Problem Set 4

CS 6347

Due: 4/7/2019 by 11:59pm

Note: all answers should be accompanied by explanations for full credit. Late homeworks cannot be accepted. All submitted code **MUST** compile/run.

Problem 1: Maximum Likelihood for Colorings (100 pts)

For this problem, we will use the same factorization as we have in past assignments. However, the weights will now be considered parameters of the model that need to be learned from samples.

1. Use the belief propagation algorithm that you wrote on Problem Set 2 to perform (approximate) maximum likelihood estimation for the coloring counting problem. Your solution should be written as a MATLAB function that takes as input an $n \times n$ matrix A corresponding to the adjacency matrix of the graph G and `samples` which is an $n \times m$ k-ary matrix where $samples_{i,t}$ corresponds to observed color for vertex i in the t^{th} sample. The output should be a k-dimensional vector of weights w corresponding to the MLE parameters for the model.

   ```
   function w = colormle(A, samples)
   ```

   Your algorithm should return the correct maximum-likelihood parameters when the graph is a tree and approximate parameters otherwise. You can test your algorithm by generating a large number of samples from a tree-structured model using the Gibbs sampler that you wrote on Problem Set 3.

2. Now, suppose that some of the vertices, $L \subseteq V$, are latent variables in the model. Given m samples of the observed variables in $V \setminus L$, what is the log-likelihood as a function of the weights? Perform MLE using the EM algorithm. Your solution should be written as a MATLAB function that takes as input an $n \times n$ matrix A corresponding to the adjacency matrix of a graph G, an n-dimensional binary vector L whose non-zero entries correspond to the latent variables, and `samples` which is an $n \times m$ k-ary matrix where $samples_{i,t}$ corresponds to observed color for vertex i in the t^{th} sample (you should discard any inputs related to the latent variables). The output should be the vector of weights w corresponding to the MLE parameters for each color from the EM algorithm. Note that you should use belief propagation to approximate the counting problem in the E-step.

   ```
   function w = colorem(A, L, samples)
   ```