

10708 Graphical Models: Homework 4

Due November 15th, beginning of class

October 27, 2006

Instructions: There are six questions on this assignment. Each question has the name of one of the TAs beside it, to whom you should direct any inquiries regarding the question. The last problem involves coding. Do *not* attach your code to the writeup. Instead, copy your implementation to

`/afs/andrew.cmu.edu/course/10/708/your_andrew_id/HW4`

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

Note: Please put your name and Andrew ID on the first page of your writeup.

1 Markov Network Representations [5 pts] [Khalid]

Figure 1 is a Markov Random Field where the potentials are defined on all cliques of three variables.

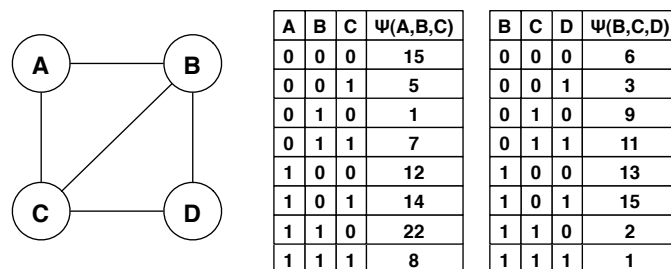


Figure 1: A chordal (triangulated) Markov network

- (a) Convert the triangle graph on (A, B, C) with potential $\Psi(A, B, C)$ into a pairwise Markov Random Field by introducing a new variable X . Show the graph, as well as the node and edge potentials in table form (*i.e.*, compute the values of the potentials in the pairwise MRF)

- (b) Convert the graph on (A, B, C, D) with potentials $\Psi(A, B, C)$ and $\Psi(B, C, D)$ into a pairwise Markov Random Field. Is the graph chordal? (*Note:* You do not have to compute the pairwise MRF potentials in your solution).

2 Hammersley-Clifford [10 pts] [Ajit]

Complete the analysis of Example 5.4.3 (Koller & Friedman, pg 199), showing that the distribution P defined in the example does not factorize over \mathcal{H} . (*Hint:* Use a proof by contradiction).

3 Importance Sampling [20 pts] [Khalid]

To do this question you need to read (Koller & Friedman, 10.2.2). The likelihood weighting of an importance sampler is defined as $w(\mathbf{x}) = P(\mathbf{x})/Q(\mathbf{x})$ where P is the distribution we want to sample from and Q is a proposal distribution.

- (a) Why is computing the probability of a complete instantiation of the variables in a Markov Random Field computationally intractable? Your answer should be brief.
- (b) Given a chordal graph, describe how to compute the likelihood weighting for an importance sampler. (*Hint:* What is the relationship between chordal graphs and junction trees?)
- (c) Given a non-chordal graph, describe how to compute the likelihood weighting for an importance sampler.
- (d) Why is it not useful to do use importance sampling for approximate inference on Markov Random Fields? Your answer should be brief.

4 Generalized Belief Propagation [20 pts] [Khalid]

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 *Understanding Belief Propagation and its Generalizations* (Yedidia et. al.) on the web page.

- (a) Draw the region graph for the undirected model in Figure 2, assuming overlapping clusters of four nodes.
- (b) Assume that this pairwise Markov Random Field has node potentials ϕ_a for all $a \in \{A, B, \dots, L\}$, and edge potentials ψ_{ab} for all $(a, b) \in E$, the edge set of the model.

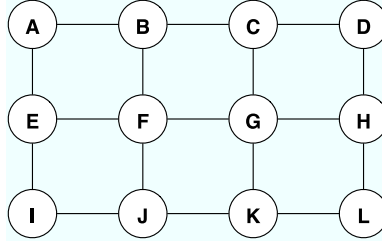


Figure 2: Markov Network for Generalized Belief Propagation Question

Write down the belief equations for b_G , b_{CG} , b_{BCFG} . These equations should be in terms of node potentials, edge potentials, and messages from regions to their subregions – *e.g.*, if region [BF] is a subregion of [ABEF], we write the message from [ABEF] to [BF] as $m_{AE \rightarrow BF}(x_B, x_F)$.

- (c) Using your answers to part (b), as well as the marginalization condition for beliefs, write the message update rule for $m_{C \rightarrow G}(x_G)$.
- (d) If you add an edge in the region graph from region [ABEF] to region [F], do the GBP fixed points change? If you add an edge in the region graph from region [ABEF] to region [G], do the GBP fixed points change? Explain briefly.

5 Variational Inference [20 pts] [Khalid]

In this problem, you will investigate *mean field* approximate inference algorithms (Koller & Friedman 11.5). Consider the Markov network in Figure 3(a). Define edge potentials $\phi_{ij}(x_i, x_j)$ for all edges (x_i, x_j) in the graph. We can write

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

- (a) Assume a fully factored mean field approximation Q (Figure 3(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 3(c)), parameterized by edge potentials $\psi_{ij}(x_i, x_j)$ for each edge (x_i, x_j) in Figure 3(c).

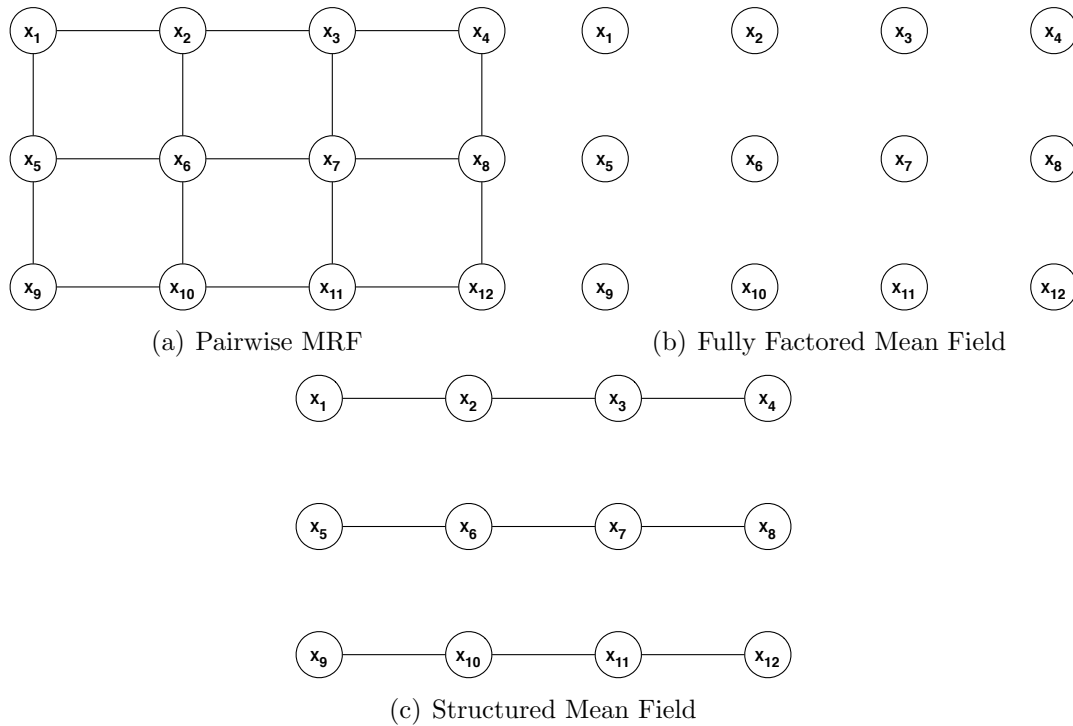


Figure 3: A pairwise Markov Random Field and the structure of two mean field approximations

- (i) Write down the update formula for $\psi_{12}(x_1, x_2)$ up to a proportionality constant. This time, you can write it in terms of expected values, but do not include unnecessary terms.
 - (ii) Write out the formula for $E_Q[\ln \phi_{23}(X_2, X_3)|x_1, x_2]$. Make sure to show how you would calculate the distribution that this expectation is over.
 - (iii) Write out the formula for $E_Q[\ln \phi_{15}(X_1, X_5)|x_1, x_2]$, Again, show how you would evaluate distribution Q .
- (c) For an $n \times n$ grid with k -ary variables, what is the computational complexity of a single message update in the fully factored model? What is the computational complexity of a single message update in the structured mean field model in figure 3(c)? (*Note:* Do not include the cost of computing the normalization constant in your answer).

6 Image Segmentation [25 pts] [Ajit]

Given an image of $l \times w$ pixels a K -ary segmentation is a clustering that assigns each pixel to one of K -classes, typically under the assumption that neighbouring pixels are more likely to belong to the same class.

The most common graphical model approach for image segmentation represents the image as a pairwise Markov random field (Figure 4) where each node corresponds to a pixel. Note that the value of a node is the cluster it belongs to.

Formally, the observed image is denoted $Y = \{Y_i\}$ and $X = \{X_i\}$, $X_i \in \{1 \dots K\}$ is the segmentation. The Markov random field has distribution

$$P(X, Y) = \frac{1}{Z} \prod_i \Phi(x_i, y_i) \prod_{(i,j) \in E} \Psi(x_i, x_j)$$

where Φ is the node potential¹, the effect that pixel y_i has on the label of x_i ; Ψ is the edge potential², how the label of x_i is influenced by the labels of its neighbours. Let

$$\Phi(x_i, y_i) = \exp \left\{ -\frac{(y_i - \mu_{x_i})^2}{2\sigma_{x_i}^2} \right\}$$

$$\Psi(x_i, x_j) = \exp \left\{ -\beta (x_i - x_j)^2 \right\}$$

6.1 Segmentation [20 pts]

Consider the image in Figure 5. We want to produce a binary segmentation ($K = 2$). You are given the following parameters: $\beta = 20$, $\mu_1 = 147$, $\sigma_1^2 = 1/2$, $\mu_2 = 150$, $\sigma_2^2 = 1/2$ - *i.e.*, pixels from segment 1 are normally distributed with mean μ_1 and variance σ_1^2 ; pixels from segment 2 are normally distributed with mean μ_2 and variance σ_2^2 . The image is located in `img.mat` (MATLAB format).

1. Produce a binary segmentation using loopy belief propagation.
2. Produce a binary segmentation using Gibbs sampling.

Produce plots of your segmentations, which must be included in your writeup. For loopy belief propagation initialize $m_{ij}(x_i) = 1$ for all $i \neq j$. Stop running loopy belief propagation once the maximum absolute difference between an old message and a new message is less than 10^{-5} . For Gibbs sampling, just flip a fair coin to select the initial value of each x_i .

Notes:

- You're free to implement this question in any programming language you like. As always, you must submit your code to the AFS directory.
- Don't use the min-cut formulation of binary MAP because (a) we want you to implement loopy belief propagation; (b) it's harder to implement than loopy belief propagation; (c) unless you're clever with data structures, min-cut will be a lot slower than loopy belief propagation on this problem.

¹Also called the observation model or likelihood.

²Also called the Ising prior.

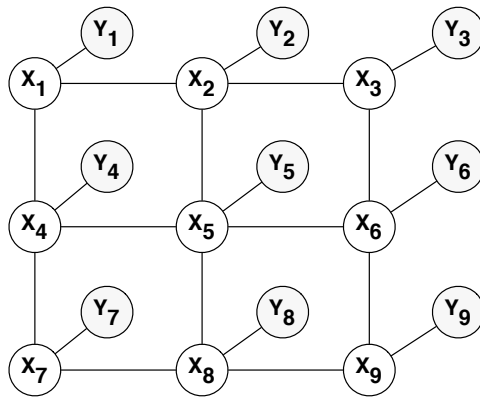


Figure 4: An example of a Markov Random Field for image segmentation

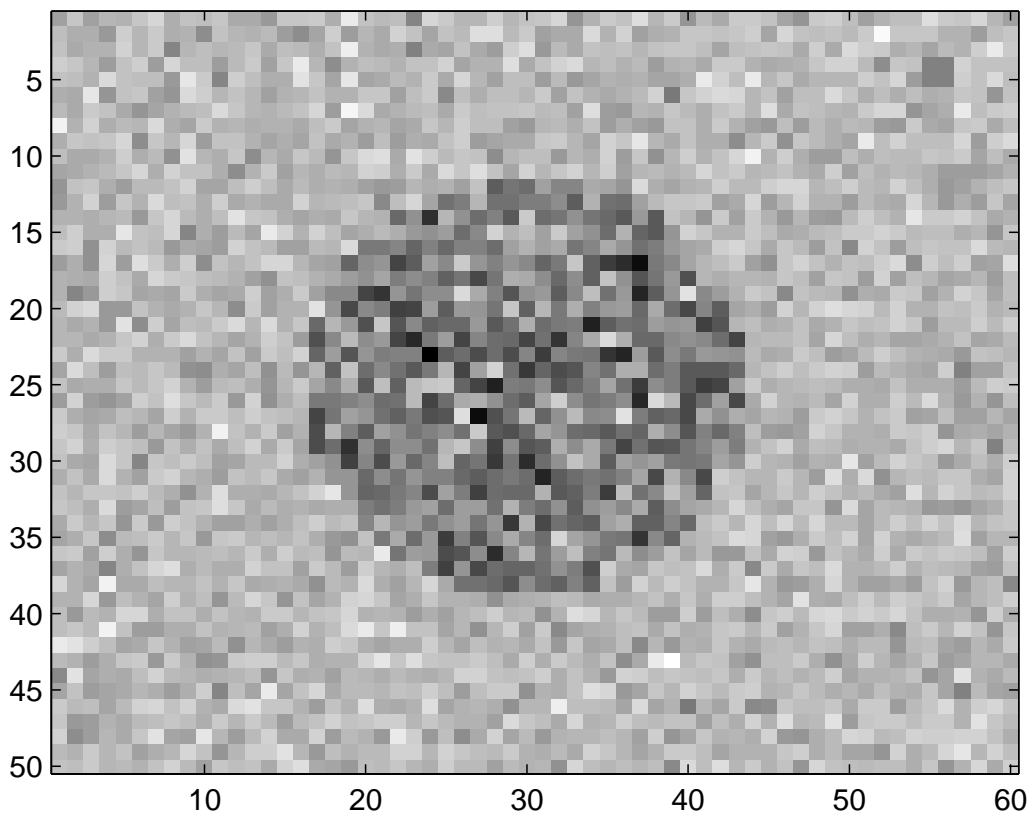


Figure 5: Image for segmentation question

- Remember to compute the messages in log-space, for numerical stability.
- You do not need to dampen messages to ensure convergence on this image.
- In Gibbs sampling, the transition corresponds to changing the value of one X_i , which for $K = 2$ involves sampling a binomial.
- In Gibbs sampling, pick a few nodes which are far apart in the graph, and check that each marginal converges. You do not need to implement any automated test for convergence of the Markov chain.

6.2 Potentials [5 pts]

Using loopy belief propagation, produce segmentations with $\beta = 2.0$ and $\beta = 5.0$. Include plots of the segmentation in your writeup. Briefly describe the effect that β has on the segmentation.