Advanced Machine Learning: HW-3

Uri Avron [uriavron@gmail.com] [308046994]
Jonathan Somer [jonathan.somer@gmail.com] [307923383]
Matan Harel [matan.harel.mh@gmail.com] [302695721]

May 3, 2018

1. Exact Solution Using The Local Marginal Polytope Approximation

Setting:

- n random variables $X_1, .., X_n$
- Graph E
- MRF defined by:
  1. $\forall ij \in E : \theta_{ij}(x_i, x_j) = \begin{bmatrix} 0 & 0 \\ 0 & s_{ij} \end{bmatrix}$ and $s_{ij} > 0$
  2. $\forall ij \in E : \theta_i(x_i) = \begin{bmatrix} 0 \\ s_i \end{bmatrix}$ and $s_i \neq 0$

(a) Show that $max_{\boldsymbol{\mu} \in M_L} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$ is equivalent to the following LP:

$$\text{Maximize: } f(\boldsymbol{\tau}) = \sum_i s_i \tau_i + \sum_{ij} s_{ij} \tau_{ij}$$

With respect to constraints:

$$\forall ij \in E : \tau_{ij} \geq 0 \tag{1}$$
$$\forall i : \tau_i \geq 0 \tag{2}$$
$$\forall ij \in E : \tau_{ij} \leq \tau_i \tag{3}$$
$$\forall ij \in E : \tau_{ij} \leq \tau_j \tag{4}$$
$$\forall ij \in E : \tau_{ij} \geq \tau_i + \tau_j - 1 \tag{5}$$

We shall start with the local marginal polytope (LMP) relaxation:

$$max_{\boldsymbol{\mu}\in M_L}\boldsymbol{\mu}\cdot\boldsymbol{\theta} = max_{\boldsymbol{\mu}}\sum_{ij}\sum_{x_i,x_j}\mu_{ij}(x_i,x_j)\theta_{ij}(x_i,x_j)+\sum_i\sum_{x_i}\mu_i(x_i)\theta_i(x_i)$$

With constraints:

$$\boldsymbol{\mu}\geq 0 \tag{6}$$

$$\sum_{x_i}\mu_i(x_i)=1 \tag{7}$$

$$\sum_{x_i,x_j}\mu_{ij}(x_i,x_j)=1 \tag{8}$$

$$\forall ij\in E, x_j:\sum_{x_i}\mu_{ij}(x_i,x_j)=\mu_j(x_j) \tag{9}$$

$$\forall ij\in E, x_i:\sum_{x_j}\mu_{ij}(x_i,x_j)=\mu_i(x_i) \tag{10}$$

First of all we shall assign $\boldsymbol{\theta}$ its values in our case, noting that its value is 0 in all cases but $(1,1),(1)$

$$max_{\boldsymbol{\mu}\in M_L}\boldsymbol{\mu}\cdot\boldsymbol{\theta} = max_{\boldsymbol{\mu}}\sum_{ij}\mu_{ij}(1,1)\theta_{ij}(1,1)+\sum_i\mu_i(1)\theta_i(1)$$

$$= max_{\boldsymbol{\mu}}\sum_{ij}\mu_{ij}(1,1)s_{ij}+\sum_i\mu_i(1)s_i$$

We now rename $\forall i:\mu_i(1):=\tau_i$ and $\forall ij:\mu_{ij}(1,1)=\tau_{ij}$

$$max_{\boldsymbol{\tau}}\sum_{ij}\tau_{ij}s_{ij}+\sum_i\tau_i s_i$$

This is the exact function we are maximizing in the LP. That is, an optimal solution for the LP problem is also an optimal solution for the LMP problem. Thus it is enough to find optimal values for the τ that appear in the new formulation. But we must take into consideration that:

1. We must be sure that the new LP constraints are not "too tight" so that we are not missing any optimal assignment to τ. We will show this by showing that the new LP constraints can be derived from the LMP constraints.

2. We must also be sure that the constraints are not "too loose", that is: that we find some optimal τ that can satisfy the original constraints on μ. We will prove this by showing that any optimal assignment to τ which satisfies LP can be extended to a valid assignment to μ according to LMP.

We will now show that the 5 constraints in the new problem, denoted by LP,can be derived from the 5 constraints in the LMP relaxation, denoted by LMP. Denote the i^{th} rule in LP by: $LP(i)$, and similarly with LMP.

(1) LP can be derived from LMP:

- $LP(1), LP(2)$ directly result from $LMP(1)$ and the way we defined $\boldsymbol{\tau}$
- $LP(3), LP(4)$ result from $LMP(1), LMP(4), LMP(5)$. To show this we will assume by contradiction and w.l.o.g that $LP(3)$ does not hold for some $ij \in E$, that is: $\tau_{ij} > \tau_i$. Note that from the way we defined τ_{ij}, τ_i we have $\mu_{ij}(1,1) > \mu_i(1)$ From $LMP(1)$:

$$\sum_{x_j} \mu_{ij}(1, x_j) \geq \tau_{ij} > \tau_i = \mu_i(1)$$

  In contradiction to $LMP(5)$.

- We will now show $LP(5)$ results from $LMP(1), LMP(3), LMP(4), LMP(5)$. Let there be some $ij \in E$. By definition:

$$\tau_i + \tau_j = \mu_i(1) + \mu_j(1)$$

  From $LMP(4), LMP(5)$:

$$= \sum_{x_j} \mu_{ij}(1, x_j) + \sum_{x_i} \mu_{ij}(x_i, 1)$$

  Add $\mu_{ij}(0,0)$, $LMP(1)$:

$$\leq \sum_{x_i x_j} \mu_{ij}(x_i, x_j) + \mu_{ij}(1,1)$$

  Definition $+$ $LMP(3)$:

$$= 1 + \tau_{ij}$$

  Subtracting 1 from both sides of the inequality we arrive at:

$$\tau_i + \tau_j - 1 \leq \tau_{ij}$$

(2) Any optimal LP-valid assignment to τ can be extended to a LMP-valid assignment to μ:

Let there be some optimal LP-valid assignment to τ.

Extending the optimal solution:

First of all, note that any change to the values of μ who do not correspond to τ (denote by $\mu_{-\tau}$) do not change the value of the tartget function. Start by assigning $0's$ to all $\mu_{-\tau}$. Note that at this point $LMP(1)$ holds from $LP(1,2)$ and the zero assignment to $\mu_{-\tau}$. From this point on we will only increase values of $\mu_{-\tau}$ and will not increase to more than 1. So we are done with $LMP(1)$.
Claim: $\forall ij : \tau_{ij} \leq 1$

Assume by contradiction that $\tau_{ij} > 1$ then express τ_{ij} as $1+\epsilon$, for some $\epsilon > 0$ It follows:

$$1+\epsilon = \tau_{ij} \geq \tau_i + \tau_j - 1 \geq 1+\epsilon+1+\epsilon-1$$

Subtract $1+\epsilon$ from both sides:

$$0 \geq \epsilon$$

Contradiction to the definition of ϵ.
Claim: $\forall i : \tau_i \leq 1$

Assume by contradiction that $\tau_i > 1$. Then:

$$\tau_{ij} \geq \tau_i + \tau_j - 1 = 1+\epsilon+\tau_j - 1 = \epsilon + \tau_j$$

In contradiction to $LP(4)$

So we now can assume that $\forall ij \in E : 0 \leq \tau_{ij} \leq 1$. and ∀i:0≤ $\tau_i \leq 1$
Let there be some $\tau_{ij}, \tau_i, \tau_j$ in our optimal solution. We will define μ_{ij}, μ_i, μ_j in such a way that all constraints of LMP hold for these values.
First of all, in order to satisfy $LMP(2)$ we must assign:

- $\mu_i(0) = 1 - \mu_j(1) = 1 - \tau_j$
- $\mu_j(0) = 1 - \mu_i(1) = 1 - \tau_i$

We will now satisfy $LMP(\{4,5\})$

- $\mu_j(1) = \tau_j = \sum_{x_i} \mu_{ij}(x_i, 1) = \mu_{ij}(0,1) + \tau_{ij}$. We know that $\tau_{ij} \leq \tau_j$ so we can assign a non-negative value smaller than 1 to $\mu_{ij}(0,1)$ s.t the equality holds.
- $\mu_j(0) = 1 - \tau_j = \sum_{x_i} \mu_{ij}(x_i, 0) = \mu_{ij}(0,0) + \mu_{ij}(1,0)$. We shall assign the value $1 - \tau_j$ to $\mu_{ij}(0,0)$

Note that $LMP(3)$ is satisfied by these assignments:
$\sum_{x_i,x_j} \mu_{ij}(x_i, x_j) = \sum_{x_i} \mu_{ij}(x_i, 0) + \sum_{x_i} \mu_{ij}(x_i, 1) = \tau_j + (1 - \tau_j) = 1$
This extends with no contradicting assignments to all other μ
So... We are done!
□