CIS 620 — Graphical Models — Spring 2009
Homework 6
(Problems from Carlos Guestrin)

March 23, 2009

1 Variational Inference

In this problem, you will investigate *mean field* approximate inference algorithms. Consider the Markov network in Figure 1(a). Define edge potentials $\phi_{ij}(x_i, x_j)$ for all edges (x_i, x_j) in the graph. We can write

$$P(x_1, \ldots, x_{12}) = \frac{1}{Z} \prod_{(i,j) \in E} \phi_{ij}(x_i, x_j)$$

(a) Assume a fully factored mean field approximation Q (Figure 1(b)), parameterized by node potentials Q_i.

(i) Write down the update formula for $Q_1(x_1)$.

(ii) Write down the update formula for $Q_6(x_6)$.

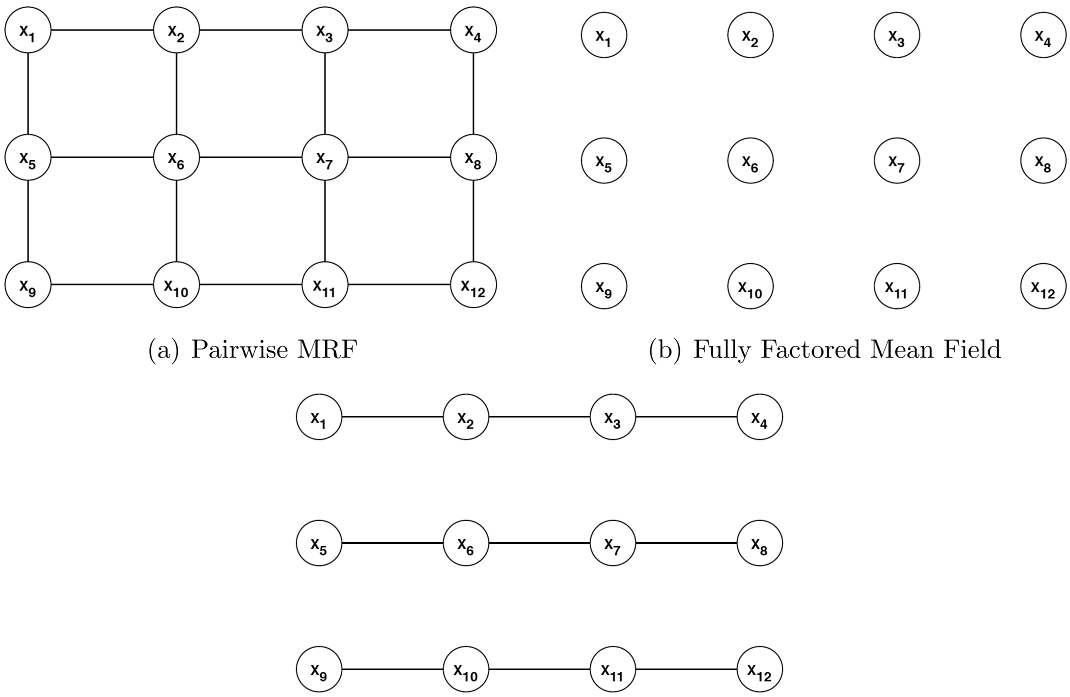
In both cases, please expand out any expectations in the formulas (your answer should be in terms of Q_i and ϕ_{ij}).

(b) Now we consider a structured mean field approximation Q (Figure 1(c)), parameterized by edge potentials $\psi_{ij}(x_i, x_j)$ for each edge (x_i, x_j) in Figure 1(c).

The update equation for potential ψ_j is,

$$\psi_j(\mathbf{c}_j) \propto \exp\Big[\sum_{\phi \in A_j} E_Q\big[\ln \phi | \mathbf{c}_j\big] - \sum_{\psi_k \in B_j} E_Q\big[\ln \psi_k | \mathbf{c}_j\big]\Big],$$

where $A_j = \{\phi \in \mathcal{F} : \text{scope}(\phi) \text{ is not independent of } \mathbf{C}_j \text{ in } Q\}$ and $B_j = \{\psi_k : \mathbf{C}_k \text{ and } \mathbf{C}_j \text{ are not independent in } Q\}$.



(a) Pairwise MRF

(b) Fully Factored Mean Field

(c) Structured Mean Field

Figure 1: A pariwise Markov Random Field and the structure of two mean field approximations

(i) Write down the update formula for $\psi_{34}(x_3, x_4)$

(ii) Write down the update formula for $\psi_{67}(x_6, x_7)$

In both cases, write the required formula up to a proportionality constant. You can write it in terms of expected values, but do not include unnecessary terms.

(iii) Write out the formula for $E_Q[\ln \phi_{12}(X_1, X_2)|x_3, x_4]$. Make sure to show how you would calculate the distribution that this expectation is over.

(iv) Repeat for $E_Q[\ln \phi_{15}(X_1, X_5)|x_3, x_4]$.

(v) Repeat for $E_Q[\ln \phi_{37}(X_3, X_7)|x_3, x_4]$.

(c) For an $n \times n$ grid with k-ary variables:

(i) What is the computational complexity of a single potential update (like $Q_6(x_6)$) in the fully factored model?

(ii) What is the computational complexity of a single potential update (like $\psi_{67}(x_6, x_7)$) in the structured mean field model?

(*Note*: Do not include the cost of computing the normalization constant in your answer).

(iii) Assuming No caching, what is the total cost in each case for a full iteration, i.e., that is computing the updates for all the potentials?

(d) We would like to use caching to speed up computations in the structured mean field approach in Figure 1(c).

(i) What are the (conditional) marginal distributions under Q needed to calculate the update for $\psi_{34}(x_3, x_4)$.

(ii) Repeat for $\psi_{34}(x'_3, x'_4)$.

(iii) Repeat for $\psi_{12}(x_1, x_2)$.

(iv) Using the above intuition, sketch a scheme to schedule the updates over all $\psi_{XY}(X, Y)$ for all possible assignments to X and Y. You may use any exact inference algorithm as a subroutine.

(v) For an $n \times n$ grid with k-ary variables, what is the commotional complexity for a single full iteration in your new scheme?

2 Generalized Belief Propagation

In Generalized Belief Propagation (GBP) we pass messages between clusters of nodes, rather than individual nodes, which can lead to better approximations. For this question, refer to (K&F Section 10.3) .

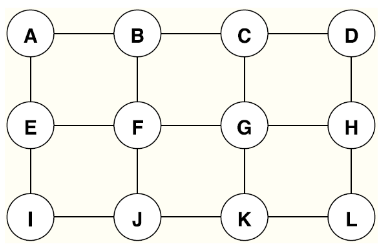


Figure 2: Markov Network for Generalized Belief Propagation Question

(a) [6 pts]Draw the region graph for the undirected model in Figure 2, assuming overlapping clusters of four nodes.

(b) [9 pts] Assume that this pairwise Markov Random Field has node potentials ϕ_a for all $a \in \{A, B, \ldots, L\}$, and edge potentials ψ_{ab} for all $(a, b) \in E$, the edge set of the model. Write down the belief equations for β[G], β[CG], β[BCFG]. These equations should be in terms of node potentials, edge potentials, and messages from regions to their subregions.

(c) [5 pts] Write down the message sent from region CG to region G.

(d) [5 pts] Use the belief equations you derived in part (b) as well as the marginalization consistency condition for beliefs (if $r \rightarrow r'$ then $\sum_{C_r - C_{r'}} \beta_r[C_r] = \beta_{r'}[C_{r'}]$), to derive the message sent from region CG to region G.