CS281A/Stat241A Homework Assignment 4 (due 5pm, October 28, 2009)

1. **(HMMs with mixtures of Poissons)** Suppose we wish to model traffic in a network using an HMM. Consider an HMM with discrete states q_t (from a set of size m) and non-negative discrete observations y_t, where the conditional distribution of y_t given q_t is a mixture of k Poisson distributions.

   (a) Draw a graphical model for this HMM, representing the observation distributions using an additional latent variable.

   (b) Write the expected complete log likelihood for the model and identify the expectations that you need to compute in the E step of the EM algorithm.

   (c) Outline an algorithm for the E step, based on the standard alpha-beta recursion.

   (d) Write the equations for the M step.

2. **(EM for HMMs)**

   (a) Implement the EM algorithm for HMMs with the observation model of Question 2, where $m = 3$ and $k = 2$.

   (b) Use your implementation with the data on the course web site (in the file `hw4-2.data`) to find maximum likelihood parameter estimates. The data file contains a single sample path of the process; do not attempt to estimate the initial state distribution. For initial values of the parameter estimates, set the Poisson parameter $\lambda_{s,i}$ for state s and mixture component i as

   $$\lambda_{1,1} = 1, \qquad \lambda_{2,1} = 50, \qquad \lambda_{3,1} = 200,$$
   $$\lambda_{1,2} = 5, \qquad \lambda_{2,2} = 100, \qquad \lambda_{3,2} = 300.$$

   and set all of the other initial distributions to be uniform.

   What are the estimated parameters?

   Evaluate the log likelihood on the training (`hw4-2.data`) and test (`hw4-2.test`) data.

   Explain how you compute the log likelihood.

   (c) Fit a mixture of Poissons with $km = 6$ components to the same data.

   What are the parameter estimates?

   Compare its performance on the training and test data with that of the HMM.

3. **(EM for hidden trees)** Suppose that we wish to model the distribution of pollutants in a large river system, using measurements taken at a set of $2n + 1$ locations. For each location i, there is a (hidden) binary state variable $x_i \in \{0, 1\}$, and an observed real-valued measurement y_i. Suppose that y_i is conditionally independent of all other measurements y_j and states x_j, given the state x_i, and that, for all locations, this conditional distribution is Gaussian with parameters (μ_0, σ_0^2) and (μ_1, σ_1^2) for state 0 and 1, respectively. Suppose also that we model the distribution of the state variables x_i using a directed graphical model, where the graph is a directed tree with node set $V = \{1, \ldots, 2n + 1\}$ and edge set E consisting of $(i, 2i)$, $(i, 2i + 1)$ for $i = 1, \ldots, n$. Define the local conditionals as

   $$p(x_1) = \frac{1}{2} \qquad \text{for } x_1 \in \{0, 1\},$$
   $$p(x_j|x_i) = \begin{cases} \alpha & \text{if } x_i \neq x_j, \\ 1 - \alpha & \text{if } x_i = x_j, \end{cases} \qquad \text{for } (i, j) \in E \text{ and } x_j \in \{0, 1\}.$$

   (a) Show how to calculate the conditional probabilities $p(x_i|y)$ using a generalization of the HMM α-β recursion: work with the conditional probabilities $p(x_i|y_{D_i})$ and $p(y_{D_i^c}|x_i)$, where D_i denotes the descendants of i, that is, the nodes in the subtree rooted at i, and D_i^c denotes the set of all other nodes.

   (b) Derive EM updates to estimate the parameters $\big(\mu_0, \sigma_0^2, \mu_1, \sigma_1^2, \alpha\big)$.

   (c) Use these updates for the data on the course web site (line $i = 1, \ldots, 2n+1$ in the file `hw4-3.data` is measurement y_i) to estimate parameters of the model.

(d) Explain how to compute a maximum likelihood configuration for the hidden states.

(e) For the data used to estimate the parameters, draw the tree (for example, by plotting node i at location $\left(i - 3 \cdot 2^{\lfloor \log_2 i \rfloor - 1} + 1/2, -\lfloor \log_2 i \rfloor\right)$ in the plane) and indicate which nodes have the same most likely hidden state.