E0:270 Homework 5

Mar 22, 2012

Due: Apr 3, 2012

Problem 1

Consider the following toy example. Professor X. Yz at IISc wants to do an analysis of the following discrete random variables: D taking values from $\{e, m, h\}$ indicating the (D)ifficulty of a course, W from $\{y, n\}$ indicating whether or not a student is hard-(W)orking, C from $\{a, b, c, d\}$ indicating the (C)ourse grade that a student gets, G from $\{ex, vg, g\}$ indicating the (G)ATE performance of the student, H from $\{y, n\}$ indicating whether or not the student is happy at the end of the course, L from $\{y, n\}$ indicating whether or not the professor writes a (L)etter of recommendation for the student, and S from $\{h, vh\}$ indicates the (S)alary level of the job that the student finally gets. (Observe that this assumes a student registers for a single course.) Assume that the department maintains a database with columns corresponding to these random variables, and one row for each student. Prof. Yz wishes to estimate the following conditional distribution $P(H|W = k)$, $k \in \{y, n\}$ of a student being happy at the end of the course given that he/she is hard-working or otherwise.

(a) Construct a fully parameterized (conditional probability tables populated with legitimate values) directed graphical model for the joint distribution over D, W, C, G, H, L and S. Clearly, there is no single correct answer for this. The graph will depend on the conditional independences that you assume. List the conditional independences that *you feel* should hold, and construct the graph that conforms with that list. What is the complexity of the inference task $P(H|W = k)$ for the graph that you have constructed?

(b) Construct a fully parameterized (completely defined potential functions, represented as tables) undirected graphical model for the same joint distribution. What is the complexity of the inference task for this graph?

(c) Two graphical models are said to be equivalent if the sets of conditional independences implied by them are equal. Are the graphical models you have constructed in (a) and (b) equivalent?

(d) Assuming that all values are observed for all students in the department database (including the happiness values!), what would be the expressions for maximum likelihood estimators of the parameters in (a)? Pick any two parameters and write the expressions for their MLE's.

(e) (Not to be submitted) Suppose the happiness values are not observed for the students. Design an algorithm for estimating the parameters for $P(H|\text{parents of } H)$ in (a).

Problem 2

(a) Consider the factorization of a joint probability mass function or density function $p(x_1 \dots x_N)$ associated with a directed *acyclic* graph G: $\prod_{i=1}^{N} f_i(x_i, x_{\pi(i)})$, where $\pi(i)$ denotes the set of parents of node i in G, and each $f_i(x, x_{\pi(i)})$ is any non-negative function such that $\sum_{x_{\pi(i)}} f_i(x, x_{\pi(i)}) = 1$ for all i. First, show that this represents a valid pmf. Then, show that each factor $f_i(x, x_{\pi(i)})$ is necessarily a conditional probability.

(b) Consider the factorization of a joint probability mass function or density function $p(x_1 \dots x_N)$ associated with an undirected graph G: $\frac{1}{Z}\prod_{C\in\mathcal{C}}\psi_{X_C}(x_C)$, where $\mathcal{C}$ is the set of maximal cliques of graph G, and Z is the normalization constant. Show that the potential functions $\psi_{X_C}(x_c)$ cannot in general be interpreted uniformly as conditional probabilities or uniformly as marginal probabilities.

Problem 3

A Hidden Markov Model is defined over $\{X_i, Z_i\}_{i=1}^T$ in the usual way as follows:

$$p(x_1, \dots, x_T, z_1, \dots, z_T) = p(z_1)p(x_1|z_1)\prod_{t=2}^{T} p(z_t|z_{t-1})p(x_t|z_t)$$

Assuming the parameters of the distribution to be known, derive the steps (messages) of the *Sum Product Algorithm* for finding the posterior distribution $p(z_t|x_1 \dots x_T)$. Compare it against the forward-backward algorithm. Explain your findings.

Problem 4

Consider an $n \times n$ image $\{y_i\}_{i=1}^{n\times n}$. Assume for simplicity that the pixels take binary values $\{1, -1\}$. You do not know any thing about the content of specific images, but know from domain knowledge that images taken under different lighting conditions have different intensity levels *overall*, and secondly, neighboring pixels *mostly* tend to take similar values. Using this prior knowledge, you decide to model the energy of an image as:

$$H(y) = -\sum_i Ay_i - \sum_{(i,j)\in\mathcal{N}} By_iy_j$$

where $\mathcal{N}$ is the set of index pairs of pixels that are either vertically or horizontally adjacent. The joint probability is given as usual by $p(y) = \frac{1}{Z}\exp\{-H(y)\}$.

Given a fully-specified joint distribution $p(y, x; \theta)$ over *hidden* variables $y = \{y_1 \dots y_n\}$ and *observed* variables x with *known* parameter θ, the basic **Gibbs Sampling algorithm** for computing $p(\bar{y}, \bar{x}; \theta)$ can be stated as follows:

0. (Randomly) Initialize $y_1 \dots y_n$
1. Repeat $t = N + M$ times
2. $\quad y_1^{t+1} \sim p(y_1|y_2^t, y_3^t \dots, y_n^t, \bar{x}; \theta)$
3. $\quad y_2^{t+1} \sim p(y_2|y_1^{t+1}, y_3^t \dots, y_n^t, \bar{x}; \theta)$
4. $\quad y_3^{t+1} \sim p(y_3|y_1^{t+1}, y_2^{t+1} \dots, y_n^t, \bar{x}; \theta)$
5. $\quad \dots$
6. $\quad y_n^{t+1} \sim p(y_n|y_1^{t+1}, y_2^{t+1} \dots, y_{n-1}^{t+1}, \bar{x}; \theta)$
7. $\tilde{p}(\bar{y}, \bar{x}) = \frac{1}{M}\sum_{t=N+1}^{N+M}\delta(y^t, \bar{y})$

where N is appropriately large (at least greater than 1000), $\delta()$ is the Kronecker delta function, and $p(y_1|y_2^t, y_3^t \dots, y_n^t, \bar{x}; \theta)$ is the *conditional* distribution of the 1^{st} hidden variable given the *current* values of *all other* variables. We will analyze this algorithm in class. For now, assume that this algorithm correctly returns the required probability mass or density $p(\bar{y}, \bar{x}; \theta)$.

(a) Assume that the task of estimating the parameters A and B of the image density has already been done for you, by considering a collection of images. You would like to infer the distribution $p(y)$ over configurations of pixel values, assuming values of A and B are specified. Design a Gibbs Sampling based inference algorithm for this problem. (More specifically, derive the conditional distributions for sampling values of all hidden variables for this problem.)

(b) (For extra credit) Suppose now you are given an image $\{x_i\}_{i=1}^{n\times m}$ corrupted by random noise. You wish to restore or de-noise the given image. How would you change the above model and algorithm to achieve this?

(c) (For extra credit) Implement your proposed approach to restore the provided images (on the course website), which are corrupted (with different degrees of random noise) versions of the same original image. Inference requires the values of A and B to be known. Experiment with simple integral values (e.g. 0,1,2) of A, B and any other parameter that you may need to introduce.