Assignment 6

Probabilistic and Unsupervised Learning / Approximate Inference

Maneesh Sahani

Due: Noon, Monday Jan 13, 2020

Note: As usual, please attempt the main questions before the bonus ones.

1. **[20 marks] EP for sign constraints**

   Consider a linear dynamical system:

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

   with each random variable being scalar. Suppose we observe only the *signs* $s_i = \pm 1$ of the outputs x_i, rather than their magnitudes. Derive two different expectation propagation algorithms to approximate the resulting posterior over the y_is.

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

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

   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 $E(m, v^2)$ and variance $V(m, v^2)$ of the truncated Gaussian:

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

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

$$g_i(y_i) = P(\text{sign}(x_i) = s_i | y_i),$$

   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.

2. **[30 marks] EP for the binary factor model**

   Now derive an EP algorithm to infer the marginals on the source variables in the binary latent factor model of question 1 of assignment 5.

   (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:

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

Relate your result to the Boltzmann Machine. [Remember that, since the sources s are binary, $s_i^2 = s_i$.] [5 marks]

(b) Next, derive a message passing scheme to find iterative approximations $\tilde{f}_i$ and $\tilde{g}_{ij}$ 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?

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. [10 marks]

(c) Rewrite your message passing approximation to use factored approximate messages. Explain how this leads to a loopy BP algorithm. [5 marks]

(d) Describe a Bayesian method for selecting K, the number of hidden binary variables using EP. Does your method pose any computational difficulties and if so how would you tackle them? [10 marks]

3. **[Bonus: 50 marks]** Implement the EP/loopy-BP algorithm that you derived in the previous question, and compare your results to those of the variational mean-field algorithm.

4. **[Bonus 10 marks] Inconsistency of Local Marginals** Loopy belief propagation approximates the distribution over a pairwise MRF using a set of locally consistent beliefs $\{b_i(x_i), b_{ij}(x_i, x_j)\}$:

$$\sum_{x_i} b_i(x_i) = 1 \qquad \text{for all } i;$$

$$\sum_{x_i} b_{ij}(x_i, x_j) = b_j(x_j) \qquad \text{for all } i, j \text{ and } x_j.$$

(a) Give an example set of beliefs that are locally consistent but not globally consistent. That is, there is no distribution $p(\mathbf{X})$ over all variables such that

$$p(X_i = x_i) = b_i(x_i) \qquad \text{for all } i, x_i;$$

$$p(X_i = x_i, X_j = x_j) = b_{ij}(x_i, x_j) \qquad \text{for all } i, j, x_i, x_j.$$

Explain why this set of beliefs is not globally consistent. [5 marks]

(b) Construct a graphical model with specific parameter settings, such that the local marginals you came up with in the previous question form a fixed point of the loopy belief propagation algorithm run on this model. [5 marks]