Inference and Representation, Fall 2015

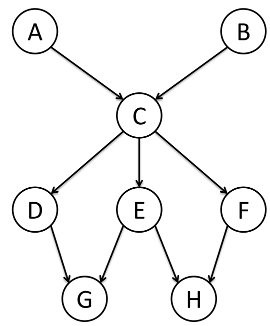
Problem Set 3: Variable Elimination & EM

Due: Weds., Oct. 14, 2015 at 4pm (as a PDF document uploaded in NYU Classes.)

Important: *See problem set policy on the course web site.*

---

1. Consider the Bayesian network shown below, and answer the following questions. You may assume that the random variables are binary-valued, i.e. take states in $\{0, 1\}$.



   (a) Moralize the Bayesian network (submit a drawing of the new graph). What edges are added?

   (b) Give a perfect elimination ordering, i.e. one that yields no fill edges.

   (c) Give an elimination ordering that results in the induced graph having ≥ 5 nodes in one (or more) cliques.

   (d) Suppose we want to compute the query $p(B = 0 \mid G = 1)$. Prove that H and F are irrelevant variables with respect to this query. That is, show that it is possible to prune the Bayesian network, removing H and F, while not changing the value of $p(B = 0 \mid G = 1)$.

   (e) Walk through the execution of the variable elimination algorithm to compute $p(B = 0 \mid G = 1)$, using as few computations as necessary (i.e., using the elimination ordering given in (b), and using the simplication given by your answer to (d)).

2. A common modification of the hidden Markov model involves using mixture models for the emission probabilities $p(\mathbf{y}_t|q_t)$, where q_t refers to the state for time t and $\mathbf{y}_t$ to the observation for time t.

   Suppose that $\mathbf{y}_t \in \mathbb{R}^n$ and that the emission distribution is given by a mixture of Gaussians for each value of the state. To be concrete, suppose that the q_t can take K discrete states and each mixture has M components. Then,

$$p(\mathbf{y}_t \mid q_t) = \sum_{j=1}^{M} b_{q_t j} \left(\frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_{q_t j}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} (\mathbf{y}_t - \mu_{q_t j})^T \Sigma_{q_t j}^{-1} (\mathbf{y}_t - \mu_{q_t j}) \right\} \right)$$

where $\mathbf{b}_i \in [0,1]^M$ denotes the mixing weights for state i ($\sum_{j=1}^{M} b_{ij} = 1$ for $i = 1, \ldots K$), $\mu_{ij} \in \mathbb{R}^n$ and $\Sigma_{ij} \in \mathbb{R}^{n \times n}$.

Let $\pi \in \mathbb{R}^K$ be the probability distribution for the initial state q_0, and $A \in \mathbb{R}^{K \times K}$ be the transition matrix of the q_t's. In this problem you will derive an EM algorithm for learning the parameters $\{b_{ij}, \mu_{ij}, \Sigma_{ij}\}$ and A, π.

(a) The EM algorithm is substantially simpler if you introduce auxiliary variables $z_t \in \{1, \ldots, M\}$ denoting which mixture component the t'th observation is drawn from.

Draw the graphical model for this modified HMM, identifying clearly the additional latent variables that are needed.

(b) Write the expected complete log likelihood for the model and identify the expectations that you need to compute in the E step. *Show all steps of your derivation.*

(c) Give an algorithm for computing the E step.

Hint: Reduce the inference problem to something you know how to do, such as sum-product belief propagation in tree-structured pairwise MRFs.

(d) Write down the equations that implement the M step.