Homework 5
CS 221 (Autumn 2012–2013)

Submission instructions: Write your answers in one PDF file named `hw5.pdf`. Remember to include your name and SUNet ID. Copy the PDF file onto `corn.stanford.edu`, ssh in to the machine, type `/usr/class/cs221/WWW/submit`, and follow the instructions.

1. CSP Warm Up (*7 points*)

A useful step before implementing any algorithm is to walk through a simple example to debug your own understanding. This problem asks you to walk through variable elimination to make sure you understand its intricacies.

Suppose you have 5 variables $X_1, \dots, X_5$. For each variable X_j, the domain is: $X_j \in \{1, \dots, 5\}$. They are subject to the following constraints:

- $X_2 > X_1$
- $X_4 = X_3 + X_2$
- $X_5 > X_4$

a. (*1 point*) Recall that conditioning on a variable X_i creates additional restrictions on each variable X_j in the Markov blanket of X_i, which effectively prunes the domain of X_j. (For example, conditioning on $X_1 = 1$ would result in $X_2 \in \{2, 3, 4, 5\}$.) Given the original CSP, only condition on $X_4 = 4$. What are the domains for X_1, X_2, X_3, X_5?

b. (*1 point*) Start with the original factor graph (not conditioned on X_4). Applying generalized arc consistency on an arbitrary k-ary constraint f and a variable X_i sets the domain of X_i to be the values v for which there is some assignment to the other $k-1$ variables along with $X_i = v$ which satisfies f (see the AI book, page 210).

Suppose we iteratively apply generalized arc consistency on each constraint until convergence (no more domains can be reduced). What are the resulting domains of each variable?

c. (*1 point*) From the original factor graph, eliminate X_4 to produce a new constraint f_{new}. Apply generalized arc consistency on just f_{new} and each of its dependent variables. What are the resulting domains of these variables?

d. (*1 point*) Starting with the original factor graph, eliminate X_4 and then X_2. Apply generalized arc consistency on the new constraints and their dependent variables. What are the resulting domains?

e. (*1 point*) Starting with the original factor graph, eliminate X_2 and then X_4. Apply generalized arc consistency on the new constraints and their dependent variables. What are the resulting domains? Did your results change?

f. (*2 points*) Consider the following two operations:

- Enforcing arc consistency on the constraint $X_2 > X_1$ and variable X_2.
- Eliminating X_1.

Briefly compare the results of these two approaches. Its not necessary to write out the results, we're more concerned that you see a similarity between the two approaches. (Note that this similarity may not hold for general factor graphs.)

2. Grids (*7 points*)

Suppose we have a Markov network consisting of a $m \times n$ grid of variables $\{X_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}$. Assume $n > m$ and both m and n are even. There is a factor between every pair variables which are Manhattan distance one away (X_{ij} and $Xi'j'$ if $|i - i'| + |j - j'| = 1$).

a. (*2 points*) What is the tree-width of this Markov network? Describe the elimination order that yields this tree-width (the maximum arity of new factors produced is exactly the tree-width).

b. (*2 points*) Create a new chain-structured Markov network with n variables so that there is an easy one-to-one mapping between assignments in the old Markov network and the new Markov network. What are the new variables and factors as a function of the old variables and factors?

c. (*1 point*) Suppose we condition on all variables X_{ij} for which $i \bmod 2 = 0$. What is the tree-width of the resulting Markov network?

d. (*1 point*) Suppose we condition on all variables X_{ij} for which $i + j \bmod 2 = 0$. What is the tree-width of the resulting Markov network?

e. (*1 point*) What is the smallest number of variables we must condition on for the resulting Markov network to have tree-width 1? You must provide the exact number. *Note we only expected a slight improvement on 2(c), but if you have a solution that is more optimal, you are eligible for 1-3 points of extra credit capped at a perfect score on the assignment. If you attempt the EC, please include an image showing your elimination pattern to make grading easy. We won't have time to read proofs.*

3. MAP versus marginal inference (*4 points*)

Suppose we have a Markov network with variables $X = (X_1, \ldots, X_n)$ and factors $f_1, \ldots, f_m$. Assume that the domain of each variable is $\text{Domain}_i = \{0, 1\}$ and that for each factor f_j and all assignments x, we have $f_j(x) > 0$. Given T be a positive number (which we will call the temperature), define a Markov network with weight $\text{Weight}_T(x) = \prod_{j=1}^{m} f_j(x)^{1/T}$ and $\mathbb{P}_T(X = x)$ be the corresponding probability distribution. Note that x is an assignment to all the variables in X. Let S be the set of maximum weight assignments x (that is $\text{Weight}_T(x) = \max_{x'} \text{Weight}_T(x')$).

Note that as T goes to ∞ ($T \to \infty$), we have that $1/T$ converges to zero ($1/T \to 0$).

a. (*2 points*) Prove that as the temperature $T \to \infty$, $\mathbb{P}_T(X = x)$ converges to $1/2^n$.

b. (*2 points*) Prove that as the temperature $T \to 0$, $\mathbb{P}_T(X = x)$ converges to $1/|S|$ if $x \in S$ and converges to 0 if $x \notin S$.

4. Spelling correction (*7 points*)

You are given a sentence, which is a sequence of n (possibly misspelled) words $w_1, \dots, w_n$, where each w_i is a sequence of letters. You have a dictionary D of possible words. Define the edit distance $D(w, w')$ as the minimum number of insertions, deletions, or replacements required to convert w into w'. Futhermore, you have a function $\text{Fluency}(w, w', w'')$, which measures how likely the words w, w', w'' are to appear next to each other in that order. Your goal is to find a sequence of real words from the dictionary $t_1, \dots, t_n$ that minimizes the sum of the edit distances $D(t_i, w_i)$ plus the sum of $-\text{Fluency}(t_i, t_{i+1}, t_{i+2})$.

a. (*3 points*) Formulate this problem as a weighted CSP—what are the variables, their domains, and factors? You should have n variables, and no factor should have arity more than 3. What is the tree-width of this weighted CSP?

b. (*2 points*) Change the domains of the n variables and redefine the factors so that no factor has arity more than 2. What happens to the size of the state space (consider both the number of variables and the domain of each variable)? What is the tree-width of this weighted CSP?

c. (*2 points*) Suppose we add an additional requirement that at least one of the words of the spelling-corrected sentence must be a verb (let $V \subset D$ be the set of verbs). Add $O(n)$ variables and factors with arity at most 3 to satisfy this requirement. What is the tree-width of this weighted CSP?