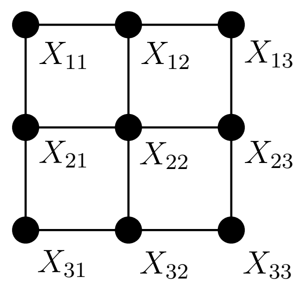
Homework 2: Graphical models

CIS 620

February 6, 2005

This homework is due noon, Thursday, February 10. It should be submitted by email to me (`pereira@cis.upenn.edu`) as a `tar` or `zip` file.

1. Consider the undirected graphical model with the following structure



   where all the variables are binary (values 0 or 1), representing a nine-pixel image. suppose that we want to represent the intuitive notion that a pixel is more likely to be 1 as more of its neighbors are 1. Let N_{ij} be the sum of the values of neighbors of variable X_{ij}. We might guess

$$\begin{aligned} P(X_{22}=1|\text{neighbors}) &= 0.1 + 0.2 \times N_{22} & \\ P(X_{ij}=1|\text{neighbors}) &= 0.2 + 0.2 \times N_{ij} & ij = 12, 21, 23, 32 \\ P(X_{ij}=1|\text{neighbors}) &= 0.25 + 0.25 \times N_{ij} & ij = 11, 13, 31, 33 \end{aligned}$$

   Is there a joint distribution on the variables with those conditional distributions? Justify your answer.

2. The "hard" direction of the Hammersley-Clifford theorem, showing that an MRF has a Gibbs distribution compatible with the graph of

the MRF, uses the assumption that $P(\mathbf{x}) > 0$ for every configuration $\mathbf{x}$. Show that this assumption is necessary by creating a counterexample, that is, an undirected graphical model with the Markov property in which some configurations have zero probability that does not have a distribution that factors as a product of clique potentials. **Hint**: four binary variables are enough.

3. In CIS 520, you learned about Bayes nets aka directed graphical models. HMMs, in particular, are usually presented as Bayes nets. As a reminder, a Bayes net is a directed acyclic graph (DAG) whose nodes are random variables. Each variable X with parents in the graph $Y_1, \ldots, Y_k$ is associated with a conditional probability $P(X|Y_1, \ldots, Y_k)$. If $\mathbf{X} = \{X_i\}_{i \in I}$ is the set of variables of a Bayes net and pa_i is the set of indices for the parents of variable X_i, then

$$P(\mathbf{X}) = \prod_i P(X_i|\mathbf{X}_{\mathrm{pa}_i}) \quad .$$

Notice that the above equation is in Gibbs form where each $P(X_i|\mathbf{X}_{\mathrm{pa}_i})$ would be a potential on the variables $\{X_i\} \cup \mathbf{X}_{\mathrm{pa}_i}$. In class, I observed that for the case of an HMM, and noted that if we just make the graph edges undirected, the above potentials form a Gibbs distribution over the now undirected graph.

   (a) The above trick (just make the edges undirected) does not work an arbitrary Bayes net. That is, the decomposition of $P(\mathbf{X})$ into the product of conditional distributions does not always make $P(\mathbf{X})$ a Gibbs distribution over the undirected version of the graph. Give a simple example that shows this.

   (b) Can you generalize from HMMs to a broader class of Bayes nets for which the trick works?

   (c) Given that the trick does not work in general, can we modify the graph by adding edges as well as making the original edges undirected to create a graph that the decomposition in terms of conditionals is Gibbs with respect to?

   (d) This conversion from directed to undirected graphical models is less clean that one would like. Alternatively, one could convert a Bayes net to a factor graph. Given informally an algorithm for converting an arbitrary Bayes net into a factor graph.

4. The second paper presented in class on February 3rd discussed experimental results on loopy belief propagation for Bayes nets. The purpose of this problem is to partially reproduce the results for the PYRAMID net discussed in the paper. The graph for the net is given in the paper.

   (a) Generate five random CPTs for the network.

   (b) For each of the five Bayes nets above, use their CPTs to generate four random partial assignments to the bottom (leaf) variables.

   (c) Using the conversion of Bayes nets to factor graphs, run the loopy sum-product algorithm discussed in class on the twenty cases (five nets, four leaf node assignments for each net) until convergence. By convergence here I mean that the ratio of the marginals for any non-leaf variable on two successive iterations is less than 10^{-2}. Limit the number of iterations to 100 on each case, whether convergence occurs or not.

   (d) Plot $P(X_i = 1)$ for each non-leaf variable obtained above against the corresponding marginal obtained by summing over all assignments to the non-leaf variables. (You can do this slightly more efficiently by computing first the 4,096 joint probabilities of each assignment under consideration, and then selecting and summing the entries in that table for $X_i = 1$ for each i).

   Deliverables: code and plot. **Note**: This question is experimental. There are details to fill in (that's the real world!), and I don't know exactly what the results should be, although Murphy et al's paper gives us an idea. If you get stuck, ask me for advice, either in class or by email.