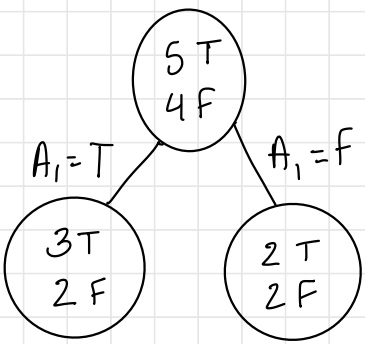
Ann Biju
axb190082
CS 4375.002

Assignment 3

(6a) $P(+) = 5/9$; $P(-) = 4/9$

Entropy $= -\frac{5}{9}\log_2\left(\frac{5}{9}\right) + \frac{4}{9}\log_2\left(\frac{4}{9}\right) =$ **-0.0489**

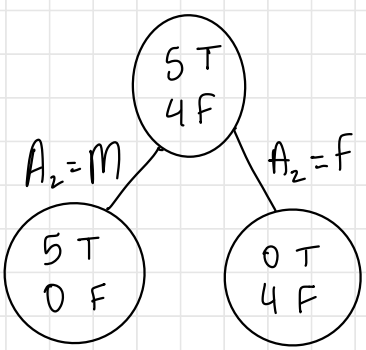
(6b)



Entropy: $4/9[-(3/5)\log_2(3/5) - (2/5)\log_2(2/5)] + 5/9[-(2/4)\log_2(2/4) - (2/4)\log_2(2/4)] =$

0.9871

$\text{Gain}_{A_1} = -0.0489 - 0.9871 = -1.036$



Entropy: $5/9[-(5/5)\log_2(5/5) - (0/5)\log_2(0/5)] + 4/9[-(0/4)\log_2(0/4) - (4/4)\log_2(4/4)] = 0$

$\text{Gain}_{A_2} = -0.0489 - 0 = -0.0489$

A3:

Sorted → 2 3 4 5 6 7 8

Split Points → 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5

Entropies →

	1.5	
	≤	>
+	0	5
−	0	4

	2.5	
	≤	>
+	1	4
−	0	4

	3.5	
	≤	>
+	2	3
−	0	4

	4.5	
	≤	>
+	3	2
−	0	4

	5.5	
	≤	>
+	3	2
−	2	2

	6.5	
	≤	>
+	3	2
−	3	1

	7.5	
	≤	>
+	4	1
−	3	1

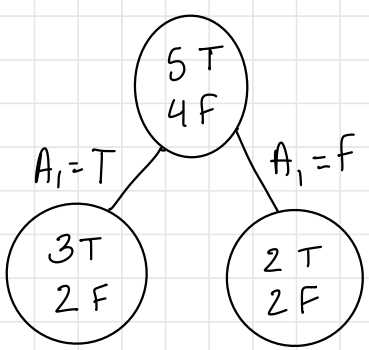
	8.5	
	≤	>
+	5	0
−	4	0

-0.0489, .22, 0.17, 0.116, 0.107, 0.064, 0.039, 0

The best split for A3 is 8.5.

6c. A3 is the best split amongst the 3.

6d.

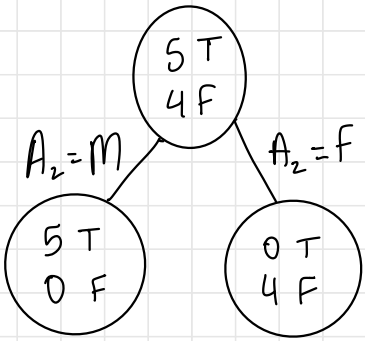


Error = 5/9 = 0.56

Error = 5/9(2/5) + 4/9(2/4)

= 0.44

gain = 0.11 → A1



$$\text{error} = \frac{4}{9}\left(\frac{5}{5}\right) + \frac{5}{9}\left(\frac{4}{4}\right)$$

$$= 1$$

$$\text{gain} = -0.44 \rightarrow A_2$$

We should split based on A_1.

6e. $GINI = 1 - (5/9)^2 - (4/9)^2 = 0.494$

$GINI = 5/9\left(1 - (3/5)^2 - (2/5)^2\right) + 4/9\left(1 - (2/4)^2 - (2/4)^2\right)$

$= 0.489$

$\text{gain}_{A_1} = 0.005$

$GINI = 5/9\left(1 - (5/5)^2\right) + 4/9\left(1 - 0^2 - (4/4)^2\right)$

$= 0$

$\text{gain}_{A_2} = .494$

So split at A_2

7a. The error on the training set is equal to the error on the testing set.

7b. Before split

resubmission error = 4/10
generalization error = 4.5/10

After split

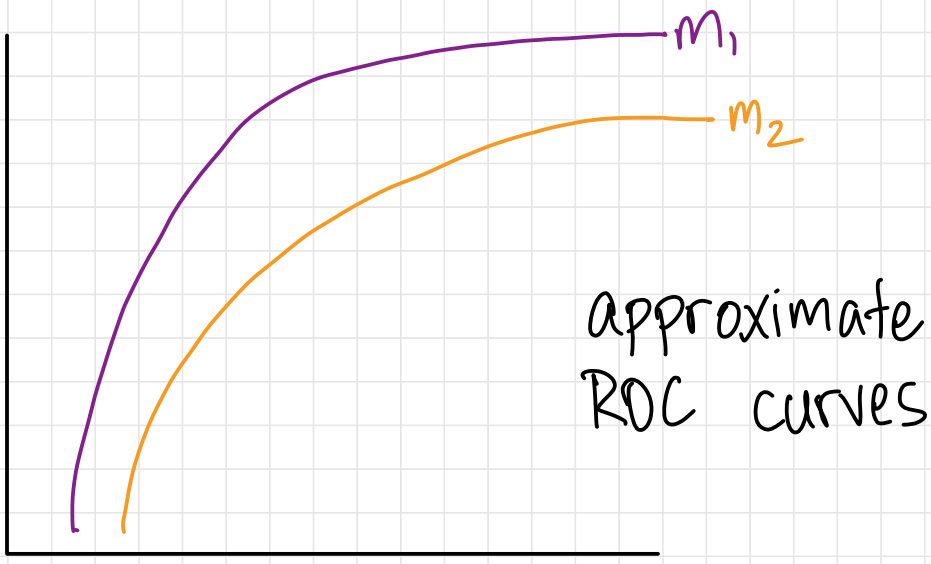
resubmission error =

$$\left(\frac{1}{1}\right)\left(\frac{1}{1}\right) + \left(\frac{1}{1}\right)\left(\frac{1}{1}\right) = 2$$

↳ trim subtree

7c. Tree must be pruned/trimmed as in part b.

8a.



approximate ROC curves

8b. Precision $= \frac{TP}{TP+FP} = \frac{1}{1+4} = \frac{1}{5}$

Recall $= \frac{TP}{TP+FN} = \frac{1}{1+2} = \frac{1}{3}$

F-measure $= \frac{2(TP)}{2(TP)+FN+FP} = \frac{2}{2+2+4} = \frac{2}{8}$