CS 281A/Stat 241A Homework Assignment 2 (due September 27)

1. **Computing edge marginals in a tree.**

   Consider an undirected tree, $T = (V, E)$, where V are the nodes and E are the edges. We can use the sum-product algorithm to compute all the single node marginals $p(x_i)$. You are to provide a modification of the sum-product algorithm that will yield the edge marginals; i.e., the probabilities $p(x_i, x_j)$ for $(i, j) \in E$. What is the running time of the algorithm?

2. **Exponential family.**

   A probability distribution in the exponential family takes the following general form:

   $$p(x \mid \eta) = h(x) \exp\{\eta^T T(x) - A(\eta)\}$$

   for a parameter vector η, often referred to as the *natural parameter*, and for given functions T, A, and h.

   (a) Show that the following distributions are in the exponential family, exhibiting the T, A, and h functions in each case.

      i. Poiss(λ)—Poisson with parameter λ.

      ii. N(μ, I)—multivariate Gaussian with mean vector μ and identity covariance matrix.

      iii. Mult(θ)—multinomial with parameter vector $\theta = (\theta_1, \theta_2, \ldots, \theta_K)$. Use the fact that $\theta_K = 1 - \sum_{k=1}^{K-1} \theta_k$ and express the distribution using a $(K-1)$-dimensional parameter η.

   (b) The function $A(\eta)$ turns out to have moment-generating properties. In particular, show the following:

   $$\nabla_\eta A = \mathrm{E}[T(X)].$$

   (c) Demonstrate that the relationship in (b) holds for the three examples in part (a).

3. **The LMS algorithm.**

   The course homepage has a data set named "lms.dat" that contains twenty rows of three columns of numbers. The first two columns are the components of an input vector x and the last column is an output y value. (We will not use a constant term for this problem; thus the input vector and the parameter vector are both two dimensional.)

   (a) Solve the normal equations for these data to find the optimal value of the parameter vector. (I recommend using MATLAB or R.)

   (b) Find the eigenvectors and eigenvalues of the covariance matrix of the input vectors and plot contours of the cost function J in the parameter space. These contours should of course be centered around the optimal value from part (a).

   (c) Initializing the LMS algorithm at $\theta = 0$ plot the path taken in the parameter space by the algorithm for three different values of the step size ρ. In particular let ρ equal the inverse of the maximum eigenvalue of the covariance matrix, one-half of that value, and one-quarter of that value.

4. **Binary Markov chains.**

   A Markov chain is a directed graphical model whose structure is a chain; that is, for nodes $(X_0, X_1, \ldots, X_N)$, the edges are $(i, i+1)$ for $i \in \{0, 1, \ldots, N-1\}$. Consider a model in which $X_i \in \{0, 1\}$. Let $p(x_0)$ be uniform over 0 and 1. Let $p(x_i = a | x_{i-1} = b)$ equal q if $a \neq b$ and $(1-q)$ if $a = b$.

   (a) Find an analytical expression for the marginal distribution $p(x_0, x_1)$.

   (b) Find an analytical expression for the marginal distribution $p(x_i, x_{i+1})$ in general.

   (c) Find an analytical expression for $p(x_0 | x_N = 0)$. Hint: Consider binomial coefficients. Your answer should be an expression that depends on N and q. It may be helpful to think of the computations in terms of the elimination algorithm.

   (d) What does $p(x_0 | x_N = 0)$ converge to as $N \to \infty$? Give an intuitive explanation of why this happens.