1.

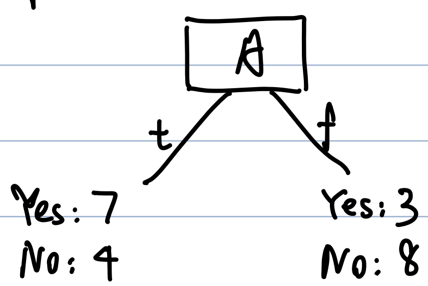
$$ENT(D) = -\sum_{i=1}^{k} P(D=d_k) \cdot \log_2 P(D=d_k)$$
$$= -\frac{10}{22} \cdot \log_2 \frac{10}{22} - \frac{12}{22} \cdot \log_2 \frac{12}{22}$$
$$= 0.9940$$

$$ENT(D|A) = -\frac{11}{22}\left(\frac{7}{11}\log_2\frac{7}{11} + \frac{4}{11}\log_2\frac{4}{11}\right) - \frac{11}{22}\left(\frac{3}{11}\log_2\frac{3}{11} + \frac{8}{11}\log\frac{8}{11}\right)$$
$$= 0.8955$$

$$ENT(D|B) = -\frac{14}{22}\left(\frac{8}{14}\log_2\frac{8}{14} + \frac{6}{14}\log_2\frac{6}{14}\right) - \frac{8}{22}\left(\frac{2}{8}\log_2\frac{2}{8} + \frac{6}{8}\log_2\frac{6}{8}\right)$$
$$= 0.9220$$

$$ENT(D|C) = -\frac{7}{22}\left(\frac{4}{7}\log_2\frac{4}{7} + \frac{3}{7}\log_2\frac{3}{7}\right) - \frac{15}{22}\left(\frac{6}{15}\log_2\frac{6}{15} + \frac{9}{15}\log_2\frac{9}{15}\right)$$
$$= 0.9755$$

We see the conditional entropy ENT(D|A) is minimized, so we should choose A as the first branch.



① when $A=t$:

$$ENT(D|B \wedge (A=t)) = -\frac{7}{11}\left(\frac{7}{7}\log\frac{7}{7} + \frac{0}{7}\log\frac{0}{7}\right) - \frac{4}{11}\left(\frac{4}{4}\log\frac{4}{4} + \frac{0}{4}\log\frac{0}{4}\right) = 0$$

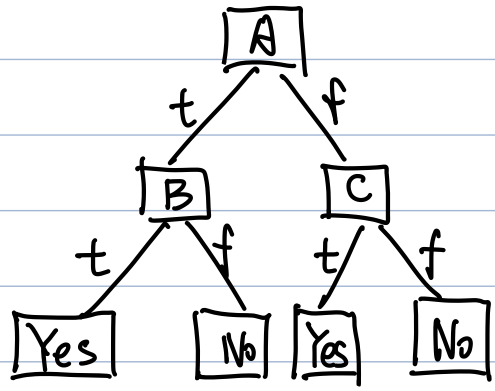
② when $A=f$:

$$ENT(D|B \wedge (A=f)) = -\frac{7}{11}\left(\frac{1}{7}\log\frac{1}{7} + \frac{6}{7}\log\frac{6}{7}\right) - \frac{4}{11}\left(\frac{2}{4}\log\frac{2}{4} + \frac{2}{4}\log\frac{2}{4}\right) = 0.7402$$

$$ENT(D|C \wedge (A=f)) = -\frac{3}{11}\left(\frac{3}{3}\log\frac{3}{3} + \frac{0}{3}\log\frac{0}{3}\right) - \frac{8}{11}\left(\frac{0}{8}\log\frac{0}{8} + \frac{8}{8}\log\frac{8}{8}\right) = 0$$

So when $A=t$, we should choose B as attribute, when $A=f$, we should use C as attribute, and they successfully classify all examples.

Decision tree:

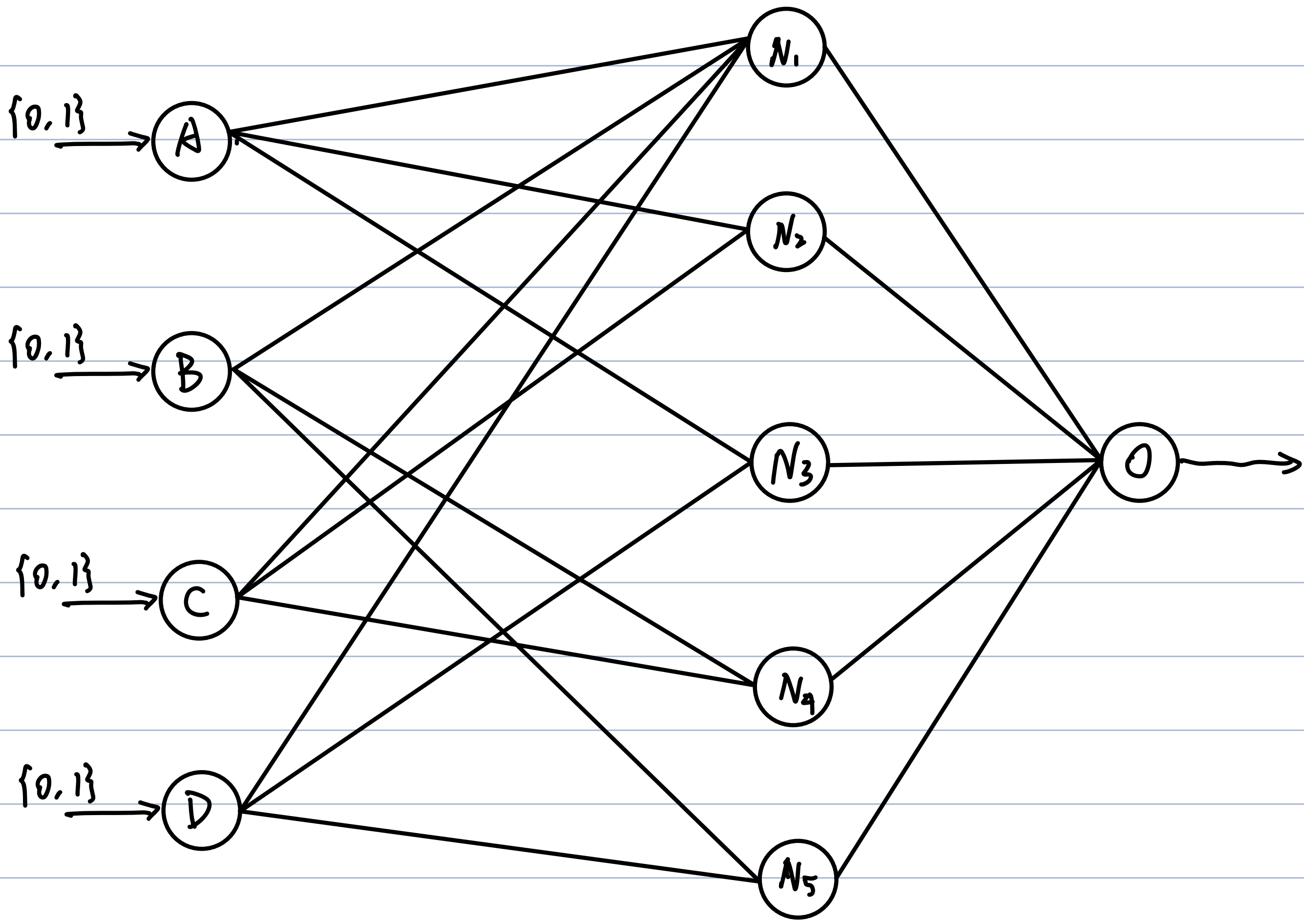


2.

$$(A \vee \neg B) \oplus (\neg C \vee D) \equiv (A \vee \neg B \vee \neg C \vee D) \wedge (\neg(A \vee \neg B) \vee \neg(\neg C \vee D))$$
$$\equiv (A \vee \neg B \vee \neg C \vee D) \wedge ((\neg A \wedge B) \vee (C \wedge \neg D))$$
$$\equiv (A \vee \neg B \vee \neg C \vee D) \wedge (\neg A \vee (C \wedge \neg D)) \wedge (B \vee (C \wedge \neg D))$$
$$\equiv (A \vee \neg B \vee \neg C \vee D) \wedge (\neg A \vee C) \wedge (\neg A \vee \neg D) \wedge (B \vee C) \wedge (B \vee \neg D)$$

Now we have converted the NNF into CNF, and there are 5 clauses.

Thus, the structure is:



For neuron N_1 : $(A \vee \neg B \vee \neg C \vee D)$

A	B	C	D	$A \vee \neg B \vee \neg C \vee D$	$A-B-C+D$
0	0	0	0	1	0
0	0	0	1	1	1
0	0	1	0	1	-1
0	0	1	1	1	0
0	1	0	0	1	-1
0	1	0	1	1	0
0	1	1	0	0	-2
0	1	1	1	1	-1
1	0	0	0	1	1
1	0	0	1	1	2
1	0	1	0	1	0
1	0	1	1	1	1
1	1	0	0	1	0
1	1	0	1	1	1
1	1	1	0	1	-1
1	1	1	1	1	0

let $W_A = 1, W_B = -1, W_C = -1, W_D = 1$,
we see for $(A \vee \neg B \vee \neg C \vee D)$ to be 0, we need $(A-B-C+D) = -2$.
for $(A \vee \neg B \vee \neg C \vee D)$ to be 1, we need $(A-B-C+D) \geq -1$
So we choose threshold t_{N_1} to be -1.5
the step function is $g_{N_1}(x) = \begin{cases} 0 & \text{if } x < -1.5 \\ 1 & \text{if } x \geq -1.5 \end{cases}$

For neuron N_2 : $(\neg A \vee C)$

A	C	$\neg A \vee C$	$-A+C$
0	0	1	0
0	1	1	1
1	0	0	-1
1	1	1	0

let $W_A = -1, W_C = 1$. we see for $(\neg A \vee C)$ to be 0, we need $(-A+C) = -1$.
for $(\neg A \vee C)$ to be 1, we need $(-A+C) \geq 0$.
So we choose threshold t_{N_2} to be -0.5
the step function is $g_{N_2}(x) = \begin{cases} 0 & \text{if } x < -0.5 \\ 1 & \text{if } x \geq -0.5 \end{cases}$

For neuron N_3 : $(\neg A \vee \neg D)$

A	D	$\neg A \vee \neg D$	$-A-D$
0	0	1	0
0	1	1	-1
1	0	1	-1
1	1	0	-2

let $W_A = -1, W_D = -1$. we see for $(\neg A \vee \neg D)$ to be 0, we need $(-A-D) = -2$.
for $(\neg A \vee \neg D)$ to be 1, we need $(-A-D) \geq -1$.
So we choose threshold t_{N_3} to be -1.5
the step function is $g_{N_3}(x) = \begin{cases} 0 & \text{if } x < -1.5 \\ 1 & \text{if } x \geq -1.5 \end{cases}$

For neuron N_4 : $(B \vee C)$

B	C	$B \vee C$	$B+C$
0	0	0	0
0	1	1	1
1	0	1	1
1	1	1	2

let $w_B = 1$, $w_C = 1$. we see for $(B \vee C)$ to be 0, we need $(B+C) = 0$.

for $(B \vee C)$ to be 1, we need $(B+C) \geq 1$.

So we choose threshold t_{N_4} to be 0.5

the step function is $g_{N_4}(x) = \begin{cases} 0 & \text{if } x < 0.5 \\ 1 & \text{if } x \geq 0.5 \end{cases}$

For neuron N_5 : $(B \vee \neg D)$

B	D	$B \vee \neg D$	$B-D$
0	0	1	0
0	1	0	-1
1	0	1	1
1	1	1	0

let $w_B = 1$, $w_D = -1$. we see for $(B \vee \neg D)$ to be 0, we need $(B-D) = -1$.

for $(B \vee \neg D)$ to be 1, we need $(B-D) \geq 0$.

So we choose threshold t_{N_4} to be -0.5

the step function is $g_{N_5}(x) = \begin{cases} 0 & \text{if } x < -0.5 \\ 1 & \text{if } x \geq -0.5 \end{cases}$

For output layer output : $(N_1 \wedge N_2 \wedge N_3 \wedge N_4 \wedge N_5)$:

Let $W_{N_1} = W_{N_2} = W_{N_3} = W_{N_4} = W_{N_5} = 1$;

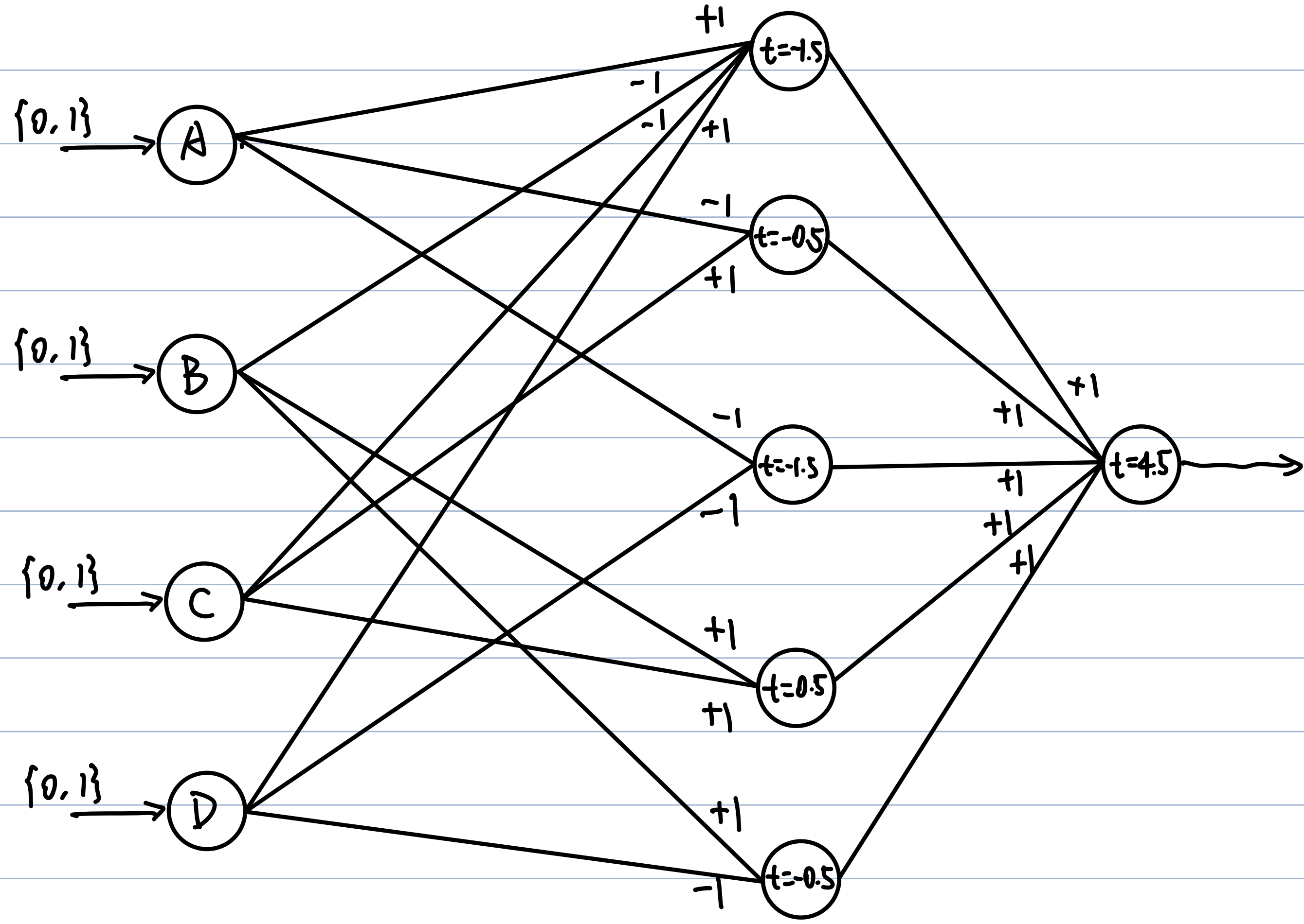
For it to be 1, we need $N_1 + N_2 + N_3 + N_4 + N_5 = 5$.

For it to be 0, we need $N_1 + N_2 + N_3 + N_4 + N_5 \leq 4$.

So we choose threshold t_0 to be 4.5

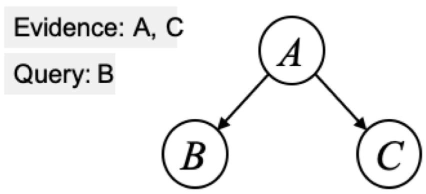
the step function is $g_0(x) = \begin{cases} 0 & \text{if } x < 4.5 \\ 1 & \text{if } x \geq 4.5 \end{cases}$

Thus, the structure is:



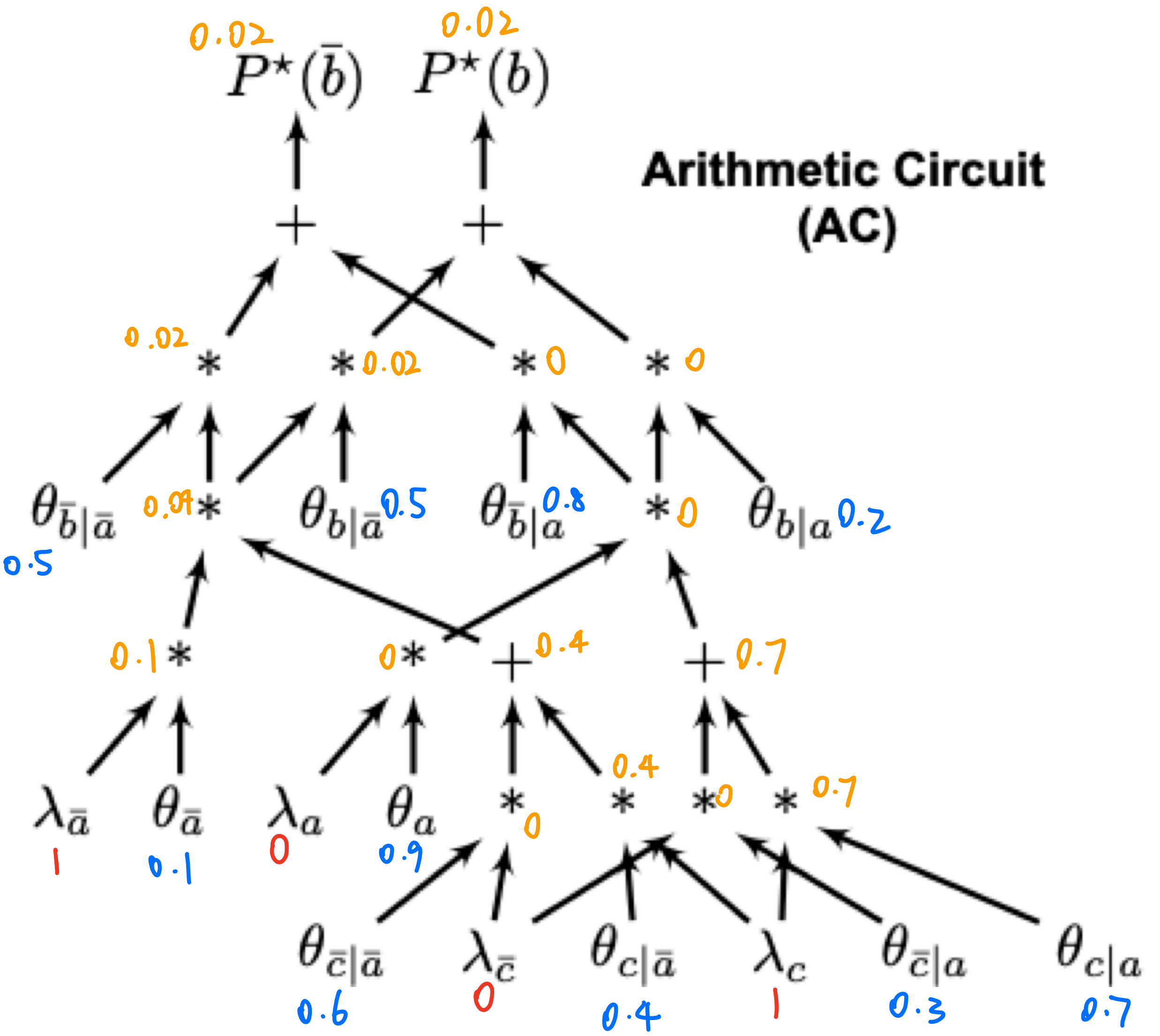
3.

(a) $\theta_a = 0.9$ $\theta_{b|a} = 0.2$ $\theta_{b|\bar{a}} = 0.5$ $\theta_{c|a} = 0.7$ $\theta_{c|\bar{a}} = 0.4$



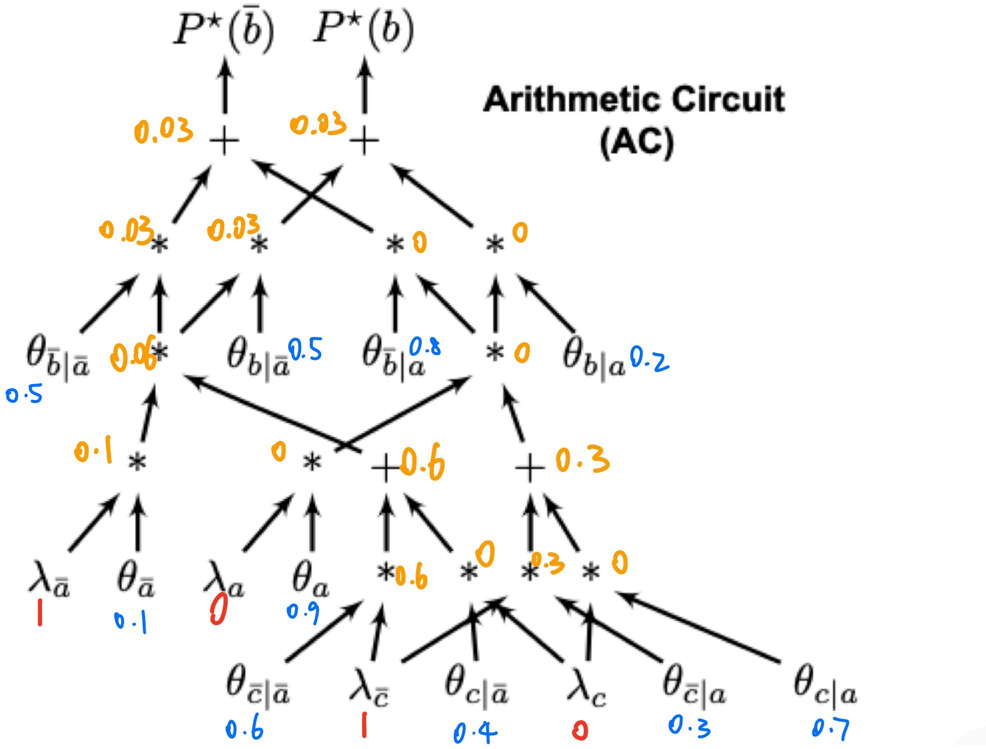
$\theta_{\bar{a}} = 0.1$	$\theta_{\bar{b}|\bar{a}} = 0.5$	$\theta_{\bar{c}|\bar{a}} = 0.6$
	$\theta_{b|\bar{a}} = 0.5$	$\theta_{c|\bar{a}} = 0.4$
$\theta_a = 0.9$	$\theta_{\bar{b}|a} = 0.8$	$\theta_{\bar{c}|a} = 0.3$
	$\theta_{b|a} = 0.2$	$\theta_{c|a} = 0.7$

① evidence $e_1 = \bar{a}, c \Rightarrow \lambda_a = 0,\ \lambda_{\bar{c}} = 0$



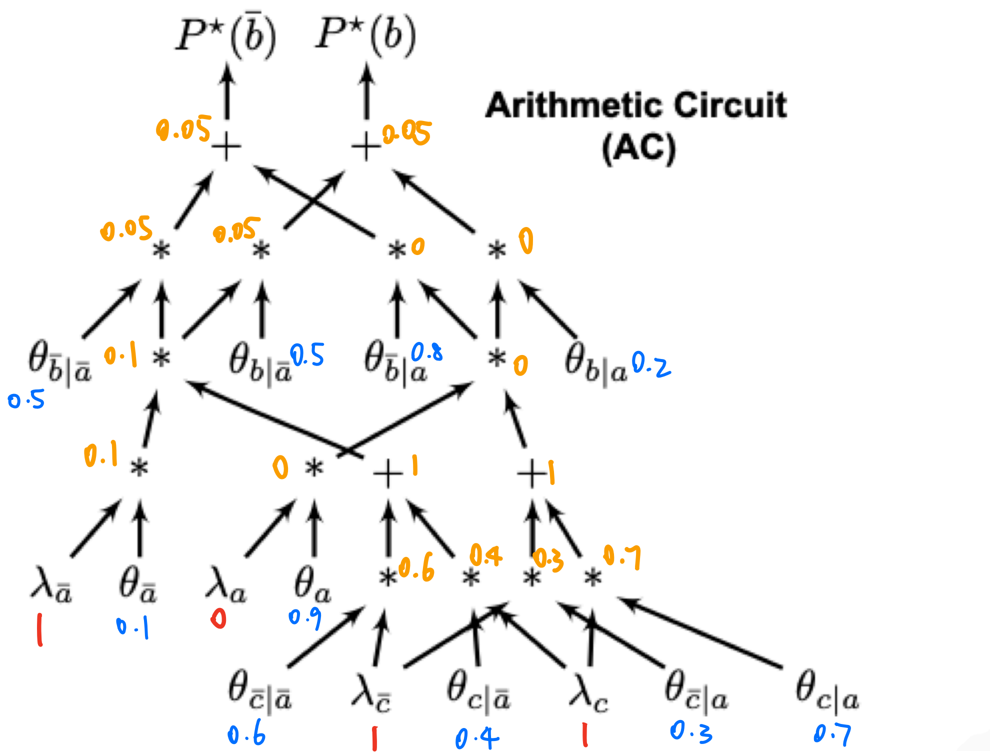
Hence, by evidence $e_1 = \bar{a}, c$, we have $P^*(\bar{b}) = 0.02$, $P^*(b) = 0.02$

② evidence $e_2 = \bar{a}, \bar{c} \Rightarrow \lambda_a = 0, \lambda_c = 0$



Hence, by evidence $e_2 = \bar{a}, \bar{c}$, we have $P^*(\bar{b}) = 0.03$, $P^*(b) = 0.03$

② evidence $e_3 = \bar{a} \Rightarrow \lambda_a = 0$



Hence, by evidence $e_3 = \bar{a}$, we have $P^*(\bar{b}) = 0.05$, $P^*(b) = 0.05$

(b).

For e_1, $p^*(\bar{b}|e_1)$ represents the unnormalized probability of $\bar{b}$ given $\bar{a}$ and c.
$p^*(b|e_1)$ represents the unnormalized probability of b given $\bar{a}$ and c.

For e_2, $p^*(\bar{b}|e_2)$ represents the unnormalized probability of $\bar{b}$ given $\bar{a}$ and $\bar{c}$.
$p^*(b|e_2)$ represents the unnormalized probability of b given $\bar{a}$ and $\bar{c}$.

For e_3, $p^*(\bar{b}|e_3)$ represents the unnormalized probability of $\bar{b}$ given $\bar{a}$.
$p^*(b|e_3)$ represents the unnormalized probability of b given $\bar{a}$.

(c)

$$Pr(\bar{b}|e_1) = \frac{Pr^*(\bar{b}|e_1)}{Pr^*(\bar{b}|e_1)+Pr^*(b|e_1)} = \frac{0.02}{0.02+0.02} = \frac{1}{2}$$

$$Pr(\bar{b}|e_2) = \frac{Pr^*(\bar{b}|e_2)}{Pr^*(\bar{b}|e_2)+Pr^*(b|e_2)} = \frac{0.03}{0.03+0.03} = \frac{1}{2}$$

$$Pr(\bar{b}|e_3) = \frac{Pr^*(\bar{b}|e_3)}{Pr^*(\bar{b}|e_3)+Pr^*(b|e_3)} = \frac{0.05}{0.05+0.05} = \frac{1}{2}$$