PAI. Exact and Approximate Inference

Anastasia Makarova

ETH Zürich

2.11.2018

Inference

Tree-structured:

- Variable elimination
- Belief propagation

Loopy networks:

- Loopy belief propagation
- Variational inference
- Gibbs sampling (Monte Carlo Sampling)

Homework problem 1

A	$g_1(A)$
f	0.7
t	0.3

C	$g_2(C)$
f	0.4
t	0.6

C	D	$g_3(C, D)$
f	f	0.10
f	t	0.90
t	f	0.25
t	t	0.75

B	E	$g_4(B, E)$
f	f	0.8
f	t	0.2
t	f	0.6
t	t	0.4

A	B	C	$g_5(A, B, C)$
f	f	f	0.8
f	f	t	0.2
f	t	f	0.2

...

D	E	F	$g_6(D, E, F)$
f	f	f	0.05
f	f	t	0.95

...

Variable elimination

Sum factor

$$g_7(D,E) = \sum_{f} \prod_{j \epsilon \{6\}} g_j = \sum_{f} g_6(D,E,f) \qquad (1)$$

Table: Intermediate factor $g_7(D,E)$

D	E	$g_7(D,E)$
f	f	$0.05 + 0.95 = 1$
f	t	$0.00 + 1.00 = 1$
t	f	$1.00 + 0.00 = 1$
t	t	$0.75 + 0.25 = 1$

Product factor

$$g_8(B, D) = \sum_e g_4(B, e) g_7(D, e) = \sum_e f_1(B, D, e) \qquad (2)$$

Table: Product factor $f_1(B, D, E)$

B	D	E	$f_1(B, D, E)$
f	f	f	$1 \times 0.8 = 0.8$
f	f	t	$1 \times 0.2 = 0.2$
f	t	f	$1 \times 0.8 = 0.8$
f	t	t	$1 \times 0.2 = 0.2$
t	f	f	$1 \times 0.6 = 0.6$
t	f	t	$1 \times 0.4 = 0.4$
t	t	f	$1 \times 0.6 = 0.6$
t	t	t	$1 \times 0.4 = 0.4$

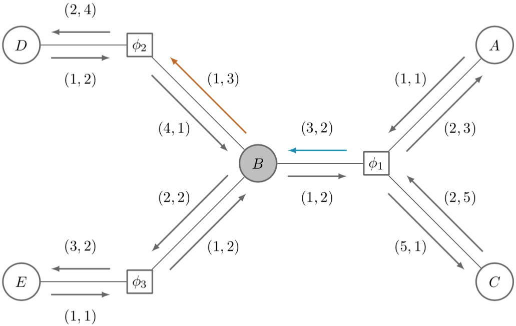
Belief propagation

$$P(A,B,C,D,E) \propto \phi_1(A,B,C)\phi_2(B,D)\phi_3(B,E).$$

All random variables are binary and the factors are defined as follows:

$$\phi_1(A=a,B=b,C=c) = \begin{cases} 0 & \text{if } c=0 \\ a+b+c & \text{if } c=1 \end{cases}$$
$$\phi_2(B=b,D=d) = b+d$$
$$\phi_3(B=b,E=e) = b+e.$$

Assume that at the t-th iteration of the belief propagation algorithm the messages shown in the figure below are exchanged between variable and factor nodes in the graph. Each vector (v_0, v_1) defines an (unnormalized) message, for which v_0 corresponds to value 0 and v_1 corresponds to value 1. For example, $\mu^{(t)}_{\phi_1 \to B}(0) = 3$, $\mu^{(t)}_{\phi_1 \to B}(1) = 2$, and $\mu^{(t)}_{B \to \phi_1}(0) = 1$, $\mu^{(t)}_{B \to \phi_1}(1) = 2$.



Belief propagation

Will belief propagation converge on the factor graph?

Solution:

1. The most important observation here is that the tree is acyclic and connected, which makes it a tree.
2. Belief propagation always converges to the exact marginals when the factor graph is a tree.
3. Therefore, for this factor graph belief propagation will converge.

Belief propagation

From the provided messages compute the approximate marginal distribution of B.

Solution:

1. From the belief propagation equations, we have:

$$P^{(t)}(B) \propto \prod_{\phi_i \in N(B)} \mu^{(t)}_{\phi_i \to B}(B)$$

$$P^{(t)}(B) \propto \mu^{(t)}_{\phi_1 \to B}(B)\mu^{(t)}_{\phi_2 \to B}(B)\mu^{(t)}_{\phi_3 \to B}(B)$$

2. We proceed to calculate the product of the messages for both values of B.

Belief propagation

Solution:

1. For $B = 0$:

$$\hat{P}^{(t)}(0) = \mu^{(t)}_{\phi_1 \to B}(0)\mu^{(t)}_{\phi_2 \to B}(0)\mu^{(t)}_{\phi_3 \to B}(0)$$
$$\hat{P}^{(t)}(0) = 3 \times 4 \times 1 = 12$$

2. For $B = 1$:

$$\hat{P}^{(t)}(1) = \mu^{(t)}_{\phi_1 \to B}(1)\mu^{(t)}_{\phi_2 \to B}(1)\mu^{(t)}_{\phi_3 \to B}(1)$$
$$\hat{P}^{(t)}(1) = 2 \times 1 \times 2 = 4$$

3. Normalizing:

$$P^{(t)}(B) = \frac{(12, 4)}{Z} = (0.75, 0.25)$$

Belief propagation

Compute the message $\mu^{(t+1)}_{B\to\phi_2}$.

Solution:

1. From the belief propagation equations:

$$\mu^{(t+1)}_{B\to\phi_2}(B) = \prod_{\phi_i \in N(B)\setminus\{\phi_2\}} \mu^{(t)}_{\phi_i\to B}(B)$$

$$\mu^{(t+1)}_{B\to\phi_2}(B) = \mu^{(t)}_{\phi_1\to B}(B)\mu^{(t)}_{\phi_3\to B}(B)$$

2. Again we calculate for both values of B. No need to normalize in the case of a factor.

Belief propagation

Solution:

1. For $B = 0$:

$$\mu^{(t+1)}_{B \to \phi_2}(0) = \mu^{(t)}_{\phi_1 \to B}(0)\mu^{(t)}_{\phi_3 \to B}(0) = 3 \times 1 = 3$$

2. For $B = 1$:

$$\mu^{(t+1)}_{B \to \phi_2}(1) = \mu^{(t)}_{\phi_1 \to B}(1)\mu^{(t)}_{\phi_3 \to B}(1) = 2 \times 2 = 4$$

3. The message is:

$$\mu^{(t+1)}_{B \to \phi_2}(B) = (3, 4)$$

Compute the message $\mu^{(t+1)}_{\phi_1 \to B}$.

Solution:

1. From the belief propagation equations:

$$\mu^{(t+1)}_{\phi_1 \to B}(B) = \sum_{v^* \in \boldsymbol{v}_{\boldsymbol{\phi_1}} \setminus \{B\}} \phi_1(v_{\phi_1}) \prod_{v^* \in N(\phi_1) \setminus \{B\}} \mu^{(t)}_{v^* \to \phi_1}(v^*)$$

$$\mu^{(t+1)}_{\phi_1 \to B}(B) = \sum_a \sum_c \phi_1(a, B, c) \mu^{(t)}_{A \to \phi_1}(a) \mu^{(t)}_{C \to \phi_1}(c)$$

2. In this case we have to marginalize over A and C which means the summation has 4 terms, and we do this for both values of B.

Belief propagation

Solution:

1. For $B = 0$:

$$\mu^{(t+1)}_{\phi_1 \to B}(0) =$$

$$\phi_1(0,0,0)\mu^{(t)}_{A\to\phi_1}(0)\mu^{(t)}_{C\to\phi_1}(0)+$$
$$\phi_1(0,0,1)\mu^{(t)}_{A\to\phi_1}(0)\mu^{(t)}_{C\to\phi_1}(1)+$$
$$\phi_1(1,0,0)\mu^{(t)}_{A\to\phi_1}(1)\mu^{(t)}_{C\to\phi_1}(0)+$$
$$\phi_1(1,0,1)\mu^{(t)}_{A\to\phi_1}(1)\mu^{(t)}_{C\to\phi_1}(1)$$

2. Note that $\phi_1(A, B, C) = 0$ if $C = 0$, therefore we just need to sum 2 terms.

$$\mu^{(t+1)}_{\phi_1 \to B}(0) = 1 \times 1 \times 5 + 2 \times 1 \times 5 = 15$$

Solution:

1. For $B = 1$:

$$\mu^{(t+1)}_{\phi_1 \to B}(1) =$$

$$\phi_1(0,1,0)\mu^{(t)}_{A \to \phi_1}(0)\mu^{(t)}_{C \to \phi_1}(0)+$$
$$\phi_1(0,1,1)\mu^{(t)}_{A \to \phi_1}(0)\mu^{(t)}_{C \to \phi_1}(1)+$$
$$\phi_1(1,1,0)\mu^{(t)}_{A \to \phi_1}(1)\mu^{(t)}_{C \to \phi_1}(0)+$$
$$\phi_1(1,1,1)\mu^{(t)}_{A \to \phi_1}(1)\mu^{(t)}_{C \to \phi_1}(1)$$

2. Note that the product of the messages from A and C didn't change, only the value of $\phi_1(A, B, C)$.

$$\mu^{(t+1)}_{\phi_1 \to B}(1) = 2 \times 1 \times 5 + 3 \times 1 \times 5 = 25$$

3. Finally:

$$\mu^{(t+1)}_{\phi_1 \to B}(B) = (15, 25)$$

Recap

Gibbs sampling

Assume that we are running a Gibss sampler on the same factor graph and the last sample we drew is $(A = 0, B = 0, C = 1, D = 1, E = 1)$. Compute the distribution from which we should draw the new value of A.

Solution:

1. Remember that a Gibbs sampler uses a conditional distribution, i.e. it samples from:

$$P(A^{(t+1)} \mid B^{(t)}, C^{(t)}, D^{(t)}, E^{(t)}) \tag{3}$$

2. Then remember the bayes theorem:

$$P(A^{(t+1)} \mid B^{(t)}, C^{(t)}, D^{(t)}, E^{(t)}) = \tag{4}$$

$$\frac{P(A^{(t+1)}, B^{(t)}, C^{(t)}, D^{(t)}, E^{(t)})}{P(B^{(t)}, C^{(t)}, D^{(t)}, E^{(t)})} \tag{5}$$

Gibbs sampling

Solution:

1. Note that for sampling A, the denominator in equation 5 is a constant, therefore we have:

$$P(A^{(t+1)} \mid B^{(t)}, C^{(t)}, D^{(t)}, E^{(t)}) \propto \tag{6}$$
$$P(A^{(t+1)}, B^{(t)}, C^{(t)}, D^{(t)}, E^{(t)}) \tag{7}$$

2. And from the factor graph we have

$$P(A^{(t+1)}, B^{(t)}, C^{(t)}, D^{(t)}, E^{(t)}) \propto \tag{8}$$
$$\phi_1(A^{(t+1)}, B^{(t)}, C^{(t)})\phi_2(B^{(t)}, D^{(t)})\phi_3(B^{(t)}, E^{(t)}) \tag{9}$$

Solution:

1. Finally, we only care about the factors that include A, the others are constants:

$$P(A^{(t+1)}, B^{(t)}, C^{(t)}, D^{(t)}, E^{(t)}) \propto \phi_1(A^{(t+1)}, B^{(t)}, C^{(t)}) \quad (10)$$

2. We can now replace with both possible values of A.

$$P(A^{(t+1)} = 0 \mid B^{(t)} = 0, C^{(t)} = 1) \propto \phi_1(0, 0, 1) = 1$$
$$P(A^{(t+1)} = 1 \mid B^{(t)} = 0, C^{(t)} = 1) \propto \phi_1(1, 0, 1) = 2$$

3. After normalizing, we find that we must draw A from the following distribution:

$$\hat{P}(A) = (0.333, 0.667) \quad (11)$$