EE512 Graphical Models
Spring 2006

University of Washington
Dept. of Electrical Engineering

Midterm 1: Due Friday, May 12, 2006, 5:00pm

Prof: Jeff A. Bilmes <bilmes@ee.washington.edu> TA: Chris Bartels <bartels@ee.washington.edu>

In all problems, make sure your answers are extremely clear, complete, concise, and neatly written. Points will be taken off if we can not read your solution, if they are excessively long, or unclear. Do not wait until the last minute to start this midterm. You may not discuss this midterm with anyone or anything except for Prof. Bilmes or the TA (Chris Bartels). All solutions must be entirely your own work.

Problem 1. Factorization

Consider the following factorization property. Let $p(X)$ be a positive distribution over n random variables, and suppose that property (F) holds true on this distribution with graph $G = (V(G), E(G))$. Let $\mathcal{C}$ be the set of *all* cliques in the graph, where $c \in \mathcal{C}$ is a set of nodes that constitute a clique.

Consider the following factorization of $p(x)$:

$$p(x) = \prod_{c \in \mathcal{C}} \phi_c(x_c)$$

where

$$\phi_c(x_c) \triangleq \frac{p(x_c)}{\prod_{c' \subset c} \phi_{c'}(x_{c'})}$$

Note that $\phi_c(x_c)$ is defined recursively.

Given this definition of $\phi_c(x_c)$, can you prove this factorization is valid using the Möbius inversion lemma? If so, then prove that the above factorization is valid. If not, very clearly argue why the Möbius inversion lemma is not applicable. Is this factorization an instance of inclusion-exclusion? Clearly state how or how not.

Problem 2. Junction Trees.

Prove or disprove: All junction trees over the maximal cliques of the same triangulated graph have exactly the same set of separators (also counting multiplicity, meaning if there are k separators that consist of the same set of underlying original graph variables, there will be k such separators of that original graph set).

Problem 3. k-trees

A k-tree is an important graph-theoretic concept, that is very important for many graphical model settings.

We define a k-tree using the following generative language. A k-tree over $k+1$ nodes is a clique with $k+1$ nodes. A k-tree with $n > k+1$ nodes is generated from a k-tree with $n-1$ nodes by connecting the n^{th} node to k fully connected previous nodes (i.e., the k previous nodes must form a completed set).

You can see that a 1-tree is just a tree.

3a) Draw a k-tree for $k = 2$ and $k = 3$.

3b) Given a k-tree over n nodes, what is the maximum cliques size? Prove that your answer is correct.

3c) How many additional edges are needed to triangulate a k-tree?

3d) Prove that all minimal separators in a k tree are exactly the same size. Next, what is the size of a minimal separator in a k tree? Last, are the minimal separators of a k tree complete?

Problem 4. Markov Properties.

Consider a graph $G = (V(G), E(G))$. Consider a set of nodes $F \subseteq V(G)$. We say that F *fragments* G into $k = k_F$ connected components $U_1, U_2, \ldots, U_k$ if it is the case that each of U_i is connected (there is a path between any pair of nodes within U_i), and that any path from U_i to U_j must intersect F for all $i \neq j$.

Consider the following two Markov properties on MRFs.

Definition 4.1: (MP-1) Suppose that S is a separator that partitions the graph G into two sets of vertices V_1 and V_2 such that S separates V_1 from V_2. Then we have $X_{V_1} \perp\!\!\!\perp X_{V_2} | S$.

Definition 4.2: (MP-2) For any set F that fragments G in to k fragments, then $X_{U_1}, X_{U_2}, \ldots, X_{U_k}$ are mutually independent given F. (see HW1-problem 5 for the notion of what we mean by mutually independent, here note that the granularity is in terms of the fragments).

Your task is to prove that these two Markov properties are mathematically equivalent.

Problem 5. The EM algorithm and gradient descent.

In class, we defined two methods to train the parameters of an HMM (and more generally a graphical model), the EM auxiliary equation and gradient descent. We recall them here in a more general context. Lets suppose that λ consists of all parameters of a graphical model. A distribution factors according to a graphical model with graph $G = (V(G), E(G))$ where $V(G) = E \cup H$ is partitioned into hidden and observed variables $\bar{x}_E$ and x_H. We assume all observed variables a bound to their observed values. Assuming a previous set of parameters λ^p, we defined the EM auxiliary equation (also called the expected complete log-likelihood) as:

$$f(\lambda) = Q(\lambda, \lambda^p) = E_{p(x_H|\bar{x}_E, \lambda^p)}[\log p(x_H, \bar{x}_E|\lambda)] = \sum_{x_H} p(x_H|\bar{x}_E, \lambda^p) \log p(x_H, \bar{x}_E|\lambda)$$

We also defined the incomplete log-likelihood function, defined as:

$$g(\lambda) = \log p(\bar{x}_E|\lambda)$$

You task is to show that the gradient of $Q(\lambda, \lambda^p)$ is the same as the gradient of $\log p(\bar{x}_E|\lambda)$, both with respect to λ^p where $p()$ obeys (F) on G. In other words, show that:

$$\frac{\partial}{\partial \lambda} Q(\lambda, \lambda^p) \bigg|_{\lambda = \lambda^p} = \frac{\partial}{\partial \lambda^p} \log p(\bar{x}_E|\lambda^p)$$

on a graphical model G.

To make this more controlled, lets assume that each random variable involved has r possible values (i.e., $x_i \in D_i$ where $|D_i| = r$ for all i). If we have a vector of variables X_c for index set c, we can say that $X_c \in D_c$ where $D_c = D_{c_1} \times D_{c_2} \times \ldots \times D_{c_{|c|}}$ is the Cartesian product of the domain sets.

We assume that $p()$ obeys (F) on G with factorization:

$$p(x|\lambda) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi_{c,\lambda}(x_c)$$

where

$$Z = \sum_x \prod_{c \in \mathcal{C}} \psi_{c,\lambda}(x_c)$$

is a normalizing constant. The quantity $\mathcal{C}$ are the maxcliques of G, so we only worry about maxcliques (not subsets of maxcliques which could be just cliques). We also assume that each clique potential is parameterized in the following simple log-linear way:

$$\psi_{c,\lambda}(x_c) = e^{\sum_{x'_c \in D_c} \lambda_{x'_c} \delta(x_c, x'_c)}$$

in other words, in each maxclique potential function we have a unique scalar parameter for every possible configuration of the clique's variable values. Lastly, we assume that G is triangulated.

extra credit: What happens if G is not triangulated?

Problem 6. Independence Axioms and Pearl's eggs.

Consider the five Independence axioms as we have defined them on the nodes of Lecture 6, page 15/16, specifically properties C1, C2, C3, C4, and C5. Consider for this problem $X \perp\!\!\!\perp Y | Z$ where $X, Y, Z \subseteq V$ is a ternary relation, meaning it is either true or false for the subsets X, Y, and Z.

Your task is to show that the properties are not redundant, meaning they are not derivable from each other. In other words, show that it is possible to have a collection V of $|V|$ objects that are defined to logically satisfy a set of ternary relations of the form $X \perp\!\!\!\perp Y | Z$ for various subsets of V, and that it is possible to define these relations such that any four of C1,C2, C3, C4, and C5 hold, but the remaining property does not hold.

Note that since this is really asking you to solve 5 problems, you only need to solve two of them, namely choose any four of of C1,C2, C3, C4, and C5, and show that there exists a set of objects with corresponding ternary relation properties, for which the four properties you choose holds, but the fifth does not. And then repeat one time for a different set of four. It will greatly simplify your solutions if you make $|V|$ as small as possible.

Make sure your answer is extremely clear and is as concise as possible.