CMPUT 651 — Assignment 1
Revised 29/Sept/08 (fixed typo)

Instructor: R. Greiner; M. Brown
Due Date: Friday, 10 Oct 2008 at 5pm

The following exercises are intended to further your understanding of belief networks — semantics, inference, learning.
Of course, be sure to explain your answers, etc etc etc.

Total points: 90 + 35 = 125

Submission: You should hand-in hardcopies of Questions 1 to 9.
For Question 10: email a gzipped tar file to `amir@cs.ualberta.ca`.

Question 1 *[12 points]* Prove or disprove (by providing a counter-example) each of the following claim about independence:

1. $(X \perp Y, W | Z)$ implies $(X \perp Y | Z)$
2. $(X \perp Y | Z)$ and $(X \perp Y | W)$ implies $(X \perp Y | \{Z, W\})$
3. $(X \perp Y, W | Z)$ and $(Y \perp W | Z)$ imply $(X, W \perp Y | Z)$

Question 2 *[3 points]* Provide an example of a distribution $P(X_1, X_2, X_3)$ where for each $i \neq j$, we have that $(X_i \perp X_j) \in I(P)$, but we also have that $(X_1, X_2 \perp X_3) \notin I(P)$.

Question 3 *[10 points] from [Koller/Friedman:Exercise 2.9]*
This question investigates the way in which conditional independence relationships affect the amount of information needed for probabilistic calculations. Let H, E_1, E_2 be three random variables, and suppose we wish to calculate $P(\, H \mid E_1, E_2\,)$.

a [5]: Suppose we have no conditional independence information. Which of the following probability distributions (think "sets of probability values") are sufficient for the calculation?

1. $P(\, E_1, E_2\,)$, $P(\, H\,)$, $P(\, E_1 \mid H\,)$, and $P(\, E_2 \mid H\,)$
2. $P(\, E_1, E_2\,)$, $P(\, H\,)$, and $P(\, E_1, E_2 \mid H\,)$
3. $P(\, E_1 \mid H\,)$, $P(\, E_2 \mid H\,)$ and $P(\, H\,)$.

For each case, justify your response either by showing how to calculate the desired answer from the numbers given, or by explaining why this is not possible.

b [5]: Now suppose we know that E_1 and E_2 are conditionally independent given H. (Think Naïve Bayes.) Now which of the above three sets are sufficient? Justify your response as above.

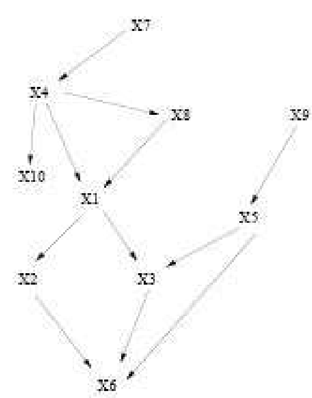
Question 4 *[10 points]* You are given a specific network over a set of n variables $\{X_1, \ldots, X_n\}$. Show how you can *efficiently* compute the distribution over a variable X_i given an assignment to *all* the other variables in the network: $P(X_i \mid x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$. Your procedure should *not* require the construction of the entire joint distribution $P(X_1, \ldots, X_n)$. Specify the computational complexity of your procedure.

Be sure to describe both the algorithms, and its complexity, in terms of the graph structure — *e.g.*, parents and children of X_i and perhaps other nodes. You may assume that each variable is discrete, and ranges over k values. **Added [20/Sept/08]:** You can retrieve the k parents of a node in $O(k)$ time, and the r children of a node in $O(r)$ time. Also: to get full marks, you need to *prove* that some specified set of nodes is sufficient.

Question 5 *[15 points]* *Representation Theory: Factorization $\Rightarrow$ I-map*
Let G be a Bayesian network graph over a set of random variables X and let P be a joint distribution over the same space. Show that if P factorizes according to G, then G is an I-map for P.

[Hint: See the example in LectureNotes (2-BeliefNet-Semantics.pdf), and one in [Koller/Friedman:Section 3.2.3.3].]

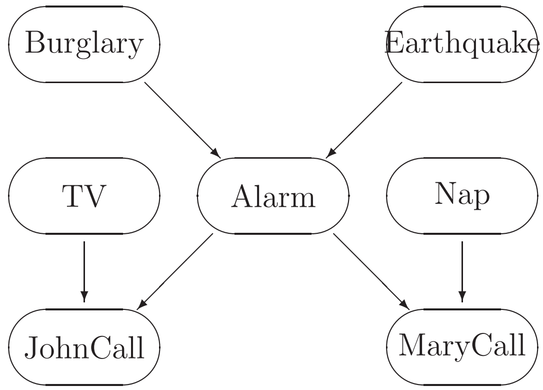
Question 6 *[5 points]* *Graph Independencies*
Let $\mathcal{X} = \{X_1, \ldots, X_{10}\}$ be a set of random variables, whose distribution is given by the following graphical model.



What is the minimal subset of the variables $\mathbf{A} \subset \mathcal{X} - \{X_1\}$ such that X_1 is independent of the rest of the variables, $\mathcal{X} - \mathbf{A} \cup \{X_1\}$, given $\mathbf{A}$?

Question 7 *[10 points]* *Marginalization*

a [3]: Consider the following `BurglarAlarm` network:



Construct a Bayesian network over all of the nodes except for `Alarm`, that is a minimal I-map for the marginal distribution over those variables defined by the above network. Be sure to include all (and only) the dependencies that remain from the original network.

b [7]: 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 of G is not in G' and G' is an I-map of the marginal distribution over the remaining variables, as defined by G.

Question 8 *[10 points]* *Belief Net gradient*

We can use a given Belief Net, with CPtable entries Θ, to compute the answer to some specific query – eg, p_Θ(cancer=true|gender=Male, headache=True) = 0.04. If we change the value of one CPtable entry by some amount (say changing $\theta_{smoke=true|gender=Male}$ from 0.3 to 0.32) this P(Cancer=true | Gender=male, Headache=true) value may change. Here, we are investigating how much – *e.g.*, given this new Θ' (differing from Θ only in this $\theta_{Smoke=true|Gender=male}$ value), will the value of P_Θ(Cancer=true|Gender=male, headache=true) remain 0.04, or change slightly – perhaps to 0.045 – or change a lot – perhaps to 0.999, or whatever.

To be more precise: Consider answering a query of the form $P(\, A = a \,|\, \mathbf{B} = \mathbf{b}\,)$, from a given belief net. Let the CPtable entry $\theta_{q|\mathbf{r}}$ designate the probability the belief net assigned to $Q = q$, given that Q's parents $\mathbf{R}$ collectively have the values $\mathbf{r}$. What is $\frac{\partial P(\, A=a \,|\, B=b\,)}{\partial \theta_{q|\mathbf{r}}}$? Your answer should be in terms of simple sums/products/quotients/... of probabilities; it should not involve $\sum$ summations, nor derivatives, ...

Here, you should assume that $\theta_{q|\mathbf{r}}$ is unrelated to $\theta_{q'|\mathbf{r}}$; *e.g.*, $\theta_{Smoke=true|Gender=Male}$ is unrelated to $\theta_{Smoke=false|Gender=Male}$.

[Hint: You may need the following information
1: $P(\, Z = z\,) = \sum_{q,r} P(\, Z = z, Q = q, R = r\,)$
2: $P(\, X = x \,|\, Y = y\,) = P(\, X = x, Y = y\,)/P(\, Y = y\,)$
3: $\frac{d\,(f(x)/g(x))}{d\,x} = \frac{gf' - fg'}{g^2}$ *]*

Question 9 *[15 points]* *Beta Distribution*

How many times do you need to flip a coin, until you see the first head. That is, what is $E[\,H_f\,] = \sum_r r \times P(\,H_f = r\,)$ where $H_f \;\in \{1, 2, \ldots\}$ is #flips until seeing first head.

Of course, this expected value depends on the coin. If you know the coin's head probability is 1/2, this is

$$E[\,H_f\,|\,\theta = 1/2] \quad = \quad \sum_{r=1}^{\infty} r \times P(\,f_r = h,\; f_1 = t, \ldots f_{r-1} = t\,) \quad = \quad \sum_{r=1}^{\infty} r \times \left(\frac{1}{2}\right)^r \quad = \quad 2$$

Here, f_i is the outcome of the i^{th} flip.

a [3]: What is $E[H_f\;|\theta = 1/3]$? $\ldots E[H_f\;|\theta = 1/4]$?

b [7]: Now suppose we know the coin is a "Beta(1, 1) coin" – that is, $\theta \sim \text{Beta}(\,1, 1\,)$. Here $P(\,f_1 = h\,) = 1/2$ (which is the expected value of θ), but $P(\,f_2 = h\,|\,f_1 = h\,) \neq 1/2$, as $\theta|\text{"1H"} \sim \text{Beta}(\,2, 1\,)$. (Of course, $\theta|\text{"2H,1T"} \sim \text{Beta}(\,3, 2\,)$, etc.) What is $E[\,H_f\,|\,\theta \sim \text{Beta}(\,1, 1\,)\,]$?

c [5]: If a coin has probability p of heads, then the chance of observing exactly r heads in k flips is $\binom{k}{r} p^r (1-p)^{k-r}$. What is the corresponding number if the coin's head probability is drawn from a Beta(1, 1) distribution – *i.e.*, $\theta \sim \text{Beta}(\,1, 1\,)$?

Question 10 *[35 points] Bayesian Network Inference*
As the owner of an online bookstore, you would like to implement a recommendation system for your customers. After poring over your records, you discover that you carry only four books. What's worse, you have only three customers. Even worse than that, you've only sold six books in the last year!

Clearly, this is a job for a Bayesian network. After thinking about the problem, you consider the two models shown in Figures 1 and 2.

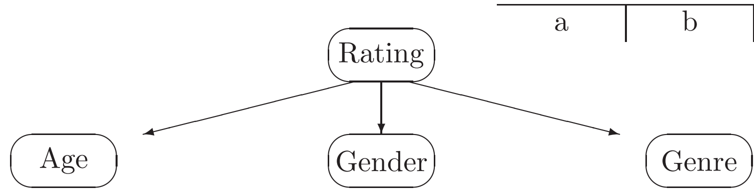
a [3]: *Conditional Independence*
Consider the elaborate model in Figure 2. If you already know r_2 and r_3, then what variables are influenced by revealing the value of r_1?

b [20]: *Inference*
In this question you will implement a representation of a general Bayesian network in MATLAB. Using this representation, implement a simple inference algorithm that iterates over all possible assignments of relevant variables.[1] We will later specify where to upload your code.

You should follow the following steps to implement this:
1. A data structure to represent a *factor*, as a mapping from an assignment of variables to a real value. Conditional probability tables can be viewed as factors. For example, in Figure 1, the conditional probability table for `Age` maps the assignment `(Rating = Dislikes, Age = Youth)` to the value c. The easiest way to encode a factor is as a multidimensional array where each dimension corresponds to a variable. See `table_factor.m`.
2. A data structure to represent a Bayesian network. The easiest way to do this is just to

[1] Do not implement variable elimination... that is overkill here! See Question 4 above.

Dislikes	Neutral	Likes
a	b	1-a-b



Rating	Youth	Adult
Dislikes	c	1-c
Neutral	d	1-d
Likes	e	1-e

Rating	Male	Female
Dislikes	f	1-f
Neutral	g	1-g
Likes	h	1-h

Rating	Self-Help	Cookbook
Dislikes	i	1-i
Neutral	j	1-j
Likes	k	1-k

Figure 1: A Naive Bayes model of recommendation, including parameters. Let $\Theta_{NB} = (a, b, \ldots, k)$ denote all the parameters in this model.

store a list of all the conditional probability tables as factors.
3. A data structure that represent an assignment to variables. The easiest way to do this is as a pair of vectors, `vars` and `vals`, where `vals(i)` is the value assigned to variable `vars(i)`. See `assignment.m`.
4. A function that takes a Bayesian network and an assignment to all the variables, and returns the probability of that assignment.
5. A *marginalization* function that iterates over all assignments to a set of variables, and accumulates the value of the joint distribution at these assignments.

Note: You will not receive full marks for an implementation that stores the full joint distribution explicitly.

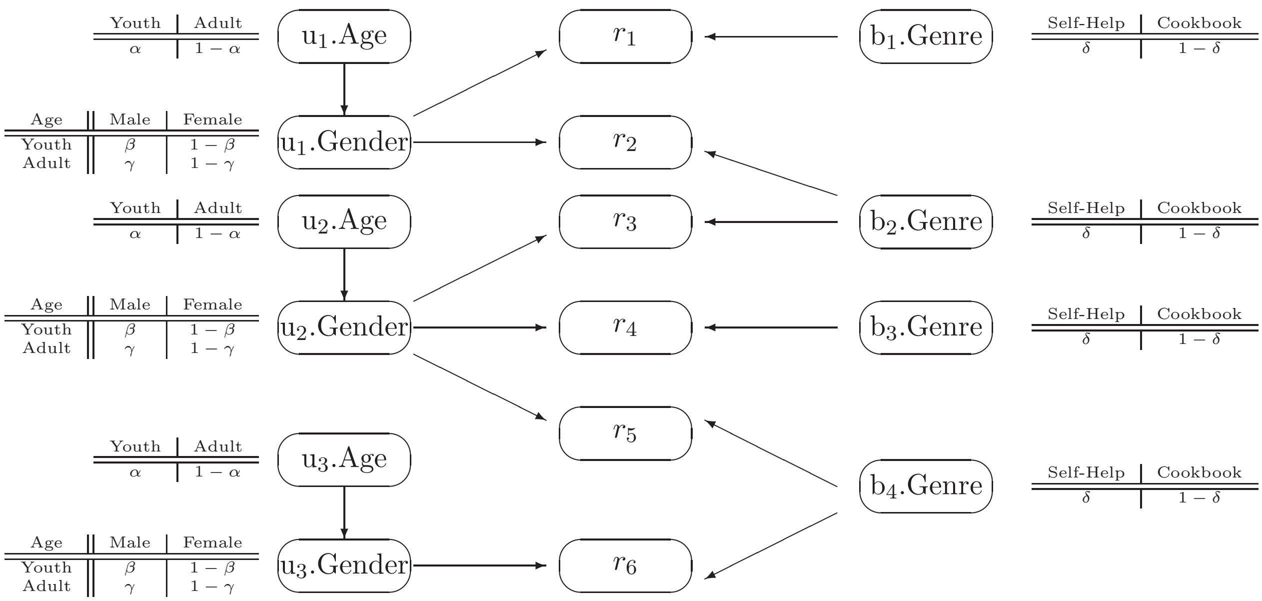
Naïve Bayes: You are given a parameterization for the Naïve Bayes model shown in Figure 1, with a = 0.15, b = 0.20, c = 0.35, d = 0.08, e = 0.45, f = 0.30, g = 0.1, h = 0.52, i = 0.7, j = 0.09, k = 0.65. Using your implementation of inference, what are the values of the following queries?

1. $P($`Rating = Likes`$|$`Age = Youth, Genre = Cookbook`$)$
2. $P($`Genre = Cookbook`$|$`Gender = Male`$)$
3. $P($`Age = Adult`$|$`Genre = Self-Help`$)$

Record the values of the above queries in your writeup, and submit your code as `infnb.m`.

Elaborate Model: You are given a parameterization for the elaborate model of recommendation, see Figure 2. Set $\alpha = 0.3$, $\beta = 0.6$, $\gamma = 0.46$, $\delta = 0.25$, $\epsilon = 0.35$, $\zeta = 0.21$, $\nu = 0.25$, $\theta = 0.15$, $\iota = 0.06$, $\kappa = 0.18$, $\lambda = 0.04$, $\mu = 0.11$. Using your implementation of inference, what are the values of the following queries?

1. $P($`u`$_1$`.Age = Youth`$|$`r`$_1$` = Likes, r`$_2$` = Likes, r`$_3$` = Dislikes,r`$_4$` = Likes, b`$_4$`.Genre = Cookbook`$)$



For each r_k node:

Gender	Genre	Dislikes	Neutral	Likes
Male	Self-Help	ϵ	ζ	$1-\epsilon-\zeta$
Male	Cookbook	ν	θ	$1-\nu-\theta$
Female	Self-Help	ι	κ	$1-\iota-\kappa$
Female	Cookbook	λ	μ	$1-\lambda-\mu$

Figure 2: An elaborate model of recommendation. Customers are labelled u_i, books are labelled b_j, ratings are labelled r_k. There is a variable for each user's age and gender, each book's genre, and each rating. Seeing that the parents of r_k are u_i and b_j means that user u_i bought book b_j and assigned it the rating r_k. Note that nodes can share the same conditional probability tables — *e.g.*, all the `Age` nodes share the same parameters. Let $\Theta_{EL} = \{\alpha, \beta, \ldots, \lambda, \mu\}$ denote all the parameters in this model.

2. $P($`u`$_2$`.Gender = Male, `u$_2$`.Age = Youth`$\,|\,$`r`$_1$` = Likes, `r$_6$` = Dislikes, ...`
   `b`$_2$`.Genre = Cookbook, `b$_3$`.Genre = Self-Help, `b$_4$`.Genre = Cookbook `$)$

3. $P($`r`$_3$` = Likes`$\,|\,$`r`$_1$` = Dislikes, `r$_2$` = Likes, `r$_4$` = Neutral, `r$_5$` = Neutral, `r$_6$` = Likes`$)$

Record the values of the above queries in your writeup, and submit your code as `infel.m`.

c [12]: *Parameter estimation*
You have cleverly negotiated a deal with a large book retailer to share sales data. You then received a list of records of the form

Age	Gender	Genre	Rating
Youth	Male	Cookbook	Likes
Adult	Female	Self-Help	Dislikes
...	...	...	...

for a large number of users; see `purchase.csv`. Look at the provided function `loaddata.m` for how to load this data into MATLAB.

Naïve Bayes Using this data, implement a function that computes the maximum likelihood estimate for Θ_{NB}. Record the estimates of the parameters in your writeup and submit your code as `mlenb.m`.

Elaborate Model Using this data, implement a function that computes the maximum likelihood estimate for Θ_{EL}. Record the estimates of the parameters in your writeup and submit your code as `mleel.m`.

Hint: In the Naïve Bayes model, the variables are `Age`, `Gender`, `Genre` and `Rating`. So estimating a probability translates into counting records that match a particular assignment to these variables. In the elaborate model there is a variable for each attribute of every entity `(user, book, rating)`. However, the data in `purchase.csv` only contains `Age`, `Gender`, `Genre`, and `Rating` attributes. The way we solved this problem is to tie parameters together -- all the `Age` nodes share the same CPT; all the `Gender` nodes share the same CPT; all the `Genre` nodes share the same CPT; and all the `Rating` nodes share the same CPT. The problem has been reduced to estimating conditional probability tables over the four variables. For example,

- α is the fraction of all records that have `Age = Youth`.
- β is the fraction of records with `Age = Youth` that also have `Gender = Male`.
- ϵ is the fraction of records with `Gender = Male` and `Genre = Self-Help` that also have `Rating = Dislikes`.
- ζ is the fraction of records with `Gender = Male` and `Genre = Self-Help` that also have `Rating = Neutral`.