Probablistic Graphical Models, Spring 2007
Homework 2

Due at the beginning of class on 10/17/07

Instructions

There are six questions in this homework. The last question involves some programming which should be done in MATLAB. Do *not* attach your code to the writeup. Make a tarball of your code and output, name it <userid>.tgz where your userid is your CS or Andrew id, and copy it to /afs/cs/academic/class/10708-f07/hw2/<userid> .

If you are not submitting this from an SCS machine, you might need to authenticate yourself first. See `http://www.cs.cmu.edu/~help/afs/cross_realm.html` for instructions. If you are not a CMU student and don't have a CS or Andrew id, email your submission to 10708-07-instr@cs.cmu.edu.

You are allowed to use any of (and only) the material distributed in class for the homeworks. This includes the slides and the handouts given in the class[1]. Refer to the web page for policies regarding collaboration, due dates, and extensions.

1 [20 pts] Markov Networks

We define the following properties for a set of conditional independencies

- **Strong Union:**
$$(X \perp Y | Z) \Rightarrow (X \perp Y | Z, W)$$
In other words, additional evidence cannot induce dependence.

- **Transitivity:** For all disjoint X, Y, Z and variable A,
$$\neg(X \perp A | Z) \& \neg(A \perp Y | Z) \Rightarrow \neg(X \perp Y | Z)$$
Intuitively, this statement asserts that if X are both correlated with A (given Z), then they are also correlated with each other(given Z).

1. Construct a simple BN $\mathcal{G}$ such that I($\mathcal{G}$) does not satisfy both of the above properties.

2. Prove that if I($\mathcal{H}$) is the set of independencies of a Markov Network $\mathcal{H}$, then I($\mathcal{H}$) satisfies the above properties.

3. Let $I_l(\mathcal{H}) = \{(X \perp \mathcal{X} - X - \mathcal{N}_{\mathcal{H}}(X) | \mathcal{N}_{\mathcal{H}}(X)) : X \in \mathcal{X}\}$, where $\mathcal{N}_{\mathcal{H}}(X)$ is the set of neighbors of X in $\mathcal{H}$, be the set of *local* Markov independencies associated with $\mathcal{H}$ and $I_p(\mathcal{H}) = \{(X \perp Y | \mathcal{X} - \{X, Y\}) : X - Y \notin \mathcal{H}\}$ be the set of *pairwise* Markov independencies associated with $\mathcal{H}$. Prove the following assertion about Markov Networks: $I_l(\mathcal{H}) \Rightarrow I_p(\mathcal{H})$.

[1] Please contact Monica Hopes(meh@cs)if you need a copy of a handout.

2 [10 pts] Constructing Junction Tree

To construct a Junction Tree from a chordal graph $\mathcal{H}$, we build an undirected graph whose nodes are the maximal cliques in $\mathcal{H}$ where every pair of nodes C_i, C_j is connected by an edge whose weight is $|C_i \cap C_j|$. We then use the maximum weight spanning tree algorithm to find a tree in this graph whose weight i.e. the sum of the weights of the edges in the graph is maximal. (See Section 10.4.2 of the handout from Koller and Friedman for more details)

Show that the Junction tree constructed using the maximum weight spanning tree procedure outlined above satisfies the running intersection property.

3 [15 pts] Modifying Junction Tree

Let $\mathcal{T}$ be a calibrated junction tree representing the unnormalized distribution $P_{\mathcal{F}} = \prod_{\phi \in \mathcal{F}} \phi$. Let ϕ' be a new factor and let $P_{\mathcal{F}'} = P_{\mathcal{F}} \times \phi'$. Let C_i be some clique such that $Scope[\phi] \in C_i$. Show that we can obtain $P'_{\mathcal{F}}$ for any clique C_j by multiypling ϕ' into π_i and then propagating messages from C_i to C_j along the path between them.

4 [10 pts]Distributive Laws

We saw in class that the sum-product and max-product versions of BP(and of Shafer-Shenoy) are virtually identical with the exception of a summation in sum-product being replaced by a maximization in max-product(leading to their names).

Now suppose, we wish to compute the *least* probable estimates instead of the most probable estimates; i.e. we wish to compute $\min_x p(x)$

Show that a similar modification lets you compute this value. You need to prove the correctness of your algorithm (you may use the correctness of max-product in your proof).

5 [15 pts]Dirichlet-multinomial prediction

Let $\theta \sim Dir(\alpha)$. Consider multinomial random variables $(X_1, X_2, ..., X_N)$, where $X_n \sim Mult(\theta)$ for each n, and where the X_n are assumed conditionally independent given θ. Now consider a random variable $X_{new} \sim Mult(\theta)$ that is assumed conditionally independent of $(X_1, X_2, ..., X_N)$ given θ.

Compute $P(x_{new}|x_1, x_2, ..., x_N)$, by integrating over θ

6 [30 pts] Linear Regression

The data file lms.dat contains twenty rows each with three columns of numbers. The first two columns are the components of an input vector x and the last column is an output value y.

1. Solve the normal equations for this data to find the optimal value of the parameter vector.
2. Find eigenvectors and eigenvalues of the covariance matrix of the input vectors and plot the contours of the cost function J in the parameter space.
3. Initializing the LMS algorithm at $\theta = 0$, plot the path taken in the parameter space by the algorithm for three different values of the step size ρ. In particular, let ρ equal the inverse of the maximum eigen value of the covariance matrix, one-twentieth that value, and one-fortieth that value.

Submit whatever code you have written for this question, along with a README briefly describing your code and submit them to the homework directory.