STA2700 - Graphical Models - Take Home 2

Luis Correia - Student No. 1006508566

November 10th 2020

Our discussions (and the derivation) of the sum-product algorithm were mainly based on the distributive law, which states that $a \cdot b + a \cdot c = a \cdot (b + c)$. In our framework, we also assumed that

$$a, b, c \in \mathbb{R}_{\geq 0}$$

where $\mathbb{R}_{\geq 0}$ is the set of non-negative real numbers.

Moreover, operation "+" denotes ordinary addition with 0 as its additive identity element, and operation "$\cdot$" denotes ordinary multiplication with 1 as its multiplicative identity.

With few examples given below, we want to show that sum-product algorithm can be generalized.

Question 1

{*Solution.*}

Using the $(+, 0)$ criteria, we have the Distributive Law can be expressed by:

$$a \cdot b + a \cdot c = a \cdot (b + c)$$

Using the new criteria, i.e., substituting $(+, 0)$ by $(\max, 0)$ we have the new derived Distributive Law still **holds** given by:

$$\max\{a \cdot b, a \cdot c\} = a \cdot \max\{b, c\}$$

Also, the Additive Identity also **holds**, as we have $a \in \mathbb{R}_{\geq 0}$:

$$\max\{a, 0\} = a$$

Question 2

{*Solution.*}

For the graph represented by the figure below:

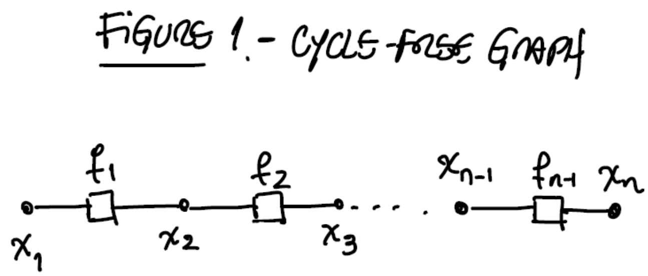

Figure 1: Cycle-free Graph

For this graph, we also have:

$$f(\boldsymbol{x}) = f_1(x_1, x_2) f_2(x_2, x_3) \dots f_{n-1}(x_{n-1}, x_n)$$

Applying the new criteria, we have:

$$\begin{aligned}\max_{\boldsymbol{x} \in \mathbb{R}^n_{\geq 0}} f(\boldsymbol{x}) &= \max_{\boldsymbol{x} \in \mathbb{R}^n_{\geq 0}} f(x_1, \dots, x_n) \\ &= \max_{x_1} \dots \max_{x_n} [f_1(x_1, x_2) \cdot \dots \cdot f_{n-1}(x_{n-1}, x_n)]\end{aligned}$$

which can be computed under the new algorithm by:

$$\implies \max_{\boldsymbol{x} \in \mathbb{R}^n_{\geq 0}} f(\boldsymbol{x}) = \max_{x_1} \left[f_1(x_1, x_2) \cdot \max_{x_2} \left[f_2(x_2, x_3) \cdot \dots \cdot \max_{x_n} [f_{n-1}(x_{n-1}, x_n)] \right] \dots \right] \tag{1}$$

The complexity of the *max-product* algorithm is composed by $(n-1)$ products and n *maximums* which implies the complexity of order $O_{MP}[n(n-1)] \approx O(n^2)$.

The complexity of the regular sum-product algorithm is given by $O_{SP}(|\mathcal{X}|^n)$ and we have:

$$O(n^2) < O_{SP}(|\mathcal{X}|^n) \tag{2}$$

By (2), the *max-product* algorithm is efficient for finding the values of $\boldsymbol{x}$ which maximizes the joint distribution $f(\boldsymbol{x})$ and also allow us to obtain the value of this joint distribution at this maximum.

Question 3

{*Solution.*}

In order to compute the marginals, lets consider the topology of the graph represented in figure below

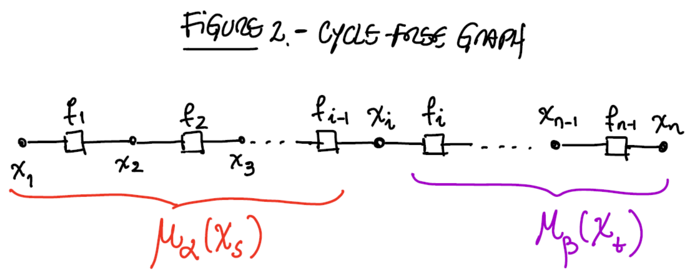

Figure 2: Marginal in Cycle-Free Graph

Using the *max-product* algorithm, follows from (1) that we can calculate the marginals from variable and factor nodes as follows:

$$f(x_i) = \underbrace{\max_{x_1}\left[f_1(x_1,x_2)\cdot \max_{x_2}\left[f_2(x_2,x_3)\cdot\ldots\cdot\max_{x_{i-1}}[f_{i-1}(x_{i-1},x_i)]\right]\ldots\right]}_{\mu_\alpha(x_i)}\times$$

$$\underbrace{\max_{x_n}\left[f_{n-1}(x_{n-1},x_n)\cdot \max_{x_{n-2}}\left[f_{n-2}(x_{n-2},x_{n-1})\cdot\ldots\cdot\max_{x_{i+1}}[f_{i+1}(x_{i+1},x_i)]\right]\ldots\right]}_{\mu_\beta(x_i)}$$

where

- $0 \leq i \leq n$
- $\mu_\alpha(x_i)$ accounts for the messages from left-to-right, i.e, from x_1 to x_{i-1}
- $\mu_\beta(x_i)$ accounts for the messages from right-to-left, i.e., x_n to x_{i+1}

Using the *message-passing* notation from left-to-right, equivalently this marginal can be represented as follows

$$\mu_{f_i\to x_i}(x_i) = \max_{x_1}\ldots\max_{x_{i-1}}\left[f_i(x_1,\ldots,x_{i-1})\cdot\mu_{x_{i+1}\to f_i}(x_i)\right]$$

... and...

$$\mu_{x_i\to f_i}(x_i) = \mu_{f_i\to x_i}(x_i)$$

Question 4

{*Solution.*}

Using the $(\min, \infty)$ criteria in substitution to $(+, 0)$ and $(+, 0)$ instead of $(\cdot\,, 1)$ and the alphabet $\mathcal{X} = \mathbb{R} \cup \{+\infty\}$, and let's also consider $a, b, c \in \mathcal{X}$. Then we have the following:

1. Product Identity

$$a \cdot 1 = a \implies a + 0 = a \text{ (Holds)}$$

2. Addictive Identity

$$a + 0 = a \implies \min\{a, +\infty\} = a \text{ (Holds)}$$

3. Distributive Law

$$a \cdot b + a \cdot c = a \cdot (b + c) \implies \min\{a + b, a + c\} = a + \min\{b, c\} \text{ (Holds)}$$

All rules holds in the new measurement system.

Question 5

{*Solution.*}

In this question we will extrapolate the 1^{st} algorithm obtained to a *general factor graph* and calculate the marginals on *factor* and *variable* nodes. The strategy here will be derive the general formulas for message-passing from *factor* and *variable* nodes and then apply the new algorithm.

Marginals in Sum-Product algorithm

By definition, the marginal is obtained by summing the joint distribution over all variables except x so that

$$p(x) = \sum_{\boldsymbol{x} \backslash x} p(\boldsymbol{x}) \tag{3}$$

we can then describe $p(\boldsymbol{x})$ considering the partition of factors into groups associated to variable node:

$$p(\boldsymbol{x}) = \prod_{s \in ne(x)} F_s(x, X_s) \tag{4}$$

where $ne(x)$ denotes the set of factor nodes that are neighbours of x, and X_s denotes the set of all variables in the sub-tree connected to the variable node x via the factor node f_s and $F_s(x, X_s)$ represents the product of all the factors in the group associated with factor f_s.

After substituting (3) in (4), some transformations and algebraic simplifications, we find that, for a general factor graph and we have:

$$\begin{aligned} p(x) &= \sum_{\boldsymbol{x} \backslash x} \Big(\prod_{s \in ne(x)} F_s(x, X_s) \Big) \\ &= \prod_{s \in ne(x)} \Big[\sum_{X_s} F_s(x, X_s) \Big] \\ &= \prod_{s \in ne(x)} \mu_{f_s \to x}(x) \end{aligned}$$

We can further derive the messages passing through the factor node f_s as follows:

$$\mu_{f_s \to x}(x) = \sum_{x_1} \cdots \sum_{x_M} f_s(x, x_1, \ldots, x_M) \prod_{m \in ne(f_s) \backslash x} \mu_{x_m \to f_s}(x_m) \tag{5}$$

where $x_1, \ldots, x_M$, are the nodes connected with f_s other than x.

Similarly the messages passing through the variable node x can be expressed by:

$$\mu_{x \to f_s}(x) = \prod_{l \in ne(x) \backslash f_s} \mu_{f_l \to x}(x) \tag{6}$$

Marginals in the New Algorithm

We noticed in this scenario it might be interesting to work with the logarithmic of the densities because the `ln` is a monotonic function and we will be able to convert the joint distribution in a bunch of sums, as well as the products on marginals derived from the traditional sum-product.

In this sense, applying the algorithm derived by substituting the `sum`'s by `max`'s (5) and (6), we have the derivation of *max-sum* algorithm[1] on factor and variable nodes as follows:

$$\mu_{f_s \to x}(x) = \max_{x_1} \ldots \max_{x_M} \Big(\ln f_s(x, x_1, \ldots, x_M) + \sum_{m \in ne(f_s) \setminus x} \mu_{x_m \to f_s}(x_m) \Big) \tag{7}$$

... and

$$\mu_{x \to f_s}(x) = \sum_{l \in ne(x) \setminus f_s} \mu_{f_l \to x}(x) \tag{8}$$

[1] Here, assuming we applied the logarithm over the joint densities.

Question 6

{*Solution.*}

Let's consider the following graphs `OR` and `AND` operations.

Graphically, these operations can be represented by the following graphs:

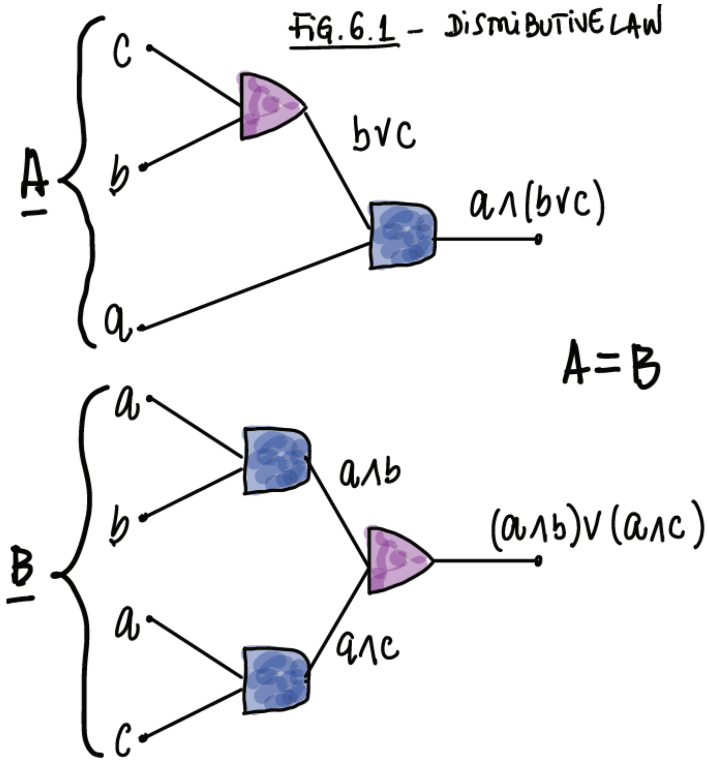

Figure 3: Distributive Law - Logical AND/OR

The identity elements for the `OR` and `AND` operations can be verified below:

- Logical-`OR`

$$a \vee 0 = a + 0 = a \implies \text{identity element is } 0.$$

- Logical-`AND`

$$a \wedge 1 = a \cdot 1 = a \implies \text{identity element is } 1.$$

To verify the Distributive Law, we will consider the alphabet $\mathcal{X} = \{0, 1\}$ then both operations with all possible results, following the graphical representation are listed below:

Table 1: $a \wedge (b \vee c)$

a	b	c	$b \vee c$	$a \wedge (b \vee c)$
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	1	0
1	0	0	0	0
1	0	1	1	1
1	1	0	1	1
1	1	1	1	1

Table 2: $(a \wedge b) \vee (a \wedge c)$

a	b	c	$a \wedge b$	$a \wedge c$	$(a \wedge b) \vee (a \wedge c)$
0	0	0	0	0	0
0	0	1	0	0	0
0	1	0	0	0	0
0	1	1	0	0	0
1	0	0	0	0	0
1	0	1	0	1	1
1	1	0	1	0	1
1	1	1	1	1	1

From tables (1) and (2) we see that Distributive Law holds.

Question 7

{*Solution.*}

According with [1], by definition, a *ring* is a set $\mathcal{S}$ together with operations $+, \cdot$ (called *addition* and *multiplication*) and a distinguished elements 0 and 1, which satisfy the following properties:

(a) $(\mathcal{S}, +, 0)$ is an *abelian* group, also called *commutative* group, is a group in which the result of applying the group operation to two group elements does not depend on the order in which they are written;

(b) $(\mathcal{S}, \cdot, 1)$ is a *monoid*, i.e., is a semigroup equipped with an *associative binary operation* and an *identity element*;

(c) $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(b + c) \cdot a = b \cdot a + c \cdot a$ for all $a, b, c \in \mathcal{S}$.

{*Proof:*}

(a) Let $a, b \in \mathcal{S} = \{0, 1\}$, then we have

$$a + b = b + a, \forall a, b \in \mathcal{S} \tag{9}$$

$$a + 0 = 0 + a = a, \forall a \in \mathcal{S} \tag{10}$$

$$a + (-a) = (-a) + a = 0, \forall a \in \mathcal{S} \tag{11}$$

... then (9), (10) and (11) $\implies$ $\mathcal{S}$ is an abelian group.

(b) Let $a, b, c \in \mathcal{S} = \{0, 1\}$, then we have

$$(a \cdot b) \cdot c = a \cdot (b \cdot c), \forall a, b, c \in \mathcal{S} \tag{12}$$

$$a \cdot 1 = 1 \cdot a = a, \forall a \in \mathcal{S} \tag{13}$$

... then (12) and (13) $\implies$ $\mathcal{S}$ is a monoid.

(c) Let $a, b, c \in \mathcal{S} = \{0, 1\}$, then we have

$$a \cdot (b + c) = a \cdot b + a \cdot c, \forall a, b, c \in \mathcal{S} \tag{14}$$

$$(b + c) \cdot a = b \cdot a + c \cdot a, \forall a, b, c \in \mathcal{S} \tag{15}$$

... then (14) and (15) $\implies$ *distributive law* still holds .

Then we conclude $(\mathcal{S}, + \cdot)$ is a ring.

References

[1] Rowen, L. H. *Ring Theory, Vol.1.* Academic Press Inc., 1988.

[2] Jacobson, N. *Structure of Rings.* American Mathematical Society, 1968.