UC Berkeley
Department of Electrical Engineering and Computer Science
Department of Statistics

EECS 281A / STAT 241A Statistical Learning Theory

Problem Set 7
Fall 2012

Issued: Tues. Nov. 13, 2012 **Due:** Thurs. Nov. 29, 2012

Reading: Sampling chapter. Notes on mean field algorithm.

Problem 7.1

Gibbs sampling and mean field: Consider the Ising model with binary variables $X_s \in \{-1, 1\}$, and a factorization of the form

$$p(x; \theta) \propto \exp \Big\{ \sum_{s \in V} \theta_s x_s + \sum_{(s,t) \in E} \theta_{st} x_s x_t \Big\}. \tag{1}$$

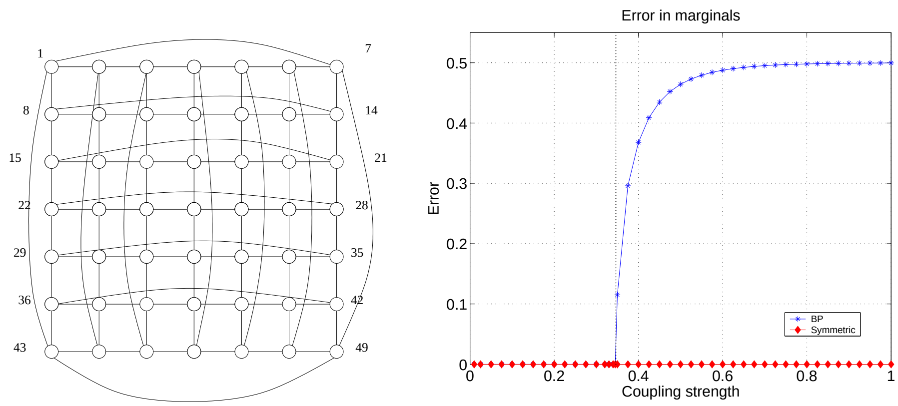
To make the problem symmetric, assume a 2-D grid with toroidal (donut-like) boundary conditions, as illustrated in Figure 2(a).

(a) Derive the Gibbs sampling updates for this model. Implement the algorithm for $\theta_{st} = 0.25$ for all edges, and $\theta_s = (-1)^s$ for all $s \in \{1, \ldots, 49\}$ (using the node ordering in Figure 2(a)). Run a burn-in period of 1000 iterations (where one iteration amounts to updating each node once). For each of 5000 subsequent iterations, collect a sample vector, and use the 5000 samples to form Monte Carlo estimates $\widehat{\mu}_s$ of the moments $\mathbb{E}[X_s]$ at each node. Output a 7×7 matrix of the estimated moments. Repeating this same experiment a few times will provide an idea of the variability in your estimate. Hand in print-outs of your code, as well as your results.

(b) Derive the naive mean field updates (based on a fully factorized approximation), and implement them for the same model. Compute the average ℓ_1 distance $\frac{1}{49} \sum_{i=1}^{49} |\tau_s - \widehat{\mu}_s|$ between the mean field estimated moments τ_s, and the Gibbs estimates $\widehat{\mu}_s$. Hand in print-outs of your code, as well as your results.

Problem 7.2

Neighborhood regression for Gaussian graphical models: In this problem, we explore properties of jointly Gaussian random vectors that underlie the



(a) (b)

Figure 1: (a) A two-dimensional grid graph with toroidal boundary conditions. (b) Break-down of the sum-product algorithm on the Ising model on the toroidal grid. For all $\gamma < \gamma^*$, the sum-product algorithm computes the correct symmetric marginals. Beyond this point, it outputs increasingly inaccurate answers.

success of the neighborhood-based Lasso approach to estimating Gaussian graphical models (as discussed in lecture).

Let $(X_1, X_2, \ldots, X_p)$ be a zero-mean jointly Gaussian random vector with positive definite covariance matrix Σ. Letting $T = \{2, 3, \ldots, p\}$, consider the conditioned random variable $Z = (X_1 \mid X_T)$.

(a) Show that there is a vector $\theta \in \mathbb{R}^{p-1}$ such that $Z = \langle \theta, X_T \rangle + W$ where W is a zero-mean Gaussian variable independent of X_T. *Hint:* Consider the best linear predictor of X_1 given X_T.

(b) Show that $\theta = (\Sigma_{TT})^{-1} \Sigma_{T1}$, where $\Sigma_{T1} \in \mathbb{R}^{p-1}$ is the vector of covariances between X_1 and X_T.

(c) Letting $N(1) = \{j \in V | (1, j) \in E\}$ be the neighborhood set of node i, show that $\theta_j = 0$ if and only if $j \notin N(1)$. *Hint:* The following elementary fact could be useful: let A be an invertible matrix, given in the block-partitioned form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Then letting $B = A^{-1}$, we have $B_{12} = A_{11}^{-1} A_{12} \big[A_{21} A_{11}^{-1} A_{12} - A_{22} \big]^{-1}$.

Problem 7.3

Sum-product on graphs with cycles In many real-world applications, the sum-product algorithm is applied to graphs with cycles. Unlike the case of trees, the sum-product updates are no longer guaranteed to converge, or to compute the correct marginal distributions. Indeed, the results can be very surprising! As an illustration of this phenomenon, consider the Ising model (1) on the toroidal grid (see Figure 2(a)), with $\theta_s = 0$ for all $s \in V$, and $\theta_{st} = \gamma$ for all edges (s,t); call this distribution $p(x;\gamma)$, since it is parameterized by $\gamma \in \mathbb{R}$.

(a) Show that the single node marginal distributions are uniform for all choices of γ (that is, $\mathbb{P}[X_s = 1; \gamma] = \mathbb{P}[X_s = -1; \gamma] = 0.5$).

(b) Figure 2(b) shows empirically that there is some critical threshold $\gamma^* > 0$ such that the sum-product algorithm, when applied to the distribution $p(x;\gamma)$, converges to uniform marginal distributions for all $\gamma \in [0, \gamma^*)$, but produces inaccurate answers for larger γ. Using the analytical form of the sum-product updates, prove that there is some γ^* that sum-product converges from any initial condition, and computes the correct uniform marginals for all $\gamma \in [0, \gamma^*)$. *Hint:* Since each node in the model looks like every other node, it is sufficient to consider a special case of the message-passing updates, in which the message $M_{t \to s}$ along each edge $t \to s$ is the same as every other message.