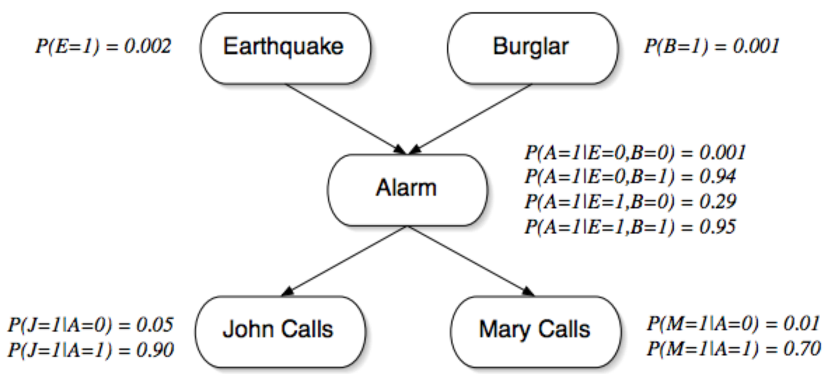
CSE 250A Assignment 2

This assignment is due at the start of class on Thursday October 18, 2012. Instructions are the same as for the first assignment. You must work again in partnership with one other student, but you may keep the same partner or change partners, as you wish. *Acknowledgment: These questions are adapted from ones written by Lawrence Saul.*

1. Reasoning with a Bayesian network

Consider the following extended version of a network that we have seen in class (drawing by Lawrence Saul):



Now consider the following conditional probabilities:

1. $p(M=1)$
2. $p(M=1|E=1)$
3. $p(M=1|E=1, B=1)$
4. $p(M=1|A=1)$
5. $p(M=1|A=1, B=1)$
6. $p(M=1|A=1, B=1, E=1)$
7. $p(M=1|J=1)$
8. $p(M=1|J=1, A=1)$
9. $p(M=1|J=1, A=1, B=1)$

(a) Which probabilities above, if any, must be equal to each other?
(b) Compute the numerical value of each probability above, exploiting conditional independence relationships as much as possible to simplify your calculations. Show your work.

2. Logistic regression

Let the binary random variable Y depend on real-valued random variables X_i as

$$p(Y = 1|X_1 = x_1, \ldots, X_k = x_k) = \sigma(\sum_{i=1}^{k} w_i x_i)$$

where $\sigma(z) = 1/(1 + e^{-z})$. The real-valued parameters w_i in this CPT are often called weights.

(a) *Revised for clarity.* If X_i in the logistic regression equation above is a binary random variable, that is a special case of it being real-valued, because its outcomes can be encoded as $X_i = 0$ and $X_i = 1$. Suppose that X_i is discrete with three or more outcomes, say { elephant, donkey, whale }. Does there also exist an encoding that makes it be a special case of a real-valued random variable?

(b) Sketch the sigmoid function $\sigma(z)$. Show that $\sigma(-z) + \sigma(z) = 1$ for all z and that the derivative $(d/dz)\sigma(z) = \sigma(z)\sigma(-z)$.

(c) The function $L(p) = \log \frac{p}{1-p}$ is called the log odds function. Show that $L(\sigma(z)) = z$.

(d) Given a training example $(x_1, \ldots, x_k, y)$ with label $y \in \{0, 1\}$, in order to learn the weights w_i, often we want to maximize $\log q$ where $q = p(Y = y|X_1 = x_1, \ldots, X_k = x_k)$. We do the maximization using partial derivatives. Evaluate $(\partial/\partial w_i) \log q$. The answer should be simple and elegant.

3. Markov blankets

Let X be any node in a Bayesian network. The Markov blanket B_X of X consists of its parents, its children, and its spouses, which are the parents of its children excluding itself.

(a) Draw a general picture of a node and its Markov blanket. Then, using the three conditions for d-separation, prove that for any node $Y \neq X$ such that $Y \notin B_X$

$$p(X, Y|B_X) = P(X|B_X)P(Y|B_X).$$

(b) Explain the name "Markov blanket." Why the word "Markov" and why the word "blanket"?