Bradley Mont
CS 161
Disc 1C
804-993-030

Homework 9

1. 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	3

total: 23

Step 1

- find attribute that minimizes the conditional entropy of D:

A:

$$ENT(D|A) = \sum_a Pr(a) ENT(D|a)$$

$$= \left(\tfrac{11}{23}\right)\left(ENT(D|a)\right) + \left(\tfrac{12}{23}\right)\left(ENT(D|\bar{a})\right)$$

$$= \tfrac{11}{23}\left(-\left[Pr(d|a)\log_2(Pr(d|a)) + Pr(\bar{d}|a)\log_2(Pr(\bar{d}|a))\right]\right)$$

$$+ \tfrac{12}{23}\left(-\left[Pr(d|\bar{a})\log_2(Pr(d|\bar{a})) + Pr(\bar{d}|\bar{a})\log_2(Pr(\bar{d}|\bar{a}))\right]\right)$$

$$= \tfrac{11}{23}\left(-\left[\left(\tfrac{7}{11}\right)\log_2\left(\tfrac{7}{11}\right) + \left(\tfrac{4}{11}\right)\log_2\left(\tfrac{4}{11}\right)\right]\right)$$

$$+ \tfrac{12}{23}\left(-\left[\left(\tfrac{3}{12}\right)\log_2\left(\tfrac{3}{12}\right) + \left(\tfrac{9}{12}\right)\log_2\left(\tfrac{9}{12}\right)\right]\right)$$

$$= .45227 + .42329$$

$$ENT(D|A) = .8755$$

B:

$$ENT(D|B) = \sum_b Pr(b) ENT(D|b)$$

$$= \left(\tfrac{14}{23}\right)\left(ENT(D|b)\right) + \left(\tfrac{9}{23}\right)\left(ENT(D|\bar{b})\right)$$

$$= \tfrac{14}{23}\left(-\left[Pr(d|b)\log_2(Pr(d|b)) + Pr(\bar{d}|b)\log_2(Pr(\bar{d}|b))\right]\right)$$

$$+ \tfrac{9}{23}\left(-\left[Pr(d|\bar{b})\log_2(Pr(d|\bar{b})) + Pr(\bar{d}|\bar{b})\log_2(Pr(\bar{d}|\bar{b}))\right]\right)$$

$$= \tfrac{14}{23}\left(-\left[\left(\tfrac{8}{14}\right)\log_2\left(\tfrac{8}{14}\right) + \left(\tfrac{6}{14}\right)\log_2\left(\tfrac{6}{14}\right)\right]\right)$$

$$+ \tfrac{9}{23}\left(-\left[\left(\tfrac{2}{9}\right)\log_2\left(\tfrac{2}{9}\right) + \left(\tfrac{7}{9}\right)\log_2\left(\tfrac{7}{9}\right)\right]\right)$$

$$= .5997 + .299$$

$$ENT(D|B) = .8987$$

C:

$$ENT(D|C) = \sum_{c} Pr(c)\, ENT(D|c)$$

$$= \left(\tfrac{7}{23}\right)\left(ENT(D|c)\right) + \left(\tfrac{16}{23}\right)\left(ENT(D|\bar{c})\right)$$

$$= \tfrac{7}{23}\left(-\left[Pr(d|c)\log_2(Pr(d|c)) + Pr(\bar{d}|c)\log_2(Pr(\bar{d}|c))\right]\right)$$

$$+ \tfrac{16}{23}\left(-\left[Pr(d|\bar{c})\log_2(Pr(d|\bar{c})) + Pr(\bar{d}|\bar{c})\log_2(Pr(\bar{d}|\bar{c}))\right]\right)$$

$$= \tfrac{7}{23}\left(-\left[\left(\tfrac{4}{7}\right)\log_2\left(\tfrac{4}{7}\right) + \left(\tfrac{3}{7}\right)\log_2\left(\tfrac{3}{7}\right)\right]\right)$$

$$+ \tfrac{16}{23}\left(-\left[\left(\tfrac{6}{16}\right)\log_2\left(\tfrac{6}{16}\right) + \left(\tfrac{10}{16}\right)\log_2\left(\tfrac{10}{16}\right)\right]\right)$$

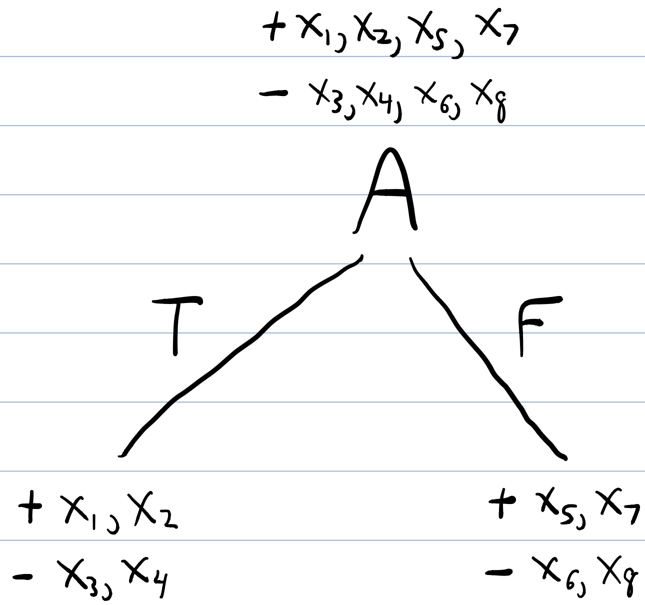
$$= .29985 + .66395$$

$$ENT(D|C) = .9638$$

Splitting on A results in the lowest conditional entropy for D, so we split on A

Step 2

- we split on A, so our decision tree looks like the following:



Step 2a: branch A=T

- our set of examples is now: $\{x_1, x_2, x_3, x_4\}$

B:

$$ENT(D|B) = \sum Pr(b)\,ENT(D|b)$$
$$= (7/11)(ENT(D|b)) + (4/11)(ENT(D|\bar{b}))$$
$$= \frac{7}{11}\left(-\left[Pr(d|b)\log_2(Pr(d|b)) + Pr(\bar{d}|b)\log_2(Pr(\bar{d}|b))\right]\right)$$
$$+ \frac{4}{11}\left(-\left[Pr(d|\bar{b})\log_2(Pr(d|\bar{b})) + Pr(\bar{d}|\bar{b})\log_2(Pr(\bar{d}|\bar{b}))\right]\right)$$
$$= \frac{7}{11}\left(-\left[(7/7)\log_2(7/7) + 0\log_2(0)\right]\right)$$
$$+ \frac{4}{11}\left(-\left[(0/4)\log_2(0) + 1\log_2(1)\right]\right)$$
$$= 0$$

Since $ENT(D|B) = 0$ gives the smallest conditional entropy possible, we split on B

Step 2b: branch A=F

- our set of examples is now $\{x_5, x_6, x_7, x_8\}$

B:

$$ENT(D|B) = \sum Pr(b)ENT(D|b)$$

$$= (7/12)(ENT(D|b)) + (5/12)(ENT(D|\bar{b}))$$

$$= 7/12\left(-\left[Pr(d|b)\log_2(Pr(d|b)) + Pr(\bar{d}|b)\log_2(Pr(\bar{d}|b))\right]\right)$$

$$+ 5/12\left(-\left[Pr(d|\bar{b})\log_2(Pr(d|\bar{b})) + Pr(\bar{d}|\bar{b})\log_2(Pr(\bar{d}|\bar{b}))\right]\right)$$

$$= 7/12\left(-\left[(1/7)\log_2(1/7) + (6/7)\log_2(6/7)\right]\right)$$

$$+ 5/12\left(-\left[(2/5)\log_2(2/5) + (3/5)\log_2(3/5)\right]\right)$$

$$= .3451 + .40456$$

$$ENT(D|B) = .7497$$

C:

$$ENT(D|C) = \sum_c Pr(c)ENT(D|c)$$

$$= (3/12)(ENT(D|c)) + (9/12)(ENT(D|\bar{c}))$$

$$= 3/12\left(-\left[Pr(d|c)\log_2(Pr(d|c)) + Pr(\bar{d}|c)\log_2(Pr(\bar{d}|c))\right]\right)$$

$$9/12\left(-\left[Pr(d|\bar{c})\log_2(Pr(d|\bar{c})) + Pr(\bar{d}|\bar{c})\log_2(Pr(\bar{d}|\bar{c}))\right]\right)$$

$$= \frac{3}{12}\left(-\left[(1)\log_2(1) + 0\log_2(0)\right]\right)$$

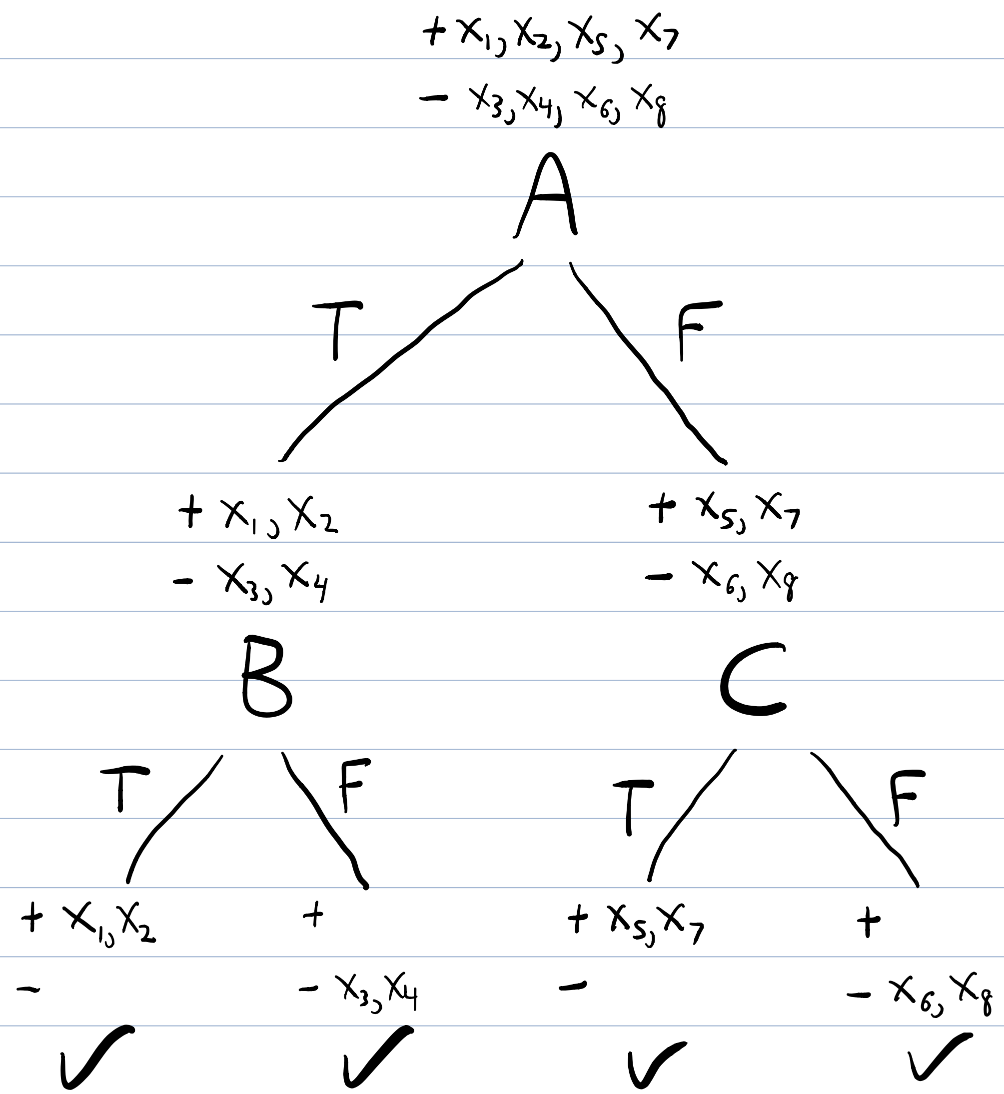
$$+ \frac{9}{12}\left(-\left[0\log_2 0 + 1\log_2(1)\right]\right)$$

$$ENT(D|C) = 0$$

Since $ENT(D|C) = 0$ gives the smallest conditional entropy possible, we split on C

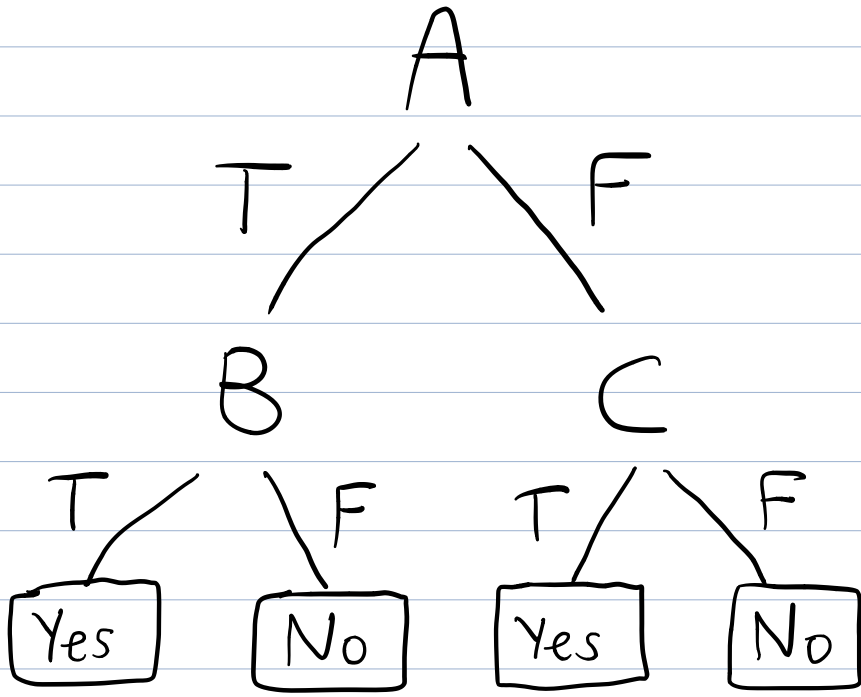
Step 3

- we split on B for the branch where A=T, and we split on C for the branch where A=F

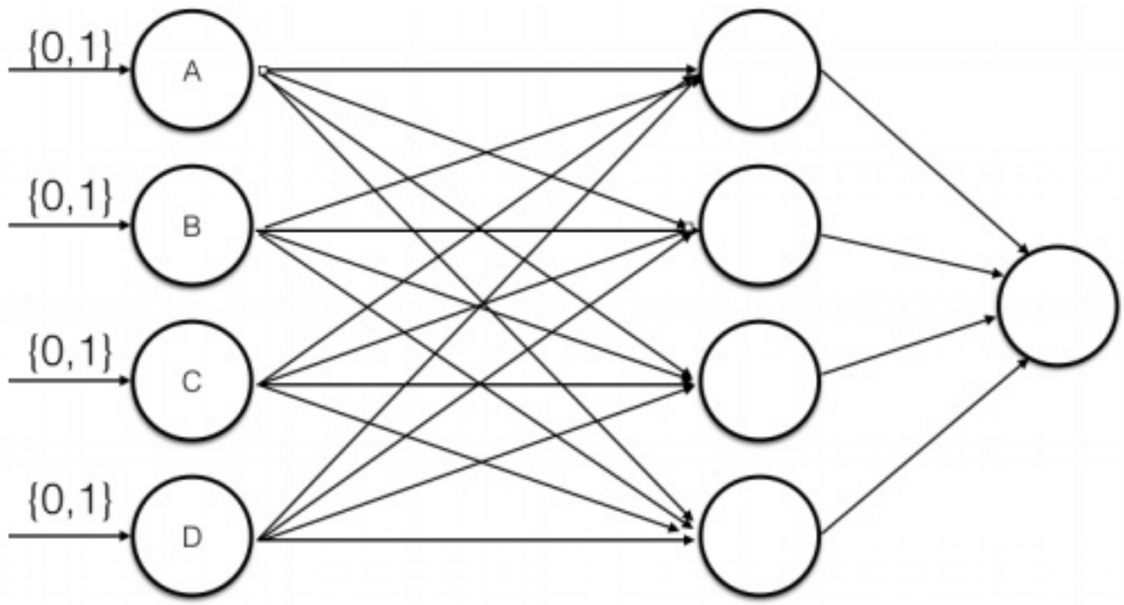


done!

Decision Tree:



2. 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 inputs of $\{0, 1\}$ for each input variable. Note that solutions with more than two layers will still receive partial credit.



$$\Delta = \underbrace{(A \vee \neg B)}_{X} \oplus \underbrace{(\neg C \vee D)}_{Y}$$

<u>Convert Δ to CNF</u>

- $X \oplus Y \approx (X \vee Y) \wedge (\neg X \vee \neg Y)$

$$\Delta = \left[(A \vee \neg B) \vee (\neg C \vee D)\right] \wedge \left[\neg(A \vee \neg B) \vee \neg(\neg C \vee D)\right]$$

$$\Delta = \left[(A \vee \neg B) \vee (\neg C \vee D)\right] \wedge \left[(\neg A \wedge B) \vee (C \wedge \neg D)\right]$$

$$\Delta = (A \vee \neg B \vee \neg C \vee D) \wedge \left[\left[\neg A \vee (C \wedge \neg D)\right] \wedge \left[B \vee (C \wedge \neg D)\right]\right]$$

$$\Delta = (A \vee \neg B \vee \neg C \vee D) \wedge \left[(\overset{\alpha}{C} \wedge \overset{\beta}{\neg D}) \vee \overset{\gamma}{\neg A}\right] \wedge \left[(C \wedge \neg D) \vee B\right]$$

$$\Delta = (A \vee \neg B \vee \neg C \vee D) \wedge \left[(C \vee \neg A) \wedge (\neg D \vee \neg A)\right] \wedge \left[(C \vee B) \wedge (\neg D \vee B)\right]$$

$$\Delta = \overset{1}{(A \vee \neg B \vee \neg C \vee D)} \wedge \overset{2}{(C \vee \neg A)} \wedge \overset{3}{(\neg D \vee \neg A)} \wedge \overset{4}{(C \vee B)} \wedge \overset{5}{(\neg D \vee B)}$$

Layer 1: individual clauses
Layer 2: connecting clauses

clauses:

1. $A \lor \neg B \lor \neg C \lor D$
2. $C \lor \neg A$
3. $\neg D \lor \neg A$
4. $C \lor B$
5. $\neg D \lor B$

Clause 1

$A \vee \neg B \vee \neg C \vee D$

$A - B - C + D$

A	B	C	D	$A \vee \neg B \vee \neg C \vee D$	g
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	$\boxed{-2}$
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

$w_A = 1, \; w_B = -1, \; w_c = -1, \; w_d = 1$

$t = -1.5$

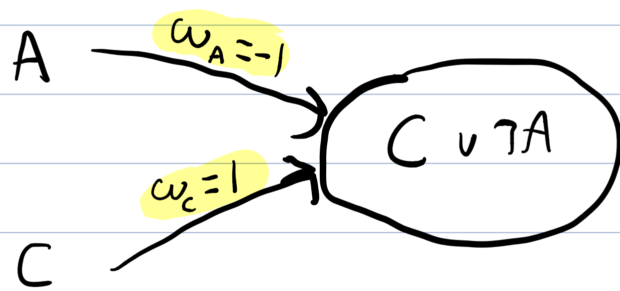
$g = A - B - C + D \geq -1.5$

Clause 2

$C \vee \neg A$

A	C		
0	0	1	✓
0	1	1	✓
1	0	0	✓
1	1	1	✓

$t = -.5$



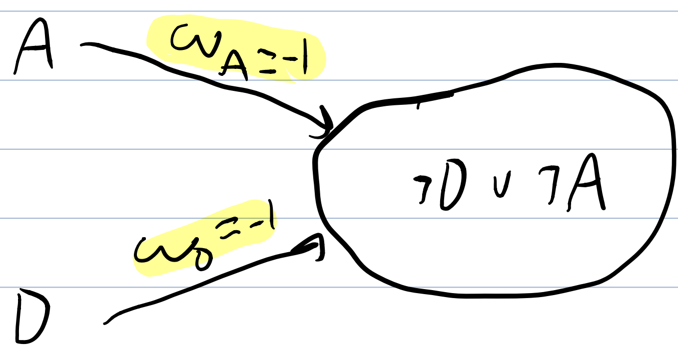
$g = -A + C \geq -.5$

Clause 3

$\neg D \vee \neg A$

or

A	D		g
0	0	1	0
0	1	1	-1
1	0	1	-1
1	1	0	-2



$t = -1.5$

$g = -A - D \geq -1.5$

Clause 4

$C \vee B$

B	C	
0	0	0
0	1	1
1	0	1
1	1	1

$w_B = 1$

$w_C = 1$

$t = .5$

Clause 5

$\neg D \vee B$

B	D		g
0	0	1	0
0	1	0	-1
1	0	1	1
1	1	1	0

$w_B = 1$

$w_D = -1$

$t = -.5$

$g = B - D \geq -.5$

2nd Layer

- since it's CNF, we need every clause to be true, so we set our threshold as 4.5. We will only meet that threshold when every clause outputs 1
- give all inputs a weight of 1

Neural Network

- each activation function is a step function with threshold t
- weights given in red