Assignment 6

Probabilistic and Unsupervised Learning

Maneesh Sahani

Due: Jan 17, 2011

1. **[30 marks] EP for the binary factor model**

   We will derive an EP algorithm to infer the marginals on the source variables in the binary latent factor model of Assignment 4.

   (a) First, write down the log-joint probability for a single observation-source pair $\log(p(\mathbf{s}, \mathbf{x}))$. Rearrange the terms to form a sum of log-factors on $\mathbf{s}$ (assuming $\mathbf{x}$ is observed), each defined either on a single source variable, or on a pair:

   $$\log(p(\mathbf{s}, \mathbf{x})) = \sum_i \log f_i(s_i) + \sum_{ij} \log g_{ij}(s_i, s_j).$$

   Do the factors correspond to a standard exponential family form? Remember that, since the sources s are binary, $s_i^2 = s_i$.

   (b) Next, derive a message passing scheme to find iterative approximations to each factor. Start your derivation from the KL divergence $\mathbf{KL}[p\|q]$ and identify clearly each time you make an approximate step. You don't need to make all of the EP approximations: which one(s) is(are) missing? Is this really an EP algorithm?

   Give the final message-passing scheme in terms of updates to the natural parameters of the site approximations. There will be two different types of update: for the $\tilde{f}_i$ and the $\tilde{g}_{ij}$ respectively.

   (c) Describe a Bayesian method for selecting K, the number of hidden binary variables in this model. Does your method pose any computational difficulties and if so how would you tackle them?

2. **[20 marks] EP for positivity constraints**

   Now consider a linear dynamical system:

   $$y_1 \sim \mathcal{N}(0, \sigma^2) \tag{1}$$
   $$y_i | y_{i-1} \sim \mathcal{N}(y_{i-1}, \sigma^2) \quad \text{for } i = 2, 3, \ldots \tag{2}$$
   $$x_i | y_i \sim \mathcal{N}(y_i, \tau^2) \quad \text{for } i = 1, 2, \ldots \tag{3}$$

   with each random variable being scalar. Suppose that we do not observe the exact values of the x_i's, but do observe that they are positive. We will now derive two different expectation propagation algorithms to approximate the resulting posterior over the y_i's.

   (a) To incorporate the positivity observations, we could include additional factors of the form:

   $$f_i(x_i) = \begin{cases} 1 & \text{if } x_i > 0, \\ 0 & \text{otherwise} \end{cases}$$

   Derive an expectation propagation algorithm to estimate the marginal distributions over all x_i and y_i in the joint distribution given by the (normalized) product of these factors with the distribution of equations (1-3). Approximate each factor with a Gaussian. You may assume access to a function which can compute the mean $E(m, s^2)$ and variance $V(m, s^2)$ of the truncated Gaussian:

$$P(z|m, v) \propto \begin{cases} e^{-\frac{(z-m)^2}{2s^2}} & \text{if } z > 0; \\ 0 & \text{otherwise} \end{cases}$$

(b) An alternative approach would be to first compute the probabilities:

$$g_i(y_i) = P(x_i > 0|y_i),$$

and then use expectation propagation to estimate the marginals of y_i's in the joint distribution given by the product of the g_i factors with the prior $P(y_1, \ldots, y_t)$ given in equations (1-2). Show that both EP algorithms are equivalent in that they should have the same fixed points.

3. **[Bonus: 50 marks]** Implement the EP algorithm you derived in question 1 above, and compare your results to those of the variational mean-field algorithm of Assignment 4.