Inference and Representation, Fall 2015

Problem Set 1: Bayesian networks

Due: Tuesday, September 22, 2015 at 3pm (as a PDF document sent to pg1338@nyu.edu)

Important: *See problem set policy on the course web site.*

---

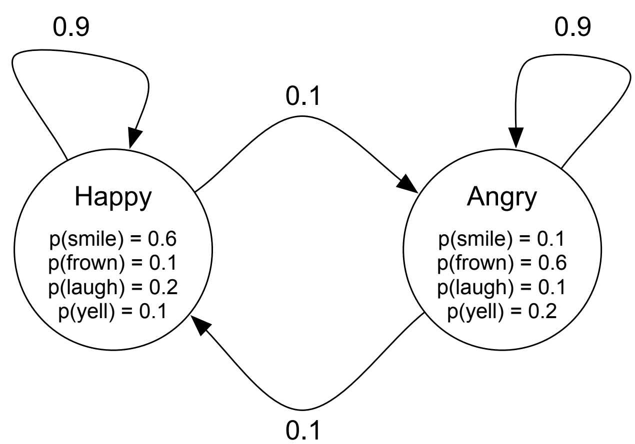
1. You go for your yearly checkup and have several lab tests performed. A week later your doctor calls you and says she has good and bad news. The bad news is that you tested positive for a marker of a serious disease, and that the test is 99% accurate (i.e. the probability of testing positive given that you have the disease is 0.99, as is the probability of testing negative given that you don't have the disease). The good news is that this is a rare disease, striking only 1 in 25,000 people. Why is it good news that the disease is rare? What are the chances that you actually have the disease?

2. **Hidden Markov models**. Harry lives a simple life. Some days he is Angry and some days he is Happy. But he hides his emotional state, and so all we can observe is whether he smiles, frowns, laughs, or yells. Harry's best friend is utterly confused about whether Harry is actually happy or angry and decides to model his emotional state using a hidden Markov model.

   Let $X_d \in \{\text{Happy, Angry}\}$ denote Harry's emotional state on day d, and let $Y_d \in \{\text{smile, frown, laugh, yell}\}$ denote the observation made about Harry on day d. **Assume that on day 1 Harry is in the Happy state**, i.e. $X_1 = \text{Happy}$. Furthermore, assume that Harry transitions between states exactly once per day (staying in the same state is an option) according to the following distribution: $p(X_{d+1} = \text{Happy} \mid X_d = \text{Angry}) = 0.1$, $p(X_{d+1} = \text{Angry} \mid X_d = \text{Happy}) = 0.1$, $p(X_{d+1} = \text{Angry} \mid X_d = \text{Angry}) = 0.9$, and $p(X_{d+1} = \text{Happy} \mid X_d = \text{Happy}) = 0.9$.

   The observation distribution for Harry's Happy state is given by $p(Y_d = \text{smile} \mid X_d = \text{Happy}) = 0.6, p(Y_d = \text{frown} \mid X_d = \text{Happy}) = 0.1, p(Y_d = \text{laugh} \mid X_d = \text{Happy}) = 0.2$, and $p(Y_d = \text{yell} \mid X_d = \text{Happy}) = 0.1$. The observation distribution for Harry's Angry state is $p(Y_d = \text{smile} \mid X_d = \text{Angry}) = 0.1, p(Y_d = \text{frown} \mid X_d = \text{Angry}) = 0.6, p(Y_d = \text{laugh} \mid X_d = \text{Angry}) = 0.1$, and $p(Y_d = \text{yell} \mid X_d = \text{Angry}) = 0.2$.

   All of this is summarized in the following figure:



Be sure to show all of your work for the below questions. Note, the goal of this question is to get you to start thinking deeply about probabilistic inference. Thus, although you could look at Chapter 17 for an overview of HMMs, try to solve this question based on first principles (also: no programming needed!).

(a) What is $p(X_2 = \text{Happy})$?

(b) What is $p(Y_2 = \text{frown})$?

(c) What is $p(X_2 = \text{Happy} \mid Y_2 = \text{frown})$?

(d) What is $p(Y_{80} = \text{yell})$?

(e) Assume that $Y_1 = Y_2 = Y_3 = Y_4 = Y_5 = \text{frown}$. What is the most likely sequence of the states? That is, compute the MAP assignment $\arg\max_{x_1,\ldots,x_5} p(X_1 = x_1, \ldots, X_5 = x_5 \mid Y_1 = Y_2 = Y_3 = Y_4 = Y_5 = \text{frown})$.

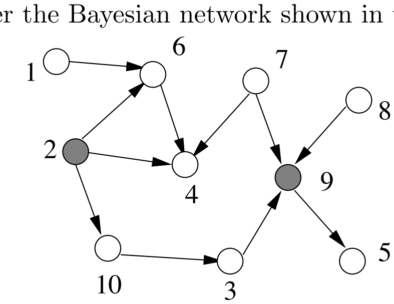
3. **Bayesian networks must be acyclic.** Suppose we have a graph $\mathcal{G} = (V, E)$ and discrete random variables $X_1, \ldots, X_n$, and define

$$f(x_1, \ldots, x_n) = \prod_{v \in V} f_v(x_v | x_{pa(v)}),$$

where $pa(v)$ refers to the parents of variable X_v in $\mathcal{G}$ and $f_v(x_v \mid x_{pa(v)})$ specifies a distribution over X_v for every assignment to X_v's parents, i.e. $0 \leq f_v(x_v \mid x_{pa(v)}) \leq 1$ for all $x_v \in \text{Vals}(X_v)$ and $\sum_{x_v \in \text{Vals}(X_v)} f_v(x_v \mid x_{pa(v)}) = 1$. Recall that this is precisely the definition of the joint probability distribution associated with the Bayesian network $\mathcal{G}$, where the f_v are the conditional probability distributions.

Show that if $\mathcal{G}$ has a directed cycle, f may no longer define a valid probability distribution. In particular, give an example of a cyclic graph $\mathcal{G}$ and distributions f_v such that $\sum_{x_1,\ldots,x_n} f(x_1, \ldots, x_n) \neq 1$. (A valid probability distribution must be non-negative and sum to one.) This is why Bayesian networks must be defined on *acyclic* graphs.

4. **D-separation.** Consider the Bayesian network shown in the below figure:



(a) For what pairs (i, j) does the statement $X_i \perp X_j$ hold? (Do not assume any conditioning in this part.)

(b) Suppose that we condition on $\{X_2, X_9\}$, shown shaded in the graph. What is the largest set A for which the statement $X_1 \perp X_A \mid \{X_2, X_9\}$ holds? The Bayes ball algorithm for d-separation given in Section 10.5.1 of Murphy's book may be helpful.

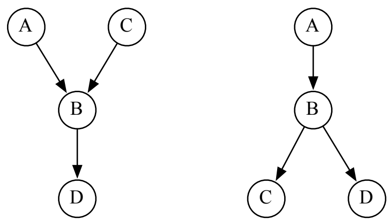
(c) What is the largest set B for which $X_8 \perp X_B \mid \{X_2, X_9\}$ holds?

5. Consider the following distribution over 3 binary variables X, Y, Z:

$$p(x,y,z) = \begin{cases} 1/12 & x \oplus y \oplus z = 0 \\ 1/6 & x \oplus y \oplus z = 1 \end{cases}$$

where $\oplus$ denotes the XOR function. Show that there is no directed acyclic graph G such that $I_{d-sep}(G) = I(p)$.

6. Consider the following two networks:



(a) (b)

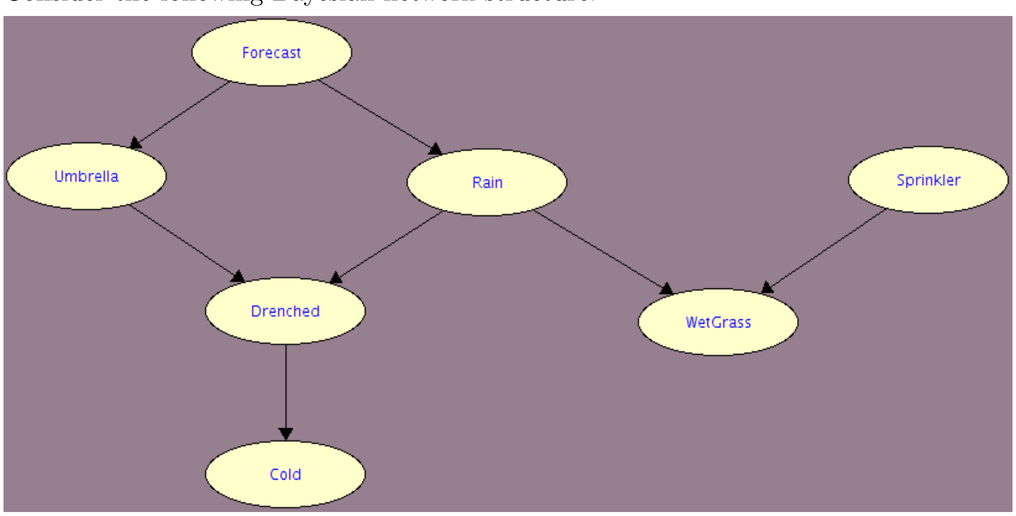
Two networks G, G' are *I-equivalent* if their structures encode exactly the same independence statements, i.e. $I(G) = I(G')$. For each of the above two networks, determine whether there can be any other Bayesian network that is I-equivalent to it. Justify your answer, listing all of the I-equivalent Bayesian networks (if any).

7. For this question, you will use **SamIam: Sensitivity Analysis, Modeling, Inference and More**, which is a free tool for inference using graphical models. The installation files, installation instructions and documentation are available at:

`http://reasoning.cs.ucla.edu/samiam/`

Note: On some systems, when you first launch SamIam, you may not be able to create new files or load existing network files. Follow these steps to resolve the issue:
- Under the Preferences menu, go to Preferences.
- Go to Global tab.
- Change 'User interface look and feel' to 'Nimbus'.

Consider the following Bayesian network structure:



The description of the nodes is given in the following table:

Node	**Values**	**Description**
Forecast	{WillRain, WillNotRain}	Prediction of whether it will rain or not
Umbrella	{Yes, No}	Whether you are carrying an umbrella along or not
Rain	{Yes, No}	Whether it actually rains or not
Sprinkler	{On, Off}	Whether the sprinkler was On or Off last night
Drenched	{Yes, No}	Whether you get drenched in the rain
WetGrass	{Yes, No}	Whether the grass is wet or not
Cold	{Yes, No}	Whether you catch cold or not

For the above network, answer the following questions using SamIam. You should use the "Shenoy-Shafer" inference algorithm, which performs exact inference. It's a good idea to set monitors on all of the variables (these show you the nodes' marginal probabilities), so that you can watch the effect on all of the variables of any changes that you make.

(a) After inputing the above network into SamIam, construct conditional probability distribution (CPD) tables for each of the nodes, assigning probabilities to events that agree with intuition. Include a print out of all of the CPD tables in your solutions, and use these CPD tables to answer the below questions.

(b) Perturb the CPD of `Forecast` and observe how the marginal probability of `Cold` changes. Give the initial and final CPDs of `Forecast` and the corresponding marginal probabilities for `Cold`. Explain your observations intuitively.

(c) Given that you observe `Cold`=Yes, how does the probability of `Rain` change (that is, compare P(Rain|Cold=Yes) with P(Rain))? Give the initial and final probability distributions of rain (P(Rain) and P(Rain|Cold=Yes)). Explain intuitively.

(d) Given that you observe that the grass is wet, how does the probability of cold change? What about if the grass is observed *not* to be wet? Give the initial and final marginal probabilities for `Cold` in both cases. Explain the change intuitively. Mention the active trail(s) which allow flow of inference in this scenario, i.e. the path(s) that the ball travels to get from `Cold` to `WetGrass` using the Bayes Ball algorithm.

(e) Given that we observe evidence that the sprinkler was off, how does the probability of cold change? What if, in addition, we observe that the grass is wet? Give the initial and final marginal distributions for `Cold` in each case. Explain your observations using terminology of Bayesian networks. Also give an intuitive explanation.

(f) Modify your CPTs to represent an almost perfect forecast (that is, P(Rain=Yes |Forecast=WillRain) $\approx$ 1.0, P(Rain=No | Forecast = WillNotRain) $\approx$ 1.0, P(Umbrella=Yes | Forecast=WillRain) $\approx$ 1.0). How does this modification change the relation between `Rain` and `Cold`? Explain intuitively. Make sure you test your theory by setting evidence on `Rain` and seeing its influence on `Cold`.