COMS 6998-01 Advanced Machine Learning

Assignment 4

April 14, 2002

Prof. Tony Jebara

The assignment is due on April 29th, before 4pm either in class, in my office CEPSR 605 or via email to jebara@cs.columbia.edu. If you email me the assignment, please use standard formats, i.e. send me plain text, latex, postscript, pdf or word and keep the file size reasonable. The assignment will be evaluated not just on your ability to get the right result but also your ability to provide reasoning, derivations and discussion for your answer.

1. HMMs with Mixture Model Emissions

We discussed in class HMMs where the emission probabilities given each state are either multinomial distributions (i.e. discrete outputs) or Gaussian distributions (i.e. a continuous output) as detailed in Chapter 11. A common modification to the HMMs involves using a full mixture model for the emission probabilities, $P(y_t|q_t)$. For instance, consider the case where $P(y_t|q_t)$ is a mixture of Gaussians as follows:

$$P(y_t|q_t) = \sum_{m=1}^{M} w_{m,q_t} \mathcal{N}(y_t|\mu_{m,q_t}, \Sigma_{m,q_t})$$

Where w_{m,q_t} are scalar weights that sum to 1 as m ranges from $1..M$, and μ_{m,q_t} are the Gaussian mean vectors and Σ_{m,q_t} are the Gaussian covariance matrices.

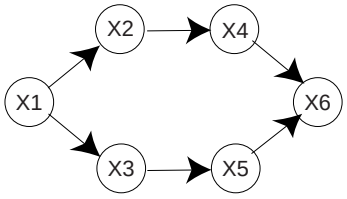
(a) Draw the graphical model for this modified HMM, and clearly show the additional latent variable nodes that are needed for the mixture model emission probability.

(b) Write the expected complete log likelihood for the model and identify the expectations that you need to compute for the E step. Then show (conceptually) an algorithm for implementing the E step above and relate it to the standard alpha-beta recursions in the simple HMM case.

(c) Write down the corresponding equations for the M step for the HMM with mixture model emissions.

2. Junction Trees

Consider the following directed graph over discrete variables $x_1, \ldots, x_6$ which have m values each:



(a) Moralize, triangulate and create a junction tree for the above graph.

(b) Are there alternative trees?

(c) Describe how the junction tree algorithm would compute $P(x_1, x_6)$.

(d) How much more expensive is it to compute $P(x_1, x_6)$ than to compute $P(x_1, x_2)$ assuming all variables each have a cardinality of m?