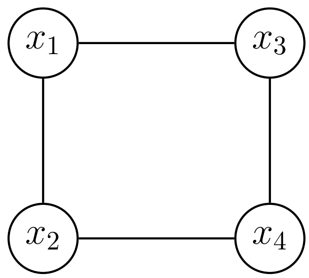
University of Illinois Spring 2018

CS 446: Machine Learning
Homework 7

Due on Tuesday, March 6, 2018, 11:59 a.m. Central Time

1. [**4 points**] Inference Methods for Discrete Markov Random Fields

For this problem, consider the following Markov Random Field, where each node can be assigned one of the values in $\{1, 2, 3, 4, 5\}$:



(a) To conduct MAP inference on this graph using exhaustive search, how many configurations must be checked?

Your answer:

$$5^4$$

(b) Can MAP inference be run on this graph using a dynamic programming algorithm? Why or why not?

Your answer: No, because the MRF has cycles in it, e.g. it is not a tree.

(c) To run MAP inference on this graph using loopy belief propagation, how many messages must be computed?

Your answer: Let's start by noticing that

$$\mathcal{R} = \{\{1\}, \{2\}, \{3\}, \{4\}, \{1,2\}, \{1,3\}, \{3,4\}, \{2,4\}\}$$

Since the number of messages is the number of Langrangian multipliers and we have one multiplier for each of the constraints $\sum_{\mathbf{y}_p \mathbf{y}_r} b_p(\mathbf{y}_b) = b_r(\mathbf{y}_r)$.
Since $P(r \in \{\{1,2\}, \{1,3\}, \{3,4\}, \{2,4\}) = \varnothing$ only the other values of r will generate a constraint.

$$\mathbf{y}_r \in \{1, 2, 3, 4, 5\} \forall r \in \{\{1\}, \{2\}, \{3\}, \{4\}\}$$

Finally, note that there are two parents for each of r, that is $|P(r)| = 2$.
Then there are $4 \times 5 \times 2 = 40$ constraints and, therefore, 40 messages.

2. **[7 points]** ILP Inference formulation in Discrete Markov Random Fields

(a) Suppose we have two variables $x_1 \in \{0,1\}$ and $x_2 \in \{0,1\}$ and their local evidence functions $\theta_1(x_1)$ and $\theta_2(x_2)$ as well as a pairwise function $\theta_{1,2}(x_1,x_2)$. Using this setup, inference solves $\arg\max_{x_1,x_2} \theta_1(x_1) + \theta_2(x_2) + \theta_{1,2}(x_1,x_2)$. Using

$$\theta_1(x_1) = \begin{cases} 1 & \text{if } x_1 = 0 \\ 2 & \text{otherwise} \end{cases} \qquad \theta_2(x_2) = \begin{cases} 1 & \text{if } x_2 = 0 \\ 2 & \text{otherwise} \end{cases}$$

$$\theta_{1,2}(x_1,x_2) = \begin{cases} -1 & \text{otherwise} \\ 2 & \text{if } x_1 = 0 \ \&\ x_2 = 1 \end{cases}$$

what is the integer linear programming formulation of the inference task? Make the cost function and constraints explicit for the given problem, i.e., do not use a general formulation.

Your answer: Lets first find the function of b_r that we are trying to maximize.

$$
\begin{aligned}
\begin{bmatrix} b_1(0) \\ b_1(1) \\ b_2(0) \\ b_2(1) \\ b_{1,2}(0,0) \\ b_{1,2}(0,1) \\ b_{1,2}(1,0) \\ b_{1,2}(1,1) \end{bmatrix}^\intercal
\begin{bmatrix} \theta_1(0) \\ \theta_1(1) \\ \theta_2(0) \\ \theta_2(1) \\ \theta_{1,2}(0,0) \\ \theta_{1,2}(0,1) \\ \theta_{1,2}(1,0) \\ \theta_{1,2}(1,1) \end{bmatrix}
&= \begin{bmatrix} b_1(0) \\ b_1(1) \\ b_2(0) \\ b_2(1) \\ b_{1,2}(0,0) \\ b_{1,2}(0,1) \\ b_{1,2}(1,0) \\ b_{1,2}(1,1) \end{bmatrix}^\intercal
\begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \\ -1 \\ 2 \\ -1 \\ -1 \end{bmatrix} \\
&= b_1(0) + 2b_1(1) + b_2(0) + 2b_2(1) - \\
&\quad b_{1,2}(0,0) + 2b_{1,2}(0,1) - b_{1,2}(1,0) - b_{1,2}(1,1) \\
&:= A_b
\end{aligned}
$$

Then the inference task can be rewritten as

$$
\max_b A_b \quad \text{s.t.} \begin{cases} b_r(\mathbf{y}_r) \geq 0 & \forall r, \mathbf{y}_r \\ \sum_{\mathbf{y}_r} b_r(\mathbf{y}_r) = 1 & \forall r \\ \sum_{\mathbf{y}_p/\mathbf{y}_r} b_p(\mathbf{y}_p) = b_r(\mathbf{y}_r) & \forall r, p \in P(r), \mathbf{y}_r \end{cases}
$$

Now let's write the constraints explicitly.

$$
\max_b A_b \quad \text{s.t.} \begin{cases} b_1(0) \geq 0 \\ b_1(1) \geq 0 \\ b_2(0) \geq 0 \\ b_2(1) \geq 0 \\ b_{1,2}(0,0) \geq 0 \\ b_{1,2}(0,1) \geq 0 \\ b_{1,2}(1,0) \geq 0 \\ b_{1,2}(1,1) \geq 0 \\ b_1(0) + b_1(1) = 1 \\ b_2(0) + b_2(1) = 1 \\ b_{1,2}(0,0) + b_{1,2}(0,1) + b_{1,2}(1,0) + b_{1,2}(1,1) = 1 \\ b_{1,2}(0,0) + b_{1,2}(0,1) = b_1(0) \\ b_{1,2}(1,0) + b_{1,2}(1,1) = b_1(1) \\ b_{1,2}(0,0) + b_{1,2}(1,0) = b_2(0) \\ b_{1,2}(0,1) + b_{1,2}(1,1) = b_2(1) \end{cases}
$$

(b) What is the solution (value and argument) to the program in part (a).

Your answer: Note that the new cost function is

$$b_1(0) + 2b_1(1) + b_2(0) + 2b_2(1) - b_{1,2}(0,0) + 2b_{1,2}(0,1) - b_{1,2}(1,0) - b_{1,2}(1,1)$$

By simply looking at the previous item we can see that the maximum value is achieved when letting $b_{1,2}(0,1) = 1$. Then the solution is The solution is:

$$\begin{aligned} b_1(0) &= 1 \\ b_1(1) &= 0 \\ b_2(0) &= 0 \\ b_2(1) &= 1 \\ b_{1,2}(0,0) &= 0 \\ b_{1,2}(0,1) &= 1 \\ b_{1,2}(1,0) &= 0 \\ b_{1,2}(1,1) &= 0 \end{aligned}$$

And then the value of our cost function is:

$$1 + 2(0) + 0 + 2(1) - 0 + 2(1) - 0 - 0 = 1 + 2 + 2 = 5$$

(c) Why do we typically not use the integer linear program for reasonably sized MRFs?

Your answer: Because it becomes very slow for large programs. The reason being that it still has to do exhaustive search on the individual restrictions' arguments. e.g. $b_r(\mathbf{y}_r)$. Which, in large programs, may still be too large. Either $|r|$ is too large for some r or there are too many restrictions, $|\mathcal{R}|$, is too large.