

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering and Computer Science

6.438 ALGORITHMS FOR INFERENCE
Fall 2014

Quiz 2

Wednesday, December 10, 2014
7:00pm–10:00pm

- This is a closed book exam, but four $8\frac{1}{2}'' \times 11''$ sheets of notes (8 sides total) are allowed.
- Calculators are **not** allowed.
- There are **3** problems.
- The problems are not necessarily in order of difficulty. We recommend that you read through all the problems first, then do the problems in whatever order suits you best.
- Record all your solutions on the exam paper. We have left enough space for each part. Extra blank sheets and stapler are available in case you need more space. You may want to first work things through on the scratch paper provided and then neatly transfer to the exam paper the work you would like us to look at. Let us know if you need additional scratch paper.
- A correct answer does not guarantee full credit, and a wrong answer does not guarantee loss of credit. You should clearly but concisely indicate your reasoning and **show all relevant work**. Your grade on each problem will be based on our best assessment of your level of understanding as reflected by what you have written in the answer booklet.
- Please be neat—we can't grade what we can't decipher!

Problem 1

Part 1: Junction Trees

Johnny, being a very bright student, is taking an advanced class on graphical models. Johnny's teacher has asked him to build a junction tree for the graph in Figure 1. In this problem, we will follow Johnny's attempts to build a Junction Tree for the given graph. On seeing the graph, Johnny immediately notices that it is not a chordal graph.

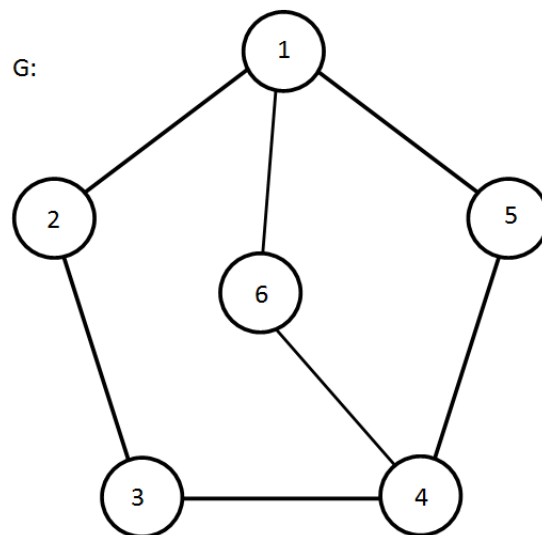


Figure 1: Graph for Problem 1

- (a) **(1 point)** Show that the graph G is non-chordal i.e. find a non-chordal cycle of length 4 or more in G .

Johnny was very attentive in class, so he knows that non-chordal graphs do not have Junction Trees. Thus, he must first add edges to G to make it chordal. Being a lazy fellow, Johnny does not wish to add too many edges. Being quite clever as well, he sees that he can triangulate the graph using only 2 edges.

- (b) (**1 point**) Triangulate G using exactly 2 edges. We will denote the resulting graph as G' . (Note: We want you to list down the edges added, as well as draw the triangulated graph.)

Johnny has now asked his friend Sarah, to help him finish the assignment. Sarah is very attentive to detail, and spells out each step of the Junction Tree algorithm clearly.

- (c) (**1 point**) Build a weighted clique graph of G' . We will call this graph G_C .

- (d) (**1 point**) Find a Maximum-weight Spanning Tree (MST) of G_C . We will call the resulting graph G_T .

(e) (**1 point**) Check whether G_T satisfies the test for junction trees, i.e. whether

$$\sum_{e \in E_{G_T}} \text{weight}(e) = \left(\sum_{C \in \mathcal{C}} |C| \right) - |V|,$$

where E_{G_T} is the set of edges in G_T .

Part 2: Sampling from Undirected Graphical Model

In the following parts, we are interested in undirected graphical models only. Moreover, assume that all the variables are binary.

- (f) (**2 points**) Give an $O(N^2)$ algorithm to sample from an N -node graphical model using the *sum-product algorithm*, given that the underlying graph is a tree.

Note: We assume that basic arithmetic operations like addition, multiplication can be done in $O(1)$ time.

- (g) (**1 point**) Extend the above algorithm to sample from any chordal graphical model G . What will be the complexity of sampling for your algorithm?
Hint: Think about part 1.

- (h) (**1 point**) Now, suppose we have a graphical model G , whose underlying graph is NOT chordal. We want to do something similar to part (g). Describe how you would do this. (In other words, how can you reduce the problem back to the one you solved in (g)?)

- (i) (**1 point**) Can you provide any useful bounds on the complexity in the case of part (h)? Justify.

Problem 2

Given an undirected graph $G = (V, \mathcal{E})$, an *independent set* I is a subset of V s.t. $\forall i, j \in I, i \neq j$, we have $(i, j) \notin \mathcal{E}$. In other words, a subset I of V is an independent set if for any edge in \mathcal{E} , at most one of the two vertices of that edge is in I .

Notice a subset of V can be represented as a sequence of binary variables $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \{0, 1\}^n$, where $n = |V|$. $x_i = 1$ if and only if node i is in the subset.

We will define a distribution over \mathbf{x} :

$$p_{\mathbf{x}}(\mathbf{x}) = \begin{cases} \frac{1}{Z} \exp\left(\sum_{i \in V} w_i x_i\right) & \text{if } \mathbf{x} \text{ corresponds to an independent set} \\ 0 & \text{otherwise} \end{cases}$$

where weights w_1, w_2, \dots, w_n are given constants, and we assume $w_i > 0, \forall i \in \{1, 2, \dots, n\}$.

We are interested in finding an independent set with the maximum total weight, i.e. we want to maximize $\sum_{i \in V} w_i x_i$ for \mathbf{x} that corresponds to an independent set. It should be obvious that this is equivalent to maximize $p_{\mathbf{x}}(\mathbf{x})$.

Part 1: Max-Product Algorithm

In this part of the problem, we apply *max-product* algorithm to solve the Weighted Maximum Independent Set problem.

- (a) **(2 points)** To be more concrete, consider the graph in Fig 2:

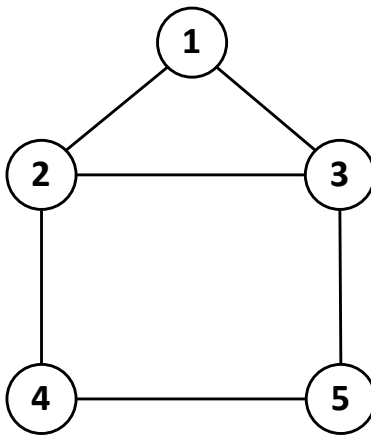


Figure 2: Graph for problem 2

Draw an undirected graphical model according to which $p_{\mathbf{x}}(\mathbf{x})$ factorizes. Specify the corresponding potential functions.

- (b) (**2 points**) Express $m_{2 \rightarrow 4}^t(x_4)$ in terms of messages at iteration $t - 1$ and the given weights w_i 's. Also, express the estimated max-marginal $\bar{p}_{x_2}^t(x_2)$ in terms of messages at iteration t and the weights. (You don't need to normalize the messages/max-marginals.)

(c) **(1 points)** At iteration t , we will make our guess of the weighted maximum independent set using the following rules:

- (1). $\hat{x}_i(\bar{p}_{x_i}^t) = 1$ if $\bar{p}_{x_i}^t(1) > \bar{p}_{x_i}^t(0)$
- (2). $\hat{x}_i(\bar{p}_{x_i}^t) = 0$ if $\bar{p}_{x_i}^t(1) < \bar{p}_{x_i}^t(0)$
- (3). $\hat{x}_i(\bar{p}_{x_i}^t) = ?$ if $\bar{p}_{x_i}^t(1) = \bar{p}_{x_i}^t(0)$

Notice we only care about the ratio of $\bar{p}_{x_i}^t(1)$ and $\bar{p}_{x_i}^t(0)$, not the actual values. Thus we define

$$\gamma_{i \rightarrow j}^t \triangleq \ln\left(\frac{m_{i \rightarrow j}^t(0)}{m_{i \rightarrow j}^t(1)}\right), \forall (i, j) \in \mathcal{E}$$

Express $\gamma_{2 \rightarrow 4}^t$ in terms of other $\gamma_{i \rightarrow j}^t$ and the weights w_i . Also express the decision rule for \hat{x}_2 in terms of $\gamma_{i \rightarrow j}^t$ and weights.

- (d) (**2 points**) For this part only, we consider a *general undirected graph*. We are interested in characterizing the fixed points of the above described max-product algorithm. Let γ^t be the set of all messages $\gamma_{i \rightarrow j}^t, \forall (i, j) \in \mathcal{E}$. Assume that γ^* is a fixed point of the max-product algorithm. Prove that $\hat{\mathbf{x}}(\gamma^*)$ does not violate the 'independent set' requirement, i.e. if $\hat{x}_i(\gamma^*) = 1$, then $\forall j \in \mathcal{N}(i), \hat{x}_j(\gamma^*) = 0$.

Hint: You can first prove the following facts:

- (i). If $\hat{x}_i(\gamma^*) = 1$, then $\gamma_{i \rightarrow j}^* > \gamma_{j \rightarrow i}^*, \forall j \in \mathcal{N}(i)$
(ii). If $\hat{x}_i(\gamma^*) = ?$, then $\gamma_{i \rightarrow j}^* \geq \gamma_{j \rightarrow i}^*, \forall j \in \mathcal{N}(i)$. Moreover, if $\gamma_{j \rightarrow i}^* \geq 0$, then equality is achieved, i.e. $\gamma_{i \rightarrow j}^* = \gamma_{j \rightarrow i}^*$

Part 2: Markov Chain Monte Carlo Methods

In this part, we consider sampling from $p_{\mathbf{x}}(\mathbf{x})$ using Metropolis-Hastings algorithm.

- (e) (**2 points**) Design a Metropolis-Hastings algorithm that samples from $p_{\mathbf{x}}(\mathbf{x})$. Make sure you explicitly describe your proposed Markov Chain and how the Metropolis-Hastings Markov Chain (i.e. the Markov Chain whose stationary distribution is $p_{\mathbf{x}}(\mathbf{x})$) is related to your proposed Markov Chain.

Note: The solution here is not unique. You only need to describe one such algorithm.

- (f) (**1 points**) Now assume you are provided with a black box that can provide independent samples from $p_{\mathbf{x}}(\mathbf{x})$. Briefly discuss how you would use the black box to approximately solve the Weighted Maximum Independent Set problem.

Note: Description of a strategy would suffice. You are not expected to come up with any theoretical guarantee of the performance of your strategy.

Problem 3

The purpose of this question is to understand how graphical model learning can be very useful in seemingly unrelated applications.

Consider a point-of-sales system like Square, Inc. (remember the payment system that works with iPhones). A candy shop uses this system for processing all of its transactions. Naturally, Square collects data about these transactions over time. Based on it, to provide it's value, it would like to inform the candy shop when it's close to running out of its inventory so that shop can order new candies without turning its customers away. The problem is that Square does not know exactly how many candies are there in a store at any point of time; it only knows whether a candy is sold or not. In what follows, we'll go through a stylized version of this question that can help Square infer the number of candies in a store at any given time.

Let time be indexed by $t \in \{0, 1, \dots, \}$. Let $X_t \in \{0, 1, \dots, C\}$ represent the number of candies in the store at time t ; C being the maximum number of candies in the store. Let $Y_t \in \{0, 1\}$ represent whether Square observes a transaction of purchase of a candy from the store at time t : $Y_t = 1$ if a candy is purchased at time t and 0 otherwise (we shall assume that no one ever purchases more than one candy).

The shop owner operates as follows. At a given time t , if $X_t = 0$ (i.e. no more candies in stock), with probability p , s/he re-fills the shop by ordering C candies (i.e. $X_{t+1} = C$); with probability $1 - p$, nothing happens (i.e. $X_{t+1} = 0$). At any given time, if $X_t = 0$, then naturally $Y_t = 0$ as there is nothing to purchase. However, if $X_t \geq 1$, then $Y_t = 1$ with probability q . Obviously, when $Y_t = 1$, the remaining stock decreases by 1 i.e. $X_{t+1} = X_t - 1$.

The question of interest for Square is to learn parameters p, q given the knowledge of C and above behavior of shop owner as well as customers. (Questions start on next page.)

(a) Suppose $X_0 = C$ and $C \geq 2$. We wish to find an estimate of q from observations $Y_t, 0 \leq t \leq T$, for T large enough.

(1) (**1 points**) For this part, assume that the stock is always full i.e. $C = +\infty$. Describe how you would estimate q using $Y_t, 0 \leq t \leq T$ so that as $T \rightarrow \infty$ your estimate finds the correct value of q .

- (2) (**2 points**) Now assume C is finite. Describe how you would estimate q using $Y_t, 0 \leq t \leq T$ so that as $T \rightarrow \infty$ your estimate finds the correct value of q .

(b) Now, suppose $X_0 = 0$ and $C \geq 2$. We wish to find an estimate of p from observations $Y_t, 0 \leq t \leq T$, for T large enough.

(1) (**1 points**) Suppose $q = 1$ i.e. an item is always sold when the stock is non-empty. Describe how you would estimate p using $Y_t, 0 \leq t \leq T$ so that as $T \rightarrow \infty$ your estimate finds the correct value of p .

- (2) (**2 points**) Describe how you would estimate p for arbitrary q so that as $T \rightarrow \infty$ your estimate finds the correct value of p .

- (c) Now suppose X_0 is distributed uniformly at random on $\{0, \dots, C\}$ at time 0, $C \geq 2$. We wish to obtain good estimates for p, q . As this part is more complex, it will be useful to write down a graphical model description.
- (1) (**2 points**) Write down a Hidden Markov Model description of the above problem.

- (2) (**2 points**) Describe how you would use EM algorithm to estimate p, q from the observations.

MIT OpenCourseWare
<http://ocw.mit.edu>

6.438 Algorithms for Inference
Fall 2014

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.