Graphical Models 2004/05- problem set 6

January 2, 2005

Submission date 5/1

In this exrecise, you will show that the EM algorithm increases the likelihood of parameters in each iteration. We defined in class the EM algorithm for learning parameters from partial evidence:
E step Estimate $\langle n(x_i, pa_i)\rangle_{\theta^k}$ and $\langle n(x_i, y_i)\rangle_{\theta^k}$
M step Choose θ^{k+1} that maximizes the likelihood of these expected counts.

Then we described the next iterative algorithm to calculate $\min_\theta \min_q(\mathcal{F}(q,\theta))$:

- $q^{k+1} = \mathrm{argmin}_q(\mathcal{F}(q, \theta^k))$
- $\theta^{k+1} = \mathrm{argmin}_\theta(\mathcal{F}(q^{k+1}; \theta))$

1. Show that the likelihood of θ^{k+1} is not lower than the likelihood of θ^k.

2. We define $Q^k(\theta)$ as

$$\begin{aligned} Q^k(\theta) &= \langle \log(p(x, y; \theta)\rangle \\ &= \sum_x (p(x|y, \theta^k) log(p(x, y; \theta)) \end{aligned}$$

   Show that if at each stage k we compute $Q^k(\theta)$ and then define $\theta^{k+1} = \mathrm{argmax}_\theta(Q^k(\theta))$ then this is equivalent to performing the EM algorithm, as defined above.

3. Show that the iterative algorithm for minimizing Gibbs free energy, is in fact equivalent to the EM algorithm as defined in (a).