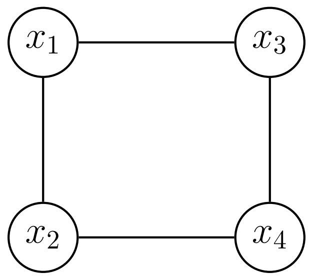
University of Illinois Spring 2018

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:

   $$\#\text{configurations} = 5^4 = 625$$

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

   Your answer: No, because the graph is cyclic, implying the graph cannot be represented as a dependency tree.

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

   Your answer: $\#\text{messages} = \|\lambda\_p, r, y_r\|$; where p is the number of parents 4, r is the number of regions 8 and y_r is the number of labels 5. $\#\text{messages} = 4 * 5 * 8 = 160$

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:

$$
\begin{aligned}
\text{define} \quad & r \in \{\{1\},\{2\},\{1,2\}\} \\
& x_1 \in \{0,1\} \\
& x_2 \in \{0,1\} \\
& x_{1,2} \in \{(0,0),(0,1),(1,0),(1,1)\}
\end{aligned}
$$

$$
\text{maximize} \quad \begin{bmatrix} b_1(0) \\ b_1(1) \\ b_2(0) \\ b_2(1) \\ b_{1,2}(0,0) \\ b_{1,2}(0,1) \\ b_{1,2}(1,0) \\ b_{1,2}(1,1) \end{bmatrix}^T \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \\ -1 \\ 2 \\ -1 \\ -1 \end{bmatrix}
$$

$$
\begin{aligned}
\text{subject to} \quad & b_r(x_r) \in \{0,1\} \quad r \\
& \sum_{x_1} b_1(x_1) = 1 \\
& \sum_{x_2} b_1(x_2) = 1 \\
& \sum_{x_{1,2}} b_{1,2}(x_{1,2}) = 1 \\
& b_{1,2}(0,0) + b_{1,2}(0,1) = b_1(0) \\
& b_{1,2}(1,0) + b_{1,2}(1,1) = b_1(1) \\
& b_{1,2}(0,0) + b_{1,2}(1,0) = b_2(0) \\
& b_{1,2}(0,1) + b_{1,2}(1,1) = b_2(1)
\end{aligned}
$$

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

Your answer:

$$\mathbf{b} = \begin{bmatrix} b_1(0) \\ b_1(1) \\ b_2(0) \\ b_2(1) \\ b_{1,2}(0,0) \\ b_{1,2}(0,1) \\ b_{1,2}(1,0) \\ b_{1,2}(1,1) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{b}^T \theta(\mathbf{x}) = 5$$

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

Your answer: As the number of features examined in an MRF increases and the set of labels increases, the number of constraints generated becomes computationally infeasible.