Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science

6.438 Algorithms For Inference
Fall 2014

Problem Set 4

Issued: Thursday, October 2, 2014 **Due:** Tuesday, October 14, 2014

Suggested Reading: Lecture notes 8–10

Problem 4.1
Consider a binary-valued Markov process $\mathsf{x}[n]$, with transition probability distribution so that each successive value of $\mathsf{x}[n]$ has a probability of $3/4$ of taking on the opposite value to that of $\mathsf{x}[n-1]$, and where $\mathsf{x}[0]$ has equal prior probability of taking on the values 0 or 1. Suppose that we also have binary measurements $\mathsf{y}[n]$, where each of these, when conditioned on $\mathsf{x}[\cdot]$ at the same time, is independent of $\mathsf{x}[\cdot]$ at all other times, and

$$\Pr[\mathsf{y}[n] = \mathsf{x}[n] | x[n]] = \frac{7}{8} \tag{1}$$

Further, suppose we observe a value of 0 for $\mathsf{y}[n]$ for $n = 0, 1, 2$.

(a) Calculate the α and β variables that would be computed, for example, if we ran the sum-product algorithm on this particular HMM.

(b) For $n = 0, 1, 2$ compute the conditional distributions $p(\mathsf{x}[n]|\mathsf{y}[k] = 0, k = 0, 1, 2)$ and from these compute most probable individual values of $\mathsf{x}[n]$ at each of these times given these observations.

(c) Compute the most probable *trajectory* given these observations, i.e., the set of values of $\mathsf{x}[n]$ for $n = 0, 1, 2$ that maximizes the joint conditional distribution $p(\mathsf{x}[0], \mathsf{x}[1], \mathsf{x}[2]|\mathsf{y}[k] = 0, k = 0, 1, 2)$.

Problem 4.2
In this problem, you will implement the sum-product algorithm with MATLAB® and analyze the behavior of S&P 500 index over a period of time: Sept 30, 2013 to Sep 28, 2014 (http://finance.yahoo.com).

For each week, we measure the price movement relative to the previous week and denote it using a binary variable (+1 indicates up and −1 indicates down). The price movements from week 1 (the week starting on September 30, 2013) to week 52 (the week ending on September 28, 2014) are plotted in Figure 4.1.

Consider a hidden Markov model in which x_t denotes the economic state (good or bad) of week t and y_t denotes the price movement (up or down) of the S&P 500 index. We assume that $\mathsf{x}_{t+1} = \mathsf{x}_t$ with probability 0.8, and $p_{\mathsf{y}_t|\mathsf{x}_t}(y_t = +1|x_t = \text{“good”}) = p_{\mathsf{y}_t|\mathsf{x}_t}(y_t = -1|x_t = \text{“bad”}) = q$. In addition, assume that $p_{\mathsf{x}_1}(x_1 = \text{“bad”}) = 0.2$.

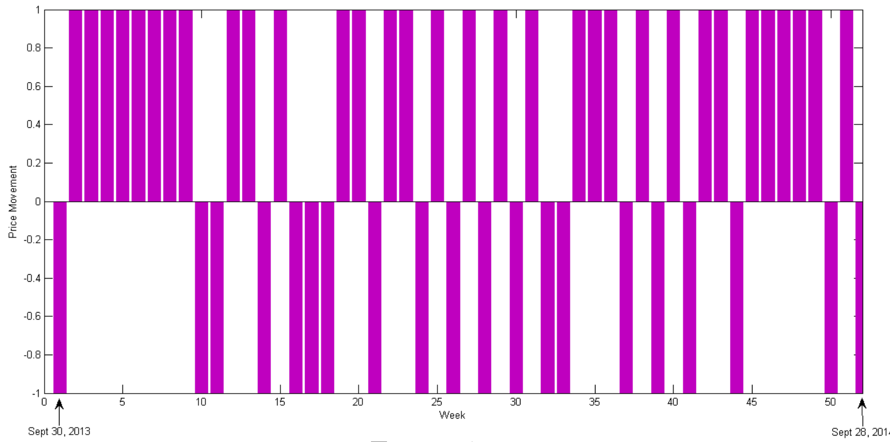

Figure 4.1

Download the file `sp500.mat` and `load` it into MATLAB. The variable `price_move` contains the binary data above. Implement the sum-product algorithm and submit a hardcopy of the code (you don't need to include the code for loading data, generating figures, etc.).

(a) Assume that $q = 0.7$. Plot $p_{\mathsf{x}_t|\mathbf{y}}(x_t = \text{"good"} \,|\mathbf{y})$ for $t = 1, 2, \ldots, 52$. What is the probability that the economy is in a good state in week 52?

(b) Repeat (a) for $q = 0.9$. Compare the results of (a) and (b).

Problem 4.3
Consider the graphical model in Figure 4.2. Draw a factor graph representing this graph-

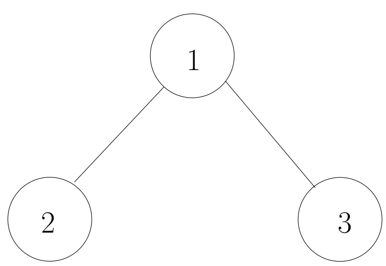

Figure 4.2

ical model. Show that you can recover the usual sum-product formulae for the graph in Figure 4.2 from the factor graph message-passing equations.

Problem 4.4
Let $G = (V, E)$ is an undirected tree graph, with factorization

$$p(\mathbf{X} = x) = \prod_{s \in V} \varphi_s(x_s) \prod_{(s,t) \in E} \psi_{st}(x_s, x_t) \tag{2}$$

Consider running sum-product algorithm on G with initialization

$$m_{s\to t}(x_t) = 1, m_{t\to s}(x_s) = 1, \forall (s,t) \in E, \forall x_s, x_t \in \mathcal{X}$$

(a) In this subproblem, we will prove by induction that the sum-product algorithm, with the parallel schedule, converges in at most diameter of the graph iterations. (Diameter of the graph is the length of the longest path.)

  (i) For $D = 1$, the result is immediate. Consider a graph of diameter D. At each time step, the message that each of the leaf nodes sends out to its neighbors is constant because it does not depend on messages from any other nodes. Construct a new undirected graphical model G' by stripping each of the leaf nodes from the original graph. How should the potentials be redefined so that the messages along the remaining edges will be the same in both graphs?

  (ii) Argue that G' has diameter strictly less than $D - 1$.

  (iii) Thus, after at most $D - 2$ time steps, the messages will all converge. Show that after "placing back" the leaf nodes into G' and running one more time step, each message will have converged to a fixed point.

(b) Prove by induction that the message fixed point m^* satisfies the following property: For any node t and $s \in N(t)$, let T_s be the tree rooted at s after the edge (s,t) is removed. Then

$$m^*_{s\to t}(x_t) = \sum_{\{x_v | v \in T_s\}} \psi(x_s, x_t) \prod_{v\in T_s} \varphi(x_v) \prod_{(i,j)\in T_s} \psi(x_i, x_j).$$

Hint: Induct on the depth of the subtree and use the definition of $m^*_{s\to t}(x_t)$.

(c) Use part (b) to show that

$$p(x_s) \propto \varphi_s(x_s) \prod_{t\in N(s)} m^*_{t\to s}(x_s)$$

for every node of the tree.

(d) Show that for each edge $(s,t) \in E$, the message fixed point m^* can be used to compute the pairwise joint distribution over (x_s, x_t) as follows:

$$p(x_s, x_t) \propto \varphi_s(x_s)\varphi_t(x_t)\psi_{st}(x_s, x_t) \prod_{u\in N(s)\setminus t} m^*_{u\to s}(x_s) \prod_{v\in N(t)\setminus s} m^*_{v\to t}(x_t).$$

Problem 4.5

Consider the graphical model in Figure 4.3.

(a) Draw a factor graph representing the graphical model and specify the factor graph message-passing equations. For this particular example, explain why the factor graph message-passing equations can be used to compute the marginals, but the sum-product equations cannot be used.

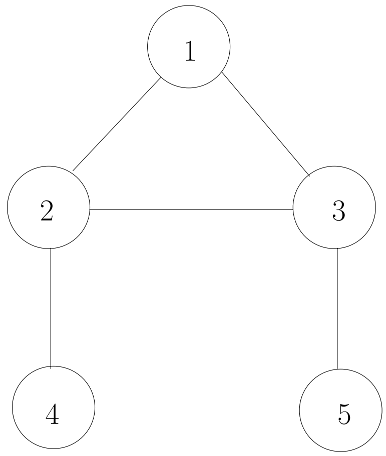

Figure 4.3

(b) Define a new random variable $x_6 = \{x_1, x_2, x_3\}$, i.e., we group variables x_1, x_2, and x_3 into one variable. Draw an undirected graph which captures the relationship between x_4, x_5, and x_6. Explain why you can apply the sum-product algorithm to your new graph to compute the marginals. Compare the belief propagation equations for the new graph with the factor graph message-passing equations you obtained in part (a).

(c) If we take the approach from part (b) to the extreme, we can simply define a random variable $x_7 = \{x_1, x_2, x_3, x_4, x_5\}$, i.e., define a new random variable which groups all five original random variables together. Explain what running the sum-product algorithm on the corresponding one vertex graph means. Assuming that we only care about the marginals for $x_1, x_2, \ldots, x_5$, can you think of a reason why we would prefer the method in part (b) to the method in this part, i.e., why it might be preferable to group a smaller number of variables together?

Problem 4.6 (Practice)

Consider a random process $x[n], n = 0, 1, 2, \ldots$ defined as follows:

$$x[0] \sim \mathcal{N}(1, 1) \tag{3}$$

$$x[n+1] = a[n]x[n] \quad , \quad n = 0, 1, 2, \ldots \tag{4}$$

where $a[n]$ is a sequence of independent, identically distributed random variables, also independent of $x[0]$, and which only take on the values ± 1, where

$$\Pr[a[n] = +1] = \Pr[a[n] = -1] = \frac{1}{2} \tag{5}$$

(a) Is $x[n]$ a Markov process? Justify your answer.

(b) What is the probability distribution for $x[1]$?

Suppose now that we have the following sequence of observations:

$$\mathsf{y}[n] = \mathsf{x}[n] + \mathsf{v}[n] \quad , \qquad n = 1, 2, \ldots \tag{6}$$

where $\mathsf{v}[n]$ is zero-mean, white, Gaussian noise, independent of $\mathsf{x}[0]$ and $\mathsf{a}[n]$, with variance of 1.

In the rest of this problem we examine the recursive computation of

$$p_{n|n}(x) \quad = \quad p_{\mathsf{x}[n]|\mathsf{y}[1],\ldots,\mathsf{y}[n]}(x)$$

and

$$p_{n+1|n}(x) \quad = \quad p_{\mathsf{x}[n+1]|\mathsf{y}[1],\ldots,\mathsf{y}[n]}(x) \ .$$

(c) Show that $p_{1|1}(x)$ is a mixture of two Gaussian distributions:

$$p_{1|1}(x) = \sum_{i=1}^{2} w_i(1|1)\mathcal{N}\big(x; \hat{x}_i(1|1), P(1|1)\big), \tag{7}$$

where the notation $\mathcal{N}\big(x; \mu, \sigma^2\big)$ indicates a Gaussian distribution with mean μ and variance σ^2 evaluated at x. Provide explicit expressions for the mean of each Gaussian distribution $\hat{x}_i(1|1)$, $i = 1, 2$, variance $P(1|1)$, and the weights $w_i(1|1), i = 1, 2$ (since these weights sum to 1, it is sufficient to specify explicit quantities to which they are proportional).

Hint: you may find it useful to write

$$\begin{aligned} p_{1|1}(x) &= p_{\mathsf{x}[1]|\mathsf{y}[1]}(x|y) \\ &= \sum_a p_{\mathsf{x}[1]|\mathsf{y}[1],\mathsf{a}[0]}(x|y,a) p_{\mathsf{a}[0]|\mathsf{y}[1]}(a|y) \end{aligned}$$

where the sum is over the two possible values, $a = \pm 1$, of $\mathsf{a}[0]$.

(d) Predicting ahead one step, the distribution $p_{2|1}(x)$ can be written as

$$p_{2|1}(x) = \sum_{i=1}^{K} w_i(2|1)\mathcal{N}\big(x; \hat{x}_i(2|1), P(2|1)\big) \ . \tag{8}$$

Specify the number of terms, K, in this sum as well as expressions for the weights, estimates, and variances in (8) in terms of the corresponding quantities in (7).

(e) In general, $p_{n|n}(x)$ and $p_{n+1|n}(x)$ will also be mixtures of Gaussian distributions.

$$p_{n|n}(x) \quad = \quad \sum_{i=1}^{K(n|n)} w_i(n|n)\mathcal{N}\big(x; \hat{x}_i(n|n), P(n|n)\big) \tag{9}$$

$$p_{n+1|n}(x) \quad = \quad \sum_{i=1}^{K(n+1|n)} w_i(n+1|n)\mathcal{N}\big(x; \hat{x}_i(n+1|n), P(n+1|n)\big) \ . \tag{10}$$

(i) What are the integers $K(n|n)$ and $K(n+1|n)$?

(ii) Specify how the parameters $P(n+1|n)$, $\hat{x}_i(n+1|n)$ and $w_i(n+1|n)$ are computed from the parameters in (9).

(iii) Determine $P(n|n)$.

(iv) Specify how the parameters at the next step $\hat{x}_i(n+1|n+1)$ and $w_i(n+1|n+1)$ are computed from the parameters in (10) and the new measurement $\mathsf{y}[n+1]$. Note that, once again, since the weights sum to 1, you can specify explicit quantities to which they are proportional.

6.438 Algorithms for Inference
Fall 2014