10708 Graphical Models: Homework 5

Due November 29th, beginning of class

November 16, 2006

Instructions: There are four 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 Gaussian Graphical Models [25 pts] [Khalid]

Consider a Gaussian graphical model over three variables x_1, x_2, x_3, represented by the following multivariate Normal distribution:

$$P(x_1, x_2, x_3) = \frac{1}{(2\pi)^{3/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right)$$

where $\mathbf{x} = (x_1\ x_2\ x_3)$ and $\mu = (\mu_1\ \mu_2\ \mu_3)$ are column vectors, and Σ is a 3×3 covariance matrix. Let

$$\Sigma_1 = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}, \ \Sigma_2 = \begin{bmatrix} 0.75 & 0.5 & 0.25 \\ 0.5 & 1.0 & 0.5 \\ 0.25 & 0.5 & 0.75 \end{bmatrix}$$

(a) Suppose $\Sigma = \Sigma_1$. Are there any marginal independencies among the components of $\mathbf{x}$?

(b) Suppose $\Sigma = \Sigma_2$. Are there any marginal independencies among the components of $\mathbf{x}$?

(c) Suppose that the distribution of $\mathbf{x}$ can be represented by the Bayes net in Figure 1 with the following parameterization:

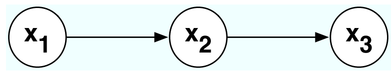

Figure 1: Structure of Gaussian graphical model.

$$
\begin{aligned}
X_1 &\sim N(0,1) \\
X_2|X_1 = x_1 &\sim N(x_1, 1) \\
X_3|X_2 = x_2 &\sim N(x_2, 1)
\end{aligned}
$$

Take the product of these three distributions to obtain the joint distribution. Express the joint in canonical form and compute the inverse covariance (*a.k.a.* precision) matrix Σ^{-1}. Referring to the structure in Figure 1, how can you interpret the zeros in this matrix in terms of conditional independencies ?

(d) Once again, suppose $\Sigma = \Sigma_1$. Are there any *conditional* independencies among the components of X?

(e) Suppose $\Sigma = \Sigma_2$. Are there any *conditional* independencies among the components of X?

2 Non-Decomposable Models [25 pts] [Khalid]

In Figure 2 we have a pairwise Markov Random Field with joint distribution

$$P(x_1, x_2, x_3, x_4) = \frac{1}{Z}\Psi_{12}(x_1, x_2)\Psi_{23}(x_2, x_3)\Psi_{34}(x_3, x_4)\Psi_{41}(x_4, x_1).$$

Let $\tilde{p}(x_1, x_2)$, $\tilde{p}(x_2, x_3)$, $\tilde{p}(x_3, x_4)$, $\tilde{p}(x_4, x_1)$ be empirical marginals. (Jordan 9.3.3) describes a procedure for finding maximum likelihood parameter estimates for decomposable graphs by inspection. In this problem, you will show that this doesn't work for the graph in Figure 2.

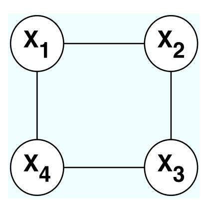

Figure 2: A non-decomposable Markov Random Field

(a) Briefly explain why this model is not decomposable

(b) Write down expressions for the parameter estimates $\hat{\Psi}_{12}, \hat{\Psi}_{23}, \hat{\Psi}_{34}, \hat{\Psi}_{41}$ assuming decomposability .

(c) Using the expressions from part (b) write down the joint estimate $\hat{p}(x_1, x_2, x_3, x_4)$.

(d) Show that the parameter estimates from part (b) cannot be maximum likelihood estimates (*Hint*: proof by contradiction).

3 Decomposable Models [25 pts] [Khalid]

Iterative proportional fitting on a decomposable model converges in at most $2n$ iterations, where n is the number of variables in the model. In this question, you will show this for the model in Figure 3.

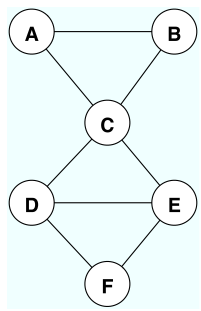

Figure 3: A decomposable Markov Random Field

(a) Show that any decomposable Markov Random Field can be represented as a junction tree.

(b) Consider the model in Figure 3. Assume that the variables are binary, and that the Markov Random Field is parameterized over maximal cliques,

$$P(A, B, C, D, E, F) = \frac{1}{Z} \Psi_{ABC}(A, B, C) \Psi_{CDE}(C, D, E) \Psi_{DEF}(D, E, F).$$

We wish to obtain maximum likelihood estimates for the parameters of the model using IPF. Assume an initialization of

$$\Psi^{(0)}_{ABC}(A, B, C) = \Psi^{(0)}_{CDE}(C, D, E) = \Psi^{(0)}_{DEF}(D, E, F) = 1$$

for all A, B, C, D, E, F.

Write out the IPF computation for this model until convergence to a fixed point. Specifically, this means we want a sequence of equations

$$\begin{array}{ccc} \Psi^{(0)}_{ABC} & \Psi^{(0)}_{CDE} & \Psi^{(0)}_{DEF} \\ \Psi^{(1)}_{ABC} & \Psi^{(1)}_{CDE} & \Psi^{(1)}_{DEF} \\ & \cdots & \\ \Psi^{(k)}_{ABC} & \Psi^{(k)}_{CDE} & \Psi^{(k)}_{DEF} \end{array}$$

until the parameter estimates stop changing. *Hint*: (Jordan 9.3.4) discusses properties of the IPF update equation.

(c) Using the maximum likelihood clique potentials learned in part (b), write out the expression for the joint probability over all the variables. What is the relationship between this expression and the junction tree in part (a) ?

(d) (Bonus Question – 10pts) Derive a closed form solution for the parameters of an arbitrary tabular decomposable model where the potentials are over maximal cliques. *Hint*: Generalize the results of parts (a)-(c).

4 Iterative Proportional Fitting [25 pts] [Ajit]

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.

Recall that we have a pairwise Markov Random Field where each node corresponds to a pixel. 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:

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

$$\Psi(x_i, x_j) = \begin{bmatrix} \theta_1 & \theta_2 \\ \theta_3 & \theta_4 \end{bmatrix}.$$

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) Assume that Φ is known ($\mu_1 = 147, \sigma_1^2 = 1/2, \mu_2 = 150, \sigma_2^2 = 1/2$). 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 image in `img.mat` as the training instance. 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.

(c) If we now assume that Ψ is fixed and Φ is unknown, can we use IPF to learn the model ? If so, write the update for Φ along with the cost of computing $\Phi^{(t+1)}$. If not, explain why and describe an alternate algorithm in terms of the parameters of Φ – you can either use the conventional parameterization $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$ or the canonical parameterization $(\lambda_1 = 1/\sigma_1^2, \eta_1 = \mu_1/\sigma_1^2, \lambda_2 = 1/\sigma_2^2, \eta_2 = \mu_2/\sigma_2^2)$.