10708 Graphical Models: Homework 5

Due November 24th, beginning of class

Note: Due sooner than usual

November 12, 2008

Instructions: There are three 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/HW5`

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 Sampling From a Markov Network [30 pts][Amr]

In this question we will explore different techniques to sample from a Markov Network.

1. [10 pts] Given a *calibrated* clique tree $\mathcal{T}$ with clique potentials $P(C_i)$ and sepset potentials $P(S_{ij})$
   (a) If clique C_i is a neighbour of clique C_j in the tree with a sepset S_{ij}, show how to calculate $P(C_j|C_i = c_i)$, that is the probability distribution over C_j given an assignment to C_i. (*Hint*: use Bayes rule and the independence assertion imposed by S_{ij}).
   (b) Show how to use forward sampling to generate a sample from the probability $P(X)$ encoded by the calibrated tree $\mathcal{T}$ (*Hint*: start by directing the tree by picking an arbitrary root).
2. [5 pts] Show how to use forward sampling to generate a sample from a *chordal* Markov network with Maximal Clique size w. What is the cost of this operation? (*Hint*: What is the relationship between chordal MNs and clique trees?)

3. [5 pts] Given a general MN (not necessarily chordal), show how to use forward sampling to generate a sample from it. Is this always efficient?

4. [5 pts] It seems that forward sampling might be prohibitive in MN. Importance sampling is a technique used when sampling from a distribution $\mathcal{P}$ is expensive but evaluating $P(X = x)$ is not expensive. Why this technique might be expensive as well for a MN? (Hint: your answer should be really 1-2 sentences)

5. [5 pts] Finally, in Gibbs sampling, the major task is to efficiently compute $P(X_i|\text{MB}(X_i))$. Briefly explain why this operation can be computed efficiently in a MN? (*Hint*: your answer should be just 1-2 sentences with the high-level intuition)

2 Generalized Belief Propagation [25 pts][Amr]

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) .

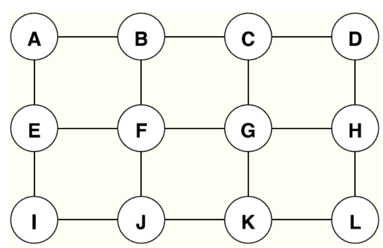


Figure 1: Markov Network for Generalized Belief Propagation Question

(a) [6 pts]Draw the region graph for the undirected model in Figure 1, assuming overlapping clusters of four nodes. (*Hint*: see Figure 10.12)

(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. (*Hint*: use Equation 10.36)

(c) [5 pts] Write down the message sent from region CG to region G. (*Hint:* use Equation 10.37)

(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.

3 Iterative Proportional Fitting [45 pts] [Dhruv]

We continue with the binary segmentation problem from homework 4. In the previous homework, you were given the parameters of the Markov Random Field and asked to produce a segmentation. In this question, you will learn the spatial prior Ψ. In a typical setting, we are given a set of training images, along with manually labelled ground-truth segmentations. We use these training images to learn parameters of our model and then proceed to segment test images with these parameters. For simplicity, in this exercise, we will only learn parameters from a single training image, shown in figure 2.

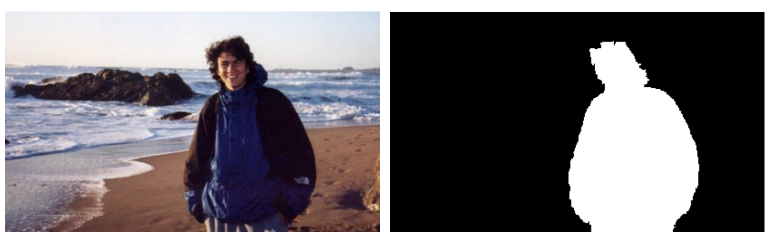
Recall that we have a pairwise Markov Random Field where each node corresponds to a superpixel. The observed image is denoted $y = \{y_i\}$ and $x = \{x_i\}$, $x_i \in \{1, 2\}$ is the segmentation. The Gibbs distribution of this model is

$$P(x, y) = \frac{1}{Z} \prod_{i \in V} \Phi(x_i, y_i) \prod_{(i,j) \in E} \Psi(x_i, x_j)$$

where the potentials are defined as follows:

$$\begin{aligned} \Phi(X_i = \text{fg}, y_i) &= P(y_i | \text{GMM}_{\text{fg}}) \\ \Phi(X_i = \text{bg}, y_i) &= P(y_i | \text{GMM}_{\text{bg}}) \\ \Psi(x_i, x_j) &= \begin{bmatrix} \theta_1 & \theta_2 \\ \theta_3 & \theta_4 \end{bmatrix}. \end{aligned}$$

For convenience we denote $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$. This MRF is not decomposable, and therefore we cannot estimate the potentials in closed form.



(a) Input image (b) Ground Truth Segmentation

Figure 2: Training Data

(a) Assume that Φ is known (computed in a manner similar to HW 4, only now all pixels are labelled). Write down the IPF update equation for $\Psi(x_i, x_j)$. What is the cost of computing $\Psi^{(t+1)}(x_i, x_j)$?

(b) Using the equation from part (a) implement IPF for Ψ using the images and data provided. Report the final value of θ. Use loopy belief propagation to compute any required probabilities. You may use your implementation or our solution from homework 4 (`lbp.m`). *Hint*: We discussed how to compute pairwise distributions $P(x_i, x_j)$ using the messages from loopy belief propagation in class.

Notes:

(a) You will need to re-learn the foreground and background GMMs. Instead of using the pixel Luv vectors like last time (which will take a lot of time now), train the GMM on the mean Luv vectors for superpixels. These are stored in the variable `modes`.

(b) In our setup above, all edges have the same edge potentials. This is known as *parameter sharing*, and requires us to compute average edge probabilities.

(c) You will experience first-hand an interesting phenomenon in graphical models, which is that learning requires inference. For example, in the above problem, you will need 20-30 iterations of IPF, and each iteration requires loopy BP which is an iterative algorithm itself! This is why fast algorithms for inference result in fast algorithms for learning.