Lecture: Dr. Tao Wu
Exercises: Yuesong Shen, Zhenzhang Ye
Summer Semester 2019

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 8

Room: 02.09.023

Wednesday, 10.07.2019, 12:15 - 14:00

MAP infer and BN learn (Due: 08.07) (14+6 Points)

Exercise 1 (8 Points)**.** (Graphcut) Consider a simple MRF with 2 connected binary nodes and the following energies:

$$E_1 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}, E_2 = \begin{bmatrix} 8 \\ 7 \end{bmatrix}, E_{1,2} = \begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix} \tag{1}$$

1) Construct new energies $\tilde{E}_1, \tilde{E}_2, \tilde{E}_{1,2}$ that can be used for flow graph while keeping the underlying joint distribution invariant.

2) Draw the flow graph and the minimum cut that you have found and conclude with the labeling of the labels that minimizes the total energy.

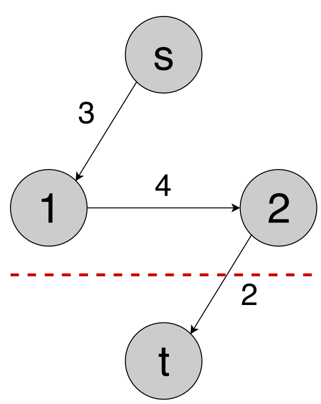
Solution.

1) Since $E_{1,2} = \begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 2 & 2 \end{bmatrix} + \begin{bmatrix} 0 & 4 \\ 0 & 0 \end{bmatrix}$, we have that

$$\tilde{E}_{1,2} = \begin{bmatrix} 0 & 4 \\ 0 & 0 \end{bmatrix}, \tilde{E}_1 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \tilde{E}_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \tag{2}$$

defines an equivalent energy parametrization which has an appropriate form for constructing a flow graph.

2) the flow graph along with the minimum cut is as follows:



Thus the minimum energy is achieved when both nodes take their first label.

Exercise 2 (6 Points)**.** (MLE of BN) In the lecture we have seen that for fully observed Bayesian network, the MLE estimate of tabular CPD parameters has a close form of ratio of counts $\frac{\#(x_{F_i})}{\#(x_{Pa(i)})}$, where $F_i = \{i\} \cup Pa(i)$. You are asked to derive this result from the general problem setting as follows:

$$\operatorname{argmin}_{\theta \in \mathbb{R}_+^{|\theta|}} -\frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{V}} \sum_{x' \in \mathcal{D}} \sum_{x_{F_i}} \log \theta(x_i | x_{Pa(i)}) \mathbf{1}\{x_{F_i} = x'_{F_i}\} \tag{3}$$

$$\text{s.t. } \forall (i, x), \sum_{x_i} \theta(x_i | x_{pa(i)}) = 1 \tag{4}$$

Hint: introducing Lagrangian will help solving this problem.

Solution. First of all, notice that the problem can be decoupled for each CPD component of the Bayesian network. We can thus consider for each i the following problem:

$$\operatorname{argmin}_{\theta_{F_i} \in \mathbb{R}_+^{|\theta_{F_i}|}} -\frac{1}{|\mathcal{S}|} \sum_{x' \in \mathcal{D}} \sum_{x_{F_i}} \log \theta_{F_i}(x_i | x_{Pa(i)}) \mathbf{1}\{x_{F_i} = x'_{F_i}\} \tag{5}$$

$$\text{s.t. } \forall x_{Pa(i)}, \sum_{x_i} \theta_{F_i}(x_i | x_{pa(i)}) = 1 \tag{6}$$

Since

$$-\frac{1}{|\mathcal{S}|} \sum_{x' \in \mathcal{D}} \sum_{x_{F_i}} \log \theta_{F_i}(x_i | x_{Pa(i)}) \mathbf{1}\{x_{F_i} = x'_{F_i}\} = -\frac{1}{|\mathcal{S}|} \sum_{x_{F_i}} \#(x_{F_i}) \log \theta_{F_i}(x_i | x_{Pa(i)}), \tag{7}$$

the Lagrangian of the decouple problem is thus

$$\mathcal{L}(\theta_{F_i}, \lambda) = -\frac{1}{|\mathcal{S}|} \sum_{x_{F_i}} \#(x_{F_i}) \log \theta_{F_i}(x_i | x_{Pa(i)}) + \sum_{x_{Pa(i)}} \lambda_{x_{Pa(i)}} (1 - \sum_{x_i} \theta_{F_i}(x_i | x_{pa(i)})). \tag{8}$$

Setting the partial derivatives to 0 givens us

$$\theta_{F_i}(x_i | x_{Pa(i)}) = -\frac{\#(x_{F_i})}{|\mathcal{S}| \lambda_{x_{Pa(i)}}} \tag{9}$$

$$\sum_{x_i} \theta_{F_i}(x_i | x_{pa(i)}) = 1 \tag{10}$$

thus $\theta_{F_i}(x_i | x_{Pa(i)}) = \frac{\#(x_{F_i})}{\sum_{x_i} \#(x_{F_i})} = \frac{\#(x_{F_i})}{\#(x_{Pa(i)})}$.

Exercise 3 (6 Points)**.** (Semiring) Interestingly, sum-product and max-sum operators are deeply linked with the semiring algebraic structure. Look up the mathematical definition of semiring on e.g. Wikipedia, then verify that $([0, +\infty), +, \times)$, $([0, +\infty), \max, \times)$ and $([-\infty, +\infty), \max, +)$ are all semirings.

Solution.

- $([0,+\infty),+,\times)$ is a semiring since
  - $([0,+\infty),+)$ is a commutative monoid with identity 0: $\forall a,b,c \in [0,+\infty), (a+b)+c = a+(b+c), 0+a = a+0 = a, a+b = b+a$;
  - $([0,+\infty),\times)$ is a monoid with identity 1: $\forall a,b,c \in [0,+\infty), (a\times b)\times c = a\times(b\times c), 1\times a = a\times 1 = a$;
  - left and right distributivity: $\forall a,b,c \in [0,+\infty), a\times(b+c) = a\times b + a\times c, (a+b)\times c = a\times c + b\times c$;
  - multiplication by zero: $\forall a \in [0,+\infty), 0\times a = a\times 0 = 0$.
- $([0,+\infty),\max,\times)$ is a semiring since
  - $([0,+\infty),\max)$ is a commutative monoid with identity 0: $\forall a,b,c \in [0,+\infty), \max(\max(a,b),c) = \max(a,\max(b,c)), \max(0,a) = \max(a,0) = a, \max(a,b) = \max(b,a)$;
  - $([0,+\infty),\times)$ is a monoid with identity 1: $\forall a,b,c \in [0,+\infty), (a\times b)\times c = a\times(b\times c), 1\times a = a\times 1 = a$;
  - left and right distributivity: $\forall a,b,c \in [0,+\infty), a\times\max(b,c) = \max(a\times b, a\times c), \max(a,b)\times c = \max(a\times c, b\times c)$;
  - multiplication by zero: $\forall a \in [0,+\infty), 0\times a = a\times 0 = 0$.
- $([-\infty,+\infty),\max,+)$ is a semiring since
  - $([-\infty,+\infty),\max)$ is a commutative monoid with identity $-\infty$: $\forall a,b,c \in [-\infty,+\infty), \max(\max(a,b),c) = \max(a,\max(b,c)), \max(-\infty,a) = \max(a,-\infty) = a, \max(a,b) = \max(b,a)$;
  - $([-\infty,+\infty),+)$ is a monoid with identity 0: $\forall a,b,c \in [-\infty,+\infty), (a+b)+c = a+(b+c), 0+a = a+0 = a$;
  - left and right distributivity: $\forall a,b,c \in [-\infty,+\infty), a+\max(b,c) = \max(a+b,a+c), \max(a,b)+c = \max(a+c,b+c)$;
  - multiplication by zero: $\forall a \in [-\infty,+\infty), -\infty + a = a + (-\infty) = -\infty$.