10708 Graphical Models: Homework 1

Due September 27th, beginning of class

September 14, 2006

Instructions: There are five questions on this assignment. The last question involves coding, which should be done in MATLAB. Do *not* attach your code to the writeup. Instead, copy your implementation to

`/afs/andrew.cmu.edu/course/10/708/Submit/your_andrew_id/HW1`

Refer to the web page for policies regarding collaboration, due dates, and extensions.

1 [15 pts] Conditional Probability

1.1 [5 pts]

Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be three disjoint sets of variables such that $\mathcal{S} = \mathcal{X} \cup \mathcal{Y} \cup \mathcal{Z}$. Prove that $P \models (\mathcal{X} \perp \mathcal{Y} | \mathcal{Z})$ if and only if we can write P in the form: $P(\mathcal{S}) = f(\mathcal{X}, \mathcal{Z}) g(\mathcal{Y}, \mathcal{Z})$

1.2 [3 pt]

Is it possible for both f and g above to be probability distributions over their respective sets of variables? Formally, is it possible for every distribution P over $(\mathcal{X} \cup \mathcal{Y} \cup \mathcal{Z})$ with the independency above, to be expressed as a product of a distribution over $(\mathcal{X} \cup \mathcal{Z})$ and a distribution over $(\mathcal{Y} \cup \mathcal{Z})$? Justify your answer.
(*Hint*: look at the marginal probability of $\mathcal{Z}$; you may assume that the variables are binary if you wish.)

1.3 [3 pts]

Prove or disprove (by providing a counter-example) each of the following properties of independence:

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

1.4 [4 pt]

Provide an example of a distribution $P(X_1, X_2, X_3)$ where for each $i \neq j$, we have that $(X_i \perp X_j) \in \mathcal{I}(P)$, but we also have that $(X_1, X_2 \perp X_3) \notin \mathcal{I}(P)$.

2 [20 pts] Graph Independencies

2.1 [5 pts]

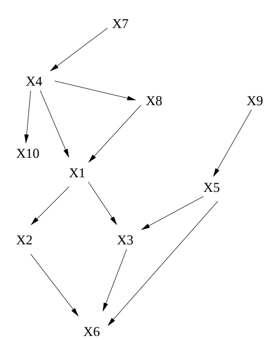

Figure 1: Graphical Model for Prob. 2

Let $\mathbf{X} = \{X_1, \ldots, X_n\}$ be a random vector with distribution given by the graphical model in Figure 1. Consider variable X_1. What is the minimal subset of the variables, $\mathbf{A} \subseteq \mathcal{X} - \{X_1\}$, such that X_1 is independent of the rest of the variables, $\mathcal{X} - \mathbf{A} \cup \{X_1\}$, given $\mathbf{A}$? Justify your answer.

2.2 [7 pts]

Now, let the distribution of $\mathbf{X}$ be given by some graphical model instance $\mathbf{B} = (\mathcal{G}, P)$. Consider variable X_i. What is the minimal subset of the variables, $\mathbf{A} \subseteq \mathcal{X} - \{X_i\}$, such that X_i is independent of the rest of the variables, $\mathcal{X} - \mathbf{A} \cup \{X_i\}$, given $\mathbf{A}$? Prove that this subset is necessary and sufficient.

2.3 [8 pts]

Show how you could efficiently compute the distribution over a variable X_i given some assignment to all the other variables in the network: $P(X_i | x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$.
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.

3 [15 pts] Factorization

Let $\mathcal{G}$ be a bayesian network graph over a set of random variables $\mathcal{X}$ and let P be a joint distribution over the same space. Show that if P factorizes according to $\mathcal{G}$, then $\mathcal{G}$ is an $I-$map for P.
(Hint: See example on page 86 of Koller and Friedman)

4 [15 pts] Marginalization

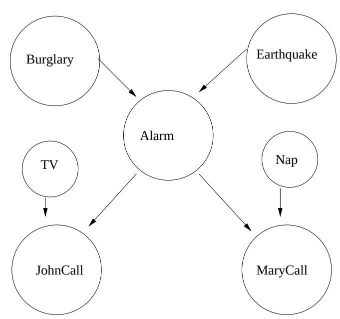

Figure 2: Burglar Alarm Network

1. Consider the *Burglar Alarm* network shown in Figure 2. Construct a Bayesian network over all of the nodes except for *Alarm*, which is a minimal $I-$map for the marginal distribution over those variables defined by the above network. Be sure to get all dependencies that remain from the original network.
2. Generalize the procedure you used to solve the above into a node-elimination algorithm. That is, define an algorithm that transforms the structure of $\mathcal{G}$ into $\mathcal{G}'$ such that one of the nodes X_ι of $\mathcal{G}$ is not in $\mathcal{G}'$ and $\mathcal{G}'$ is an $I-$map of the marginal distribution over the remaining variables as defined by $\mathcal{G}$.

5 [35 pts] Bayesian Network Inference

As the owner of an online bookstore, you would like to implement a recommendation system for your customers. After pouring 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 come up with the models in figures 3 and 4.

5.1 [3 pt] Conditional Independence

Consider the elaborate model in figure 4. If you already know r_2 and r_3, then what variables are influenced by revealing the value of r_1 ?

5.2 [20 pts] 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]. Upload all your code to

`/afs/andrew.cmu.edu/course/10/708/Submit/your_andrew_id/HW1`

Hint: The steps you should take in implementing this are as follows:

1. A data structure to represent a *factor*, a mapping from an assignment of variables to a real value. Conditional probability tables can be viewed as factors. For example, in figure 5, the conditional probability table for Age maps the assignment (Rating = Dislikes $\cup$ 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 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 $vals$. Note that $vals(i)$ 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.

[1] Do not implement variable elimination.

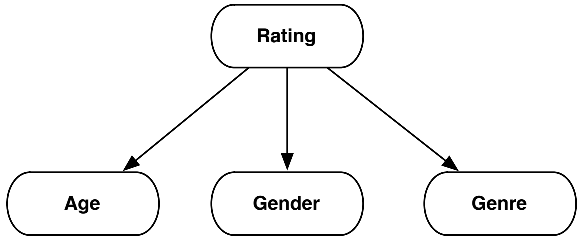

Figure 3: A Naïve Bayes model of recommendation.

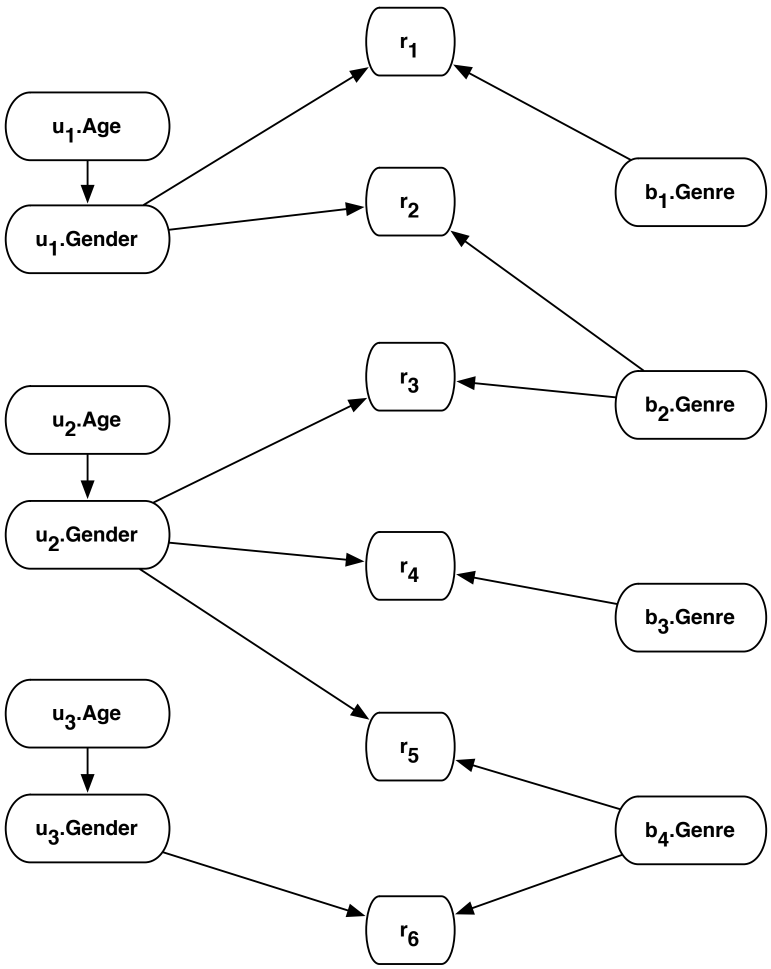

Figure 4: A elaborate model of recommendation. Customers are labelled u_i, books are labelled b_j, ratings are labelled r_k. There is a variable for each user's age and gender, each book's genre, and each rating. If the parents are r_k are u_i and b_j this means that user i bought book j and the rating he assigns to it is r_k.

5. A function that iterates over all assignments to a set of variables, and accumulates the value of the joint distribution at these assignments. This is marginalization.

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

5.2.1 Naïve Bayes

You are given a parameterization for the Naïve Bayes model in figure 5. Set $a = 0.15, b = 0.20, c = 0.35, d = 0.08, e = 0.45, f = 0.30, g = 0.1, h = 0.52, i = 0.7, j = 0.09, k = 0.65$. Using your implementation of inference, what are the values of the following queries ?

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

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

5.2.2 Elaborate Model

You are given a parameterization for the elaborate model of recommendation (figure 6). Set $\alpha = 0.3, \beta = 0.6, \gamma = 0.46, \delta = 0.25, \epsilon = 0.35, \zeta = 0.21, \eta = 0.25, \theta = 0.15, \iota = 0.06, \kappa = 0.18, \lambda = 0.04, \mu = 0.11$. Using your implementation of inference, what are the values of the following queries ?

1. $P(u_1.\text{Age} = \text{Youth}|r_1 = \text{Likes}, r_2 = \text{Likes}, r_3 = \text{Dislikes}, r_4 = \text{Likes}, b_4.\text{Genre} = \text{Cookbook})$
2. $P(u_2.\text{Gender} = \text{Male}, u_2.\text{Age} = \text{Youth}|r_1 = \text{Likes}, r_6 = \text{Dislikes}, \ldots$
   $b_2.\text{Genre} = \text{Cookbook}, b_3.\text{Genre} = \text{Self-Help}, b_4.\text{Genre} = \text{Cookbook})$
3. $P(r_3 = \text{Likes}|r_1 = \text{Dislikes}, r_2 = \text{Likes}, r_4 = \text{Neutral}, r_5 = \text{Neutral}, r_6 = \text{Likes})$

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

5.3 [12 pts] Parameter estimation

You have cleverly negotiated a deal with a large book retailer to share sales data. A list of records of the form

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

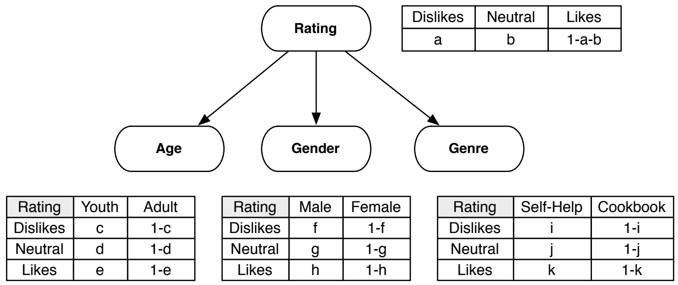

Figure 5: Parameters for the naïve Bayes model. Let $\Theta_{NB} = (a, b, \ldots, j, k)$ denote all the parameters in this model.

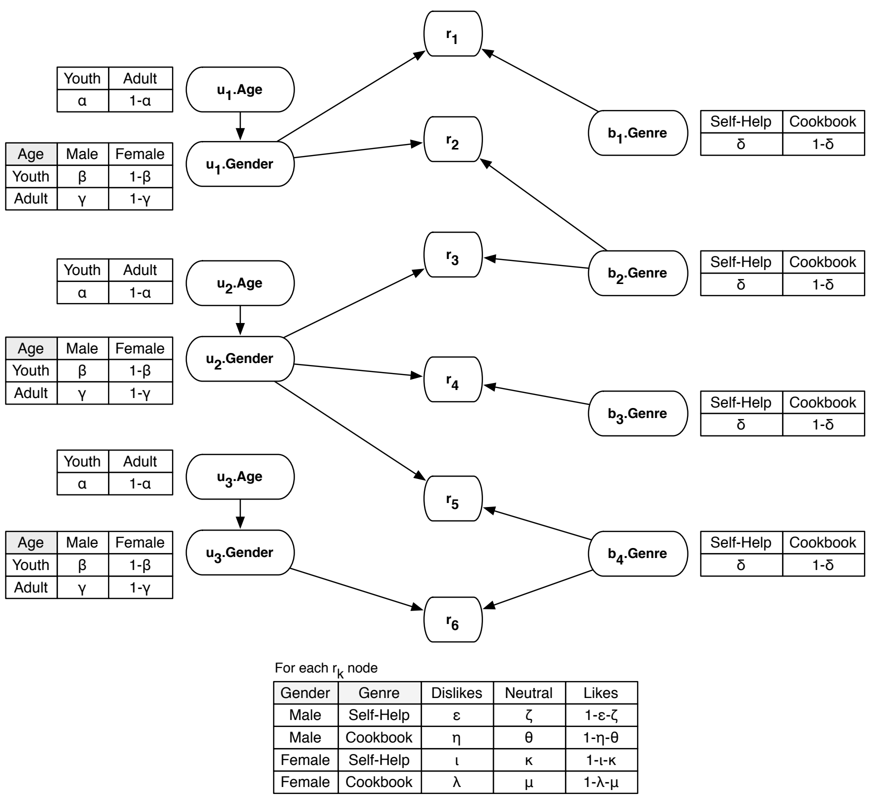

Figure 6: Parameters for the model. Note that nodes can share the same conditional probability table, *e.g.*, all the Age nodes share the same parameters. Let $\Theta_{EL} = (\alpha, \beta \ldots \lambda, \mu)$ denote all the parameters in this model

for a large number of users has been delivered (the data is located in `purchase.csv`). Look at the provided function `loaddata.m` for how to load this data into MATLAB.

5.3.1 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`.

5.3.2 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.