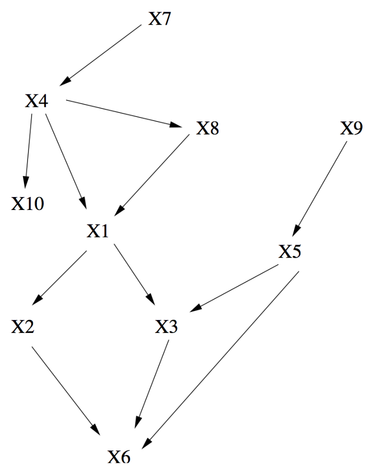
Probabilistic Graphical Models, Spring 2013

Problem Set 2: Bayesian networks

Due: Thursday, February 14, 2013 at 5pm

---

1. **Markov blanket.** Let $\mathcal{X} = \{X_1, ..., X_n\}$ be a set of random variables with distribution p given by the following graph.



   (a) Consider the variable X_1. What is the minimal subset of the variables, $A \subseteq \mathcal{X} - \{X_1\}$, such that $(X_1 \perp \mathcal{X} - A - \{X_1\} | A)$? Justify your answer.

   (b) Now, generalize this to any BN defined by (G, p). Specifically, consider variable X_i. What is the *Markov blanket* of X_i? Namely, the minimal subset of variables $A \subseteq \mathcal{X} - \{X_i\}$ such that $(X_i \perp \mathcal{X} - A - \{X_i\} \mid A)$? Prove that this subset is necessary and sufficient.

   (Hint: Think about the variables that X_i cannot possibly be conditionally independent of, and then think some more).

2. 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 $I_{d-sep}(G) = I(p)$.

3. Exercise 3.11 from Koller & Friedman.
   In addition: (c) What happens if the algorithm that you gave in part (b) is used to remove the class variable in the naive Bayes model?

Hint: Recall that the minimal I-map for a distribution is not necessarily unique, and some may have more edges than others (minimal here does not mean the fewest number of edges). Let $I_{-A}(G) \subseteq I(G)$ denote all independence statements that do not involve variable A. Formally, we are looking for a graph G_{new} such that $I(G_{new}) \subseteq I_{-A}(G)$ and such that removing any edge e from G_{new} would make $I(G_{new} \backslash e) \nsubseteq I_{-A}(G)$.

4. Exercise 3.15 from Koller & Friedman. Justify your answer, and list all of the I-equivalent Bayesian networks (if any).

5. Exercise 3.2 from Koller & Friedman.

6. Consider the Markov model given by $X_1 \rightarrow X_2 \rightarrow \ldots \rightarrow X_{n-1} \rightarrow X_n$, where $X_i \in \{0, 1\}$. The distributions $p(X_1), p(X_2 \mid X_1), \ldots, p(X_n \mid X_{n-1})$ are provided to us as tables.

   (a) Give an algorithm to compute $p(X_i = 1)$ for all $i = 1 \ldots n$

   (b) Give an algorithm to compute $p(X_i = 1 \mid X_1 = 1)$ for all $i = 2 \ldots n$

   (c) Give an algorithm to compute $p(X_1 = 1 \mid X_i = 1)$ for all $i = 2 \ldots n$
   (Hint: combine the results of (a) and (b))

   All algorithms should have a running time that is $O(n)$. This is our first example of a probabilistic inference algorithm! Notice how you were able to take advantage of the graphical model structure to come up with a more efficient algorithm than naive marginalization.