1. (35 pts) Consider the table below which represents a dataset by listing each unique example with the number of times it appears in the dataset. Construct the decision tree learned from this data by finding the most discriminating attribute at each step. Show precisely how you decided on the most discriminating attribute at each step by computing the expected entropies of the remaining attributes.

Example	Input Attributes			Class	#
	A	B	C	D	
$\mathbf{x}_1$	t	t	t	Yes	1
$\mathbf{x}_2$	t	t	f	Yes	6
$\mathbf{x}_3$	t	f	t	No	3
$\mathbf{x}_4$	t	f	f	No	1
$\mathbf{x}_5$	f	t	t	Yes	1
$\mathbf{x}_6$	f	t	f	No	6
$\mathbf{x}_7$	f	f	t	Yes	2
$\mathbf{x}_8$	f	f	f	No	2

Step 1:

Find split that minimizes conditional entropy of D

$ENT(D|A) = Pr(a)\,ENT(D|a) + Pr(\bar{a})\,ENT(D|\bar{a})$

$= 11/22 \cdot ENT(D|a) + 11/22 \cdot ENT(D|\bar{a})$

$ENT(D|a) = -[Pr(d|a)\log_2(Pr(d|a)) + Pr(\bar{d}|a)\log_2(Pr(\bar{d}|a))]$

$= -[7/11 \cdot \log_2 7/11 + \frac{4}{11} \cdot \log_2 4/11] = 0.946$

$ENT(D|\bar{a}) = -[Pr(d|\bar{a})\log_2(Pr(d|\bar{a})) + Pr(\bar{d}|\bar{a})\log_2(Pr(\bar{d}|\bar{a}))]$

$= -[3/11 \cdot \log_2 3/11 + 8/11 \cdot \log_2 8/11] = 0.845$

$ENT(D|A) = 0.47283 + 0.422668$

$= 0.8955$

$ENT(D|B) = Pr(b)\,ENT(D|b) + Pr(\bar{b})\,ENT(D|\bar{b})$

$= 14/22 \cdot ENT(D|b) + 8/22 \cdot ENT(D|\bar{b})$

$ENT(D|b) = -[Pr(d|b)\log_2(Pr(d|b)) + Pr(\bar{d}|b)\log_2(Pr(\bar{d}/b))]$

$= -[8/14 \cdot \log_2 8/14 + 6/14 \cdot \log_2 6/14] = 0.985257$

$ENT(D|\bar{b}) = -[Pr(d|\bar{b})\log_2(Pr(d|\bar{b})) + Pr(\bar{d}|\bar{b})\log_2(Pr(\bar{d}/\bar{b}))]$

$= -[2/8 \cdot \log_2 2/8 + 6/8 \cdot \log_2 6/8] = 0.81125$

$ENT(D|B) = 0.92198$

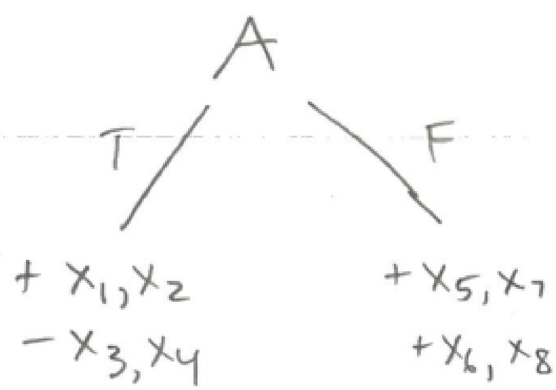
$ENT(D|C) = \overbrace{7/22(-[4/7 \log_2 4/7 + 3/7 \log_2 3/7])}^{ENT(D|c)} +$

$15/22(-[6/15 \log_2 6/15 + 9/15 \log_2 9/15])$

$= 0.31349 + 0.662$ $\quad ENT(D|\bar{c})$

$= 0.9755$

Splitting on A has the lowest conditional entropy:



Step 2: A=T branch
We only look at the examples where A=T $\{x_1, x_2, x_3, x_4\}$

$$ENT(D|B) = Pr(b)\, ENT(D|b) + Pr(\bar{b})\, ENT(D|\bar{b})$$
$$= 7/11\, ENT(D|b) + 4/11\, ENT(D|\bar{b})$$
$$ENT(D|b) = -[7/7 \log_2 7/7 + 0] = 0$$
$$ENT(D|\bar{b}) = -[0 + 4/4 \log_2 4/4] = 0$$

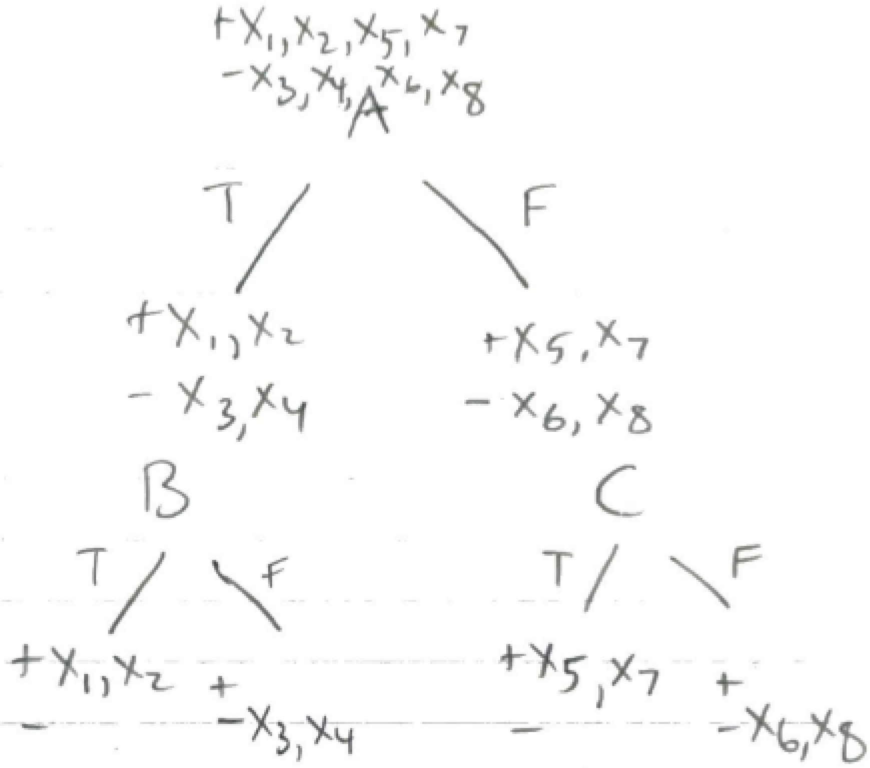
$ENT(D|B)=0$. This is the smallest conditional entropy possible, so we split on B for the A=T branch

A=F Branch
Look at examples where A=F $\{x_5, x_6, x_7, x_8\}$

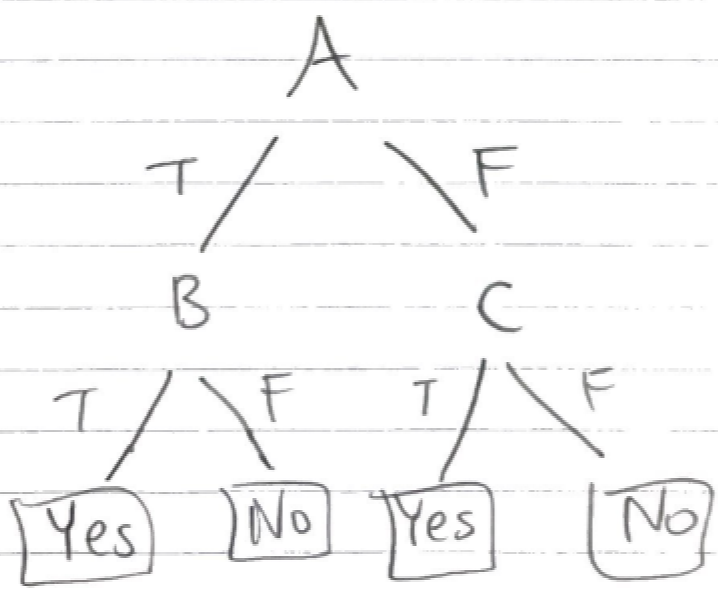
$$ENT(D|B) = 7/11\left(-[1/7 \log_2 1/7 + 6/7 \log_2 6/7]\right) +$$
$$4/11\left(-[2/4 \log_2 2/4 + 2/4 \log_2 2/4]\right)$$
$$= 0.3765 + 0.1369$$
$$= 0.5134$$

$$ENT(D|C) = 3/11\left(-[3/3 \log_2 3/3 + 0]\right) +$$
$$8/11\left(-[0 + 8/8 \log_2 8/8]\right)$$
$$= 0$$

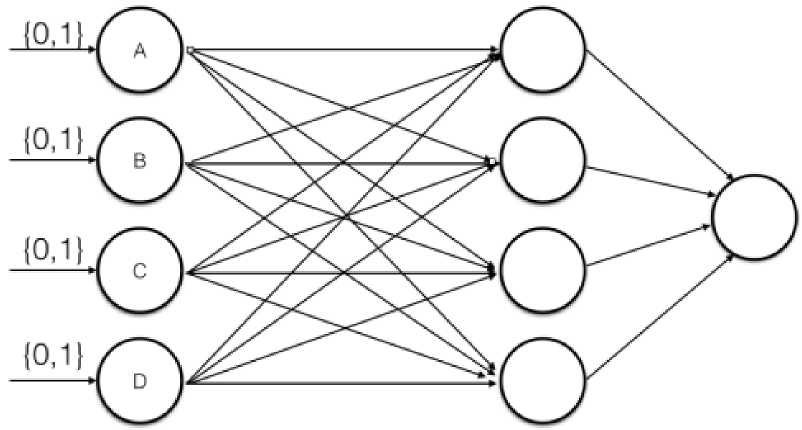
We split on C for the A=F branch.



Decision Tree:



2. (35 pts) Create a two layer neural network that uses the step function to implement $(A \vee \neg B) \oplus (\neg C \vee D)$, where $\oplus$ is the XOR function. You can either use the network structure provided below or another structure you construct. After drawing your network, clearly show the weights and activation function for each node. Assume values of $\{0, 1\}$ for each input variable. Note that solutions with more than two layers will not receive credit.

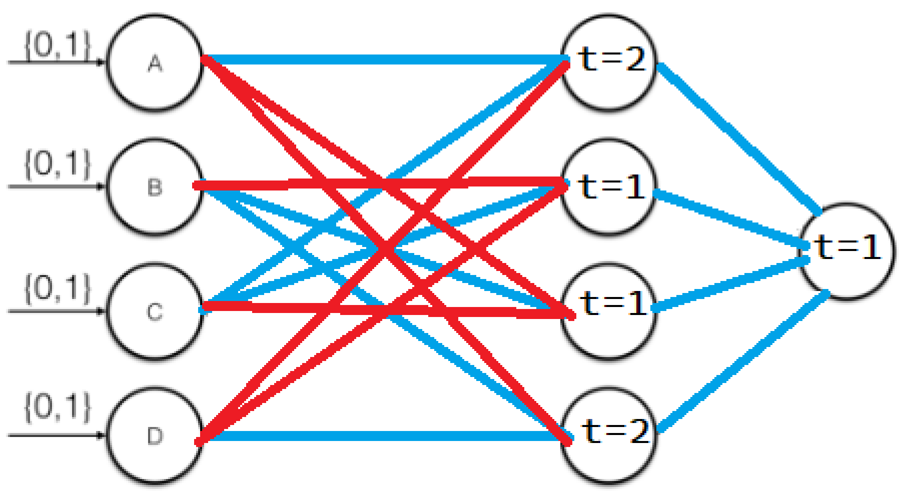


(A OR ~B) XOR (~C OR D)
The above equation can be written as:
(A AND C AND ~D) OR (~B AND C AND ~D) OR (~A AND B AND ~C) OR (~A AND B AND D)
1010, 1110 0010 1010 0100, 0101 0101, 0111
Truth table: (the inputs have numbers next to them corresponding to what sentence satisfies it)

ABCD	(A OR ~B) XOR (~C OR D)
0000	0
0001	0
0010 (2)	1
0011	0
0100 (3)	1
0101 (3, 4)	1
0110	0
0111 (4)	1
1000	0
1001	0
1010 (1, 2)	1

1011	0
1100	0
1101	0
1110 (1)	1
1111	0

For each node in the neural network, we have activation function g(x), where g(x) = 1 if x >= t, and g(x) = 0 if x < t.
A weight of 1 is represented by a blue arrow, and a weight of -1 is represented by a red arrow on the neural network below.
(A AND C AND ~D) OR (~B AND C AND ~D) OR (~A AND B AND ~C) OR (~A AND B AND D)
The output will be on as long as any of their hidden layer nodes have been activated (this represents the OR). The hidden layer nodes represent the AND clauses respectively.

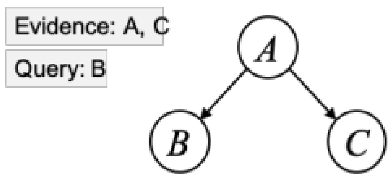


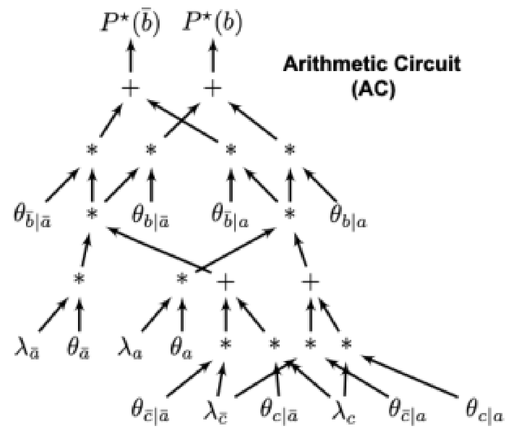
3. (30 pts) Consider the Arithmetic Circuit (AC) below and suppose the parameters have the following values: $\theta_a = .9$, $\theta_{b|a} = .2$, $\theta_{b|\bar{a}} = .5$, $\theta_{c|a} = .7$ and $\theta_{c|\bar{a}} = .4$. Consider three pieces of evidence: $\mathbf{e}_1 = \bar{a}, c$; $\mathbf{e}_2 = \bar{a}, \bar{c}$ and $\mathbf{e}_3 = \bar{a}$.

(a) (10 pts) Evaluate the AC under each piece of evidence. That is, compute the two circuit outputs under each evidence.

(b) (10 pts) What do the two circuit outputs represent for each piece of evidence $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$?

(c) (10 pts) Using the numbers obtained in the first part, compute $Pr(\bar{b}|\mathbf{e}_1)$, $Pr(\bar{b}|\mathbf{e}_2)$ and $Pr(\bar{b}|\mathbf{e}_3)$.

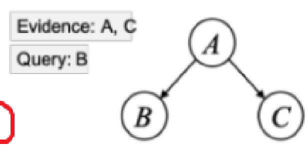




a)

e$_1$:

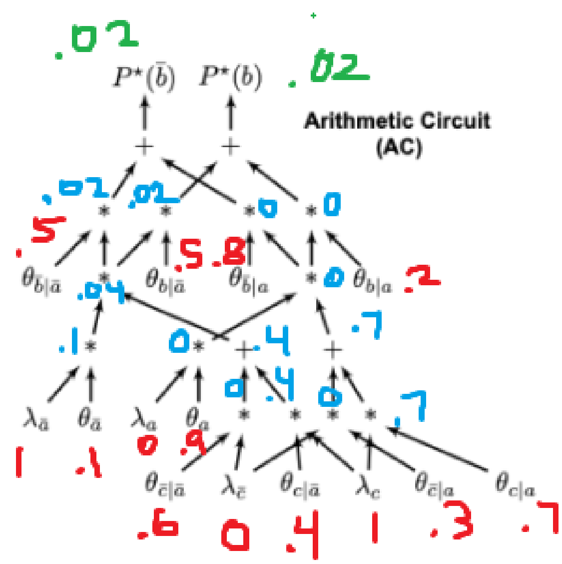
$e_1 : \bar{a}, c$



$\lambda_a = 0$

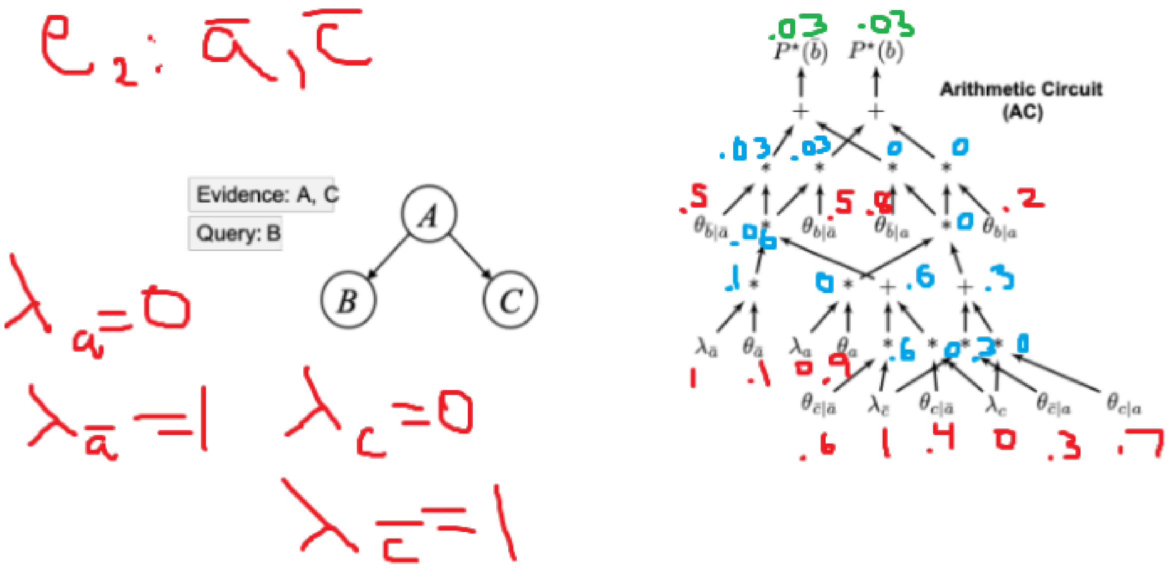
$\lambda_{\bar{a}} = 1$ $\lambda_c = 1$

$\lambda_{\bar{c}} = 0$



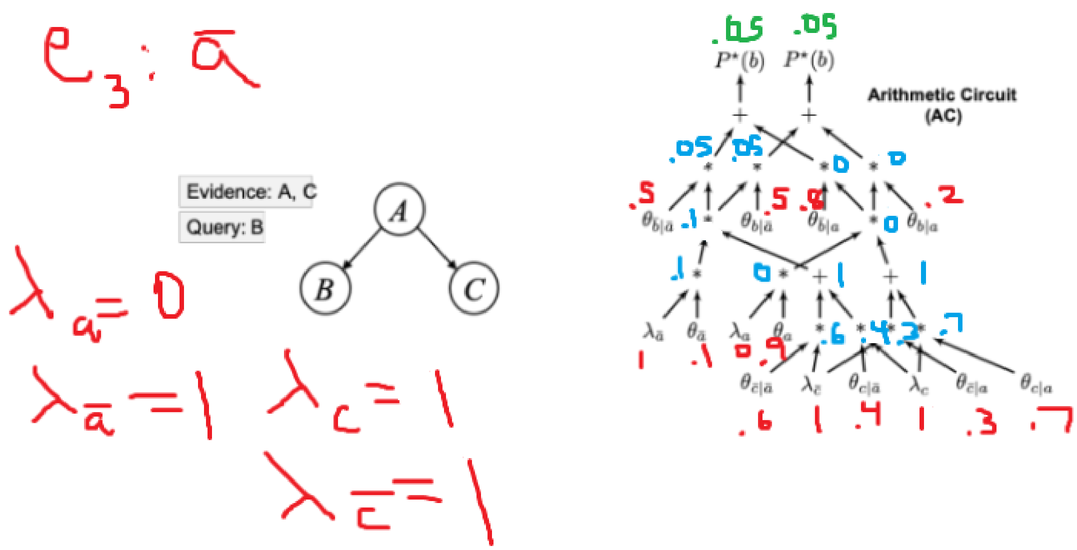
P*(~b) = 0.02, P*(b) = 0.02

$\mathbf{e}_2$:



P*(~b) = 0.03, P*(b) = 0.03

$\mathbf{e}_3$:



P*(~b) = 0.05, P*(b) = 0.05

b) The two circuit outputs represent the distribution on B. P*(b) is P(b, **e**) and P*(~b) is P(~b, **e**).

c) Pr(~b | $\mathbf{e}_1$) = Pr(~b, $\mathbf{e}_1$) / Pr($\mathbf{e}_1$) = ?

Pr($\mathbf{e}_1$) = Pr(b, $\mathbf{e}_1$) + Pr(~b, $\mathbf{e}_1$) = 0.02 + 0.02 = 0.04

Pr(~b | $\mathbf{e}_1$) = 0.02 / 0.04 = **0.5**

Pr(~b | $\mathbf{e}_2$) = Pr(~b, $\mathbf{e}_2$) / Pr($\mathbf{e}_2$)

$Pr(\mathbf{e}_2) = Pr(b, \mathbf{e}_2) + Pr(\sim b, \mathbf{e}_2) = 0.03 + 0.03 = 0.06$
$Pr(\sim b \mid \mathbf{e}_2) = 0.03 / 0.06 = \mathbf{0.5}$

$Pr(\sim b \mid \mathbf{e}_3) = Pr(\sim b, \mathbf{e}_3) / Pr(\mathbf{e}_3)$
$Pr(\mathbf{e}_3) = Pr(b, \mathbf{e}_3) + Pr(\sim b, \mathbf{e}_3) = 0.05 + 0.05 = 0.1$
$Pr(\sim b \mid \mathbf{e}_3) = 0.05 / 0.01 = \mathbf{0.5}$