lowd.

1. Suppose you wish to do predictive text entry on a cell phone without a built-in dictionary. In predictive text entry, the observations are numbers and each number is associated with 3 or 4 letters (e.g., 2 $\rightarrow$ 'a', 'b', or 'c'). The goal is to decode the intended word given the corresponding sequence of digits. Formulate this problem as a hidden Markov model, answering the following.

   (a) What are the hidden states?

   (b) What are the emission probabilities? How could you estimate parameters for the transition probabilities from freely available data on the internet? Describe what data you would use and how.

   (c) Suppose that, with probability p, a user accidentally presses a random digit instead of the correct one. How can you modify your model to handle this?

   (d) What are the limitations of this HMM as a model for predictive text entry?

2. Consider an HMM with three states, three outputs, and the following transition and emission probabilities. Assume a uniform distribution for the initial state, π_0.

π_i π_{i+1}	a	b	c
a	0.5	0.4	0.1
b	0.1	0.5	0.4
c	0.4	0.1	0.5

π_i x_i	p	q	r
a	0.7	0.1	0.2
b	0.2	0.7	0.1
c	0.1	0.2	0.7

   (a) Compute the most likely sequence of hidden states for the observed sequence (p, p, r, r, q, r) by stepping through the Viterbi algorithm by hand.

   (b) Use the forward-backward algorithm to compute the probability distribution over states at position 3.

   (c) Is the most likely state the same as the state in the most likely sequence? Will this always be the case? Why?

3. Consider a robotic blimp with an altimeter. In each second, the robot moves up approximately 1.0 meters and then takes a (noisy) reading from its altimeter. Assume that the standard deviation of the robot's moves is 0.1 meters, that the altimeter is unbiased with a standard deviation of 0.5 meters, and that the robot's initial altitude (at time zero) is zero (known with certainty). Over the first five seconds, the altimeter readings were as follows: 1.2m,

1.5m, 2.5m, 4.5m, 6m. Use Kalman filtering to compute the mean and variance of the robot's altitude from the current and past observations at each time step.

4. Consider the following statements:

   - People send email more often when indoors.
   - Spending time indoors is a consequence of the weather outside and whether or not you're a vampire. (Vampires spend more time indoors, especially when it's sunny.)
   - Both the current season and time of day affect the weather.
   - Being bitten by a vampire causes you to be a vampire.
   - Vampires are preternaturally beautiful.

   (a) Use these statements to come up with a Bayesian network over variables Email, Indoors, Weather, Time, Season, Vampire, Bitten, and Beauty.

   (b) In this network, what is the Markov blanket of Vampire?

   (c) According to this network, are beautiful people more likely to send email?

   (d) According to this network, are time of day and beauty independent?

   (e) According to this network, are time of day and beauty independent, given email?

   (f) How many different Bayesian network structures are there that specify this same set of independencies?

5. Consider the Bayesian network in Figure 1. It covers symptoms of colds, whether or not the patient chooses to take medicine for the symptoms, and an alternate cause of sneezing (sunlight). All variables are Boolean, with distributions defined by the given conditional probability tables.

   (a) What is the Markov blanket of Light in the Bayesian network?

   (b) Convert the Bayesian network to an equivalent Markov network using one potential for each maximal clique.

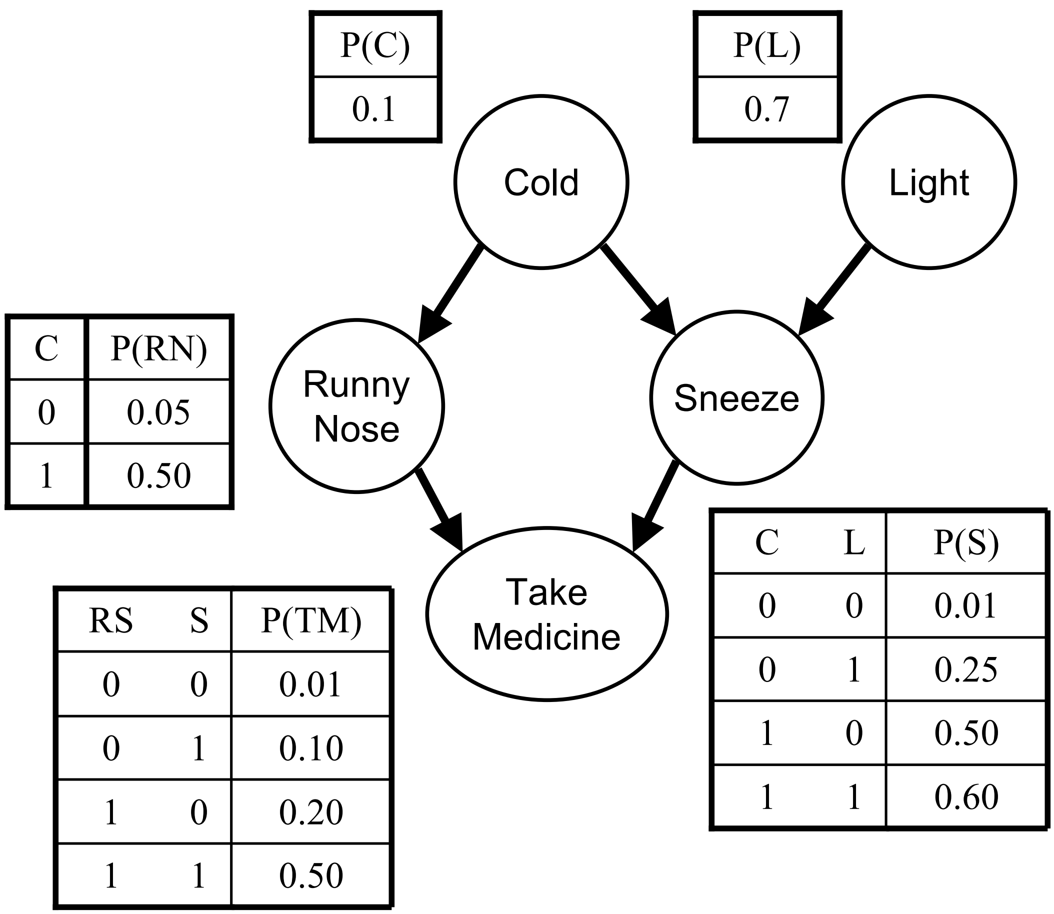
   (c) Let G be the graph of this Markov network. What is the Markov blanket of Light in G?

   (d) Is Light independent of Cold in G?

6. (Based on Koller & Friedman, 3.10.) One operation on Bayesian networks that arises in many settings is the marginalization of some node in the network.

   (a) Consider the Burglary Alarm network shown in Figure 3.23 of Koller & Friedman. Construct a Bayesian network over all of the nodes except for Alarm that is consistent with the any probability distribution over the network after marginalizing out Alarm. Your network should be the minimal consistent network, i.e., preserve as many independencies as possible while consistently representing *all* necessary dependencies.

   (b) Generalize the procedure you used to solve the above into a node-elimination algorithm. That is, define an algorithm that transforms the structure of G into G' such that one of the nodes X_i of G is not in G' and the independencies in G' are consistent with the marginal distribution over the remaining variables as defined in G.



P(C)
0.1

P(L)
0.7

C	P(RN)
0	0.05
1	0.50

C	L	P(S)
0	0	0.01
0	1	0.25
1	0	0.50
1	1	0.60

RS	S	P(TM)
0	0	0.01
0	1	0.10
1	0	0.20
1	1	0.50

Figure 1: Bayesian network for Problem 5.