1).

a) No because $P(B|A)$ is calculated using the probability of B, or $P(B)$ which we do not know, or $P(A)$. Because we lack $P(A \cap B)$ we cannot calculate for $P(B)$

b) No, because we don't know the probability of B, $P(B)$ we cannot calculate for $P(A)$ which is needed

c) $P(A \cap B) + P(A \cap \sim B) = P(A)$

$P(A|B) = \frac{P(A \cap B)}{P(B)}$ $\quad \frac{3}{4} \cdot \frac{1}{4} = \frac{P(A \cap B)}{1/4} \cdot \frac{1}{4}$

$P(A \cap B) = \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{16}$

$P(\sim B) = 1 - P(B) = 1 - \frac{1}{4} = \frac{3}{4}$

$P(A|\sim B) = \frac{P(A \cap \sim B)}{P(\sim B)}$ $\quad \frac{1}{4} = \frac{P(A \cap \sim B)}{\frac{3}{4}} \cdot \frac{3}{4}$

$P(A \cap \sim B) = \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16}$ $\quad P(A) = \frac{3}{16} + \frac{3}{16} = \frac{6}{16} = \frac{3}{8}$

$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)} = \frac{\frac{3}{16}}{\frac{3}{8}} = \frac{3}{16} \cdot \frac{8}{3} = \frac{8}{16} = 0.5$

~~$P(B|A) > 0.5$~~

d) $P(B|A) = \frac{P(A|B)\,P(B)}{P(A)} = \frac{(\frac{3}{4})(\frac{1}{4})}{(\frac{3}{8})} = \frac{\frac{3}{16}}{\frac{3}{8}} = \frac{3}{16} \cdot \frac{8}{3} = 0.5$ *

2)

a) for $a \le x \le b$, for $c \le y \le b$

$a=0, b=0$
$a=0, b=1$
$a=0, b=2$
$a=1, b=1$
$a=1, b=2$
$a=2, b=2$ } 6

$c=0\ d=0$
⋮
$c=2, d=2$ } 6

$6 \cdot 6 = \boxed{36}$

b) for $a \le x \le b$

$a=-2\ b=-2$
$a=-2\ b=-1$
$a=-2\ b=0$
$a=-2\ b=1$
$a=-2\ b=2$
$a=-1\ b=-1$
$a=-1\ b=0$
$a=-1\ b=1$
$a=-1\ b=2$ } 9

$a=0$ ⋮ $a=2\ b=2$ } 6

$9+6=15$

$15 \cdot 6 = 9$

assume $a, b, c, d \in \{0,1,2\}$

c)

$a=0\ b=0\ c=0\ d=0$
$a=1\ b=1\ c=1\ d=1$
$a=2\ b=2\ c=2\ d=2$ } 3

$a=0\ b=1$
$a=1\ b=2$
$c=0\ d=0$
$c=0\ d=1$
$c=1\ d=1$
$c=1\ d=2$
$c=2\ d=2$ } 5

$a=0\ b=2$
$c=0\ d=0$
⋮
$c=0\ d=2$ } 6

$3(3) + 2(5) + 1(6) = 9+10+6 = \boxed{25}$

d) When H is used in conjunction with C, one could simply use the list of all such ML algorithms on H to find a subset of H that is consistent with C1. This is because C1 is simply a subset of C due to the added constraint, which means H1 is simply some of a subset of H.

e) When H1 is used in conjunction with C, one could use candidate elimination, using the most general list to simply $a, b, c, d \in \{0,1,2\}$ and the most specified list to + to H1. From there, the only examples in C will cause the algorithm to generalize H1, as it's more constrained from H. These examples will only be ones such that $H(x)=1$ but $H1(x)=0$, such as $a=0\ b=0\ c=1\ d=2$.

3) (A OR B) AND (C OR ~D) = F

#HG See coffee.ppt to evaluate.

- P(+) = 9/16
- P(-) = 7/16

Entropy = H(9/16, 7/16) = $-\left(\log_2(9/16)\cdot 9/16 + \log_2(7/16)\cdot 7/16\right)$

= 0.9887

$E_{\{A=true\}}$ ... ; P(+) = 6/8 ; 2 3 $E_{\{A=F\}}$; P(+) = 3/8 P(-) = 5/8

Entropy$_{A=true}$ = 0.8113 Entropy$_{A=F}$ = 0.9528

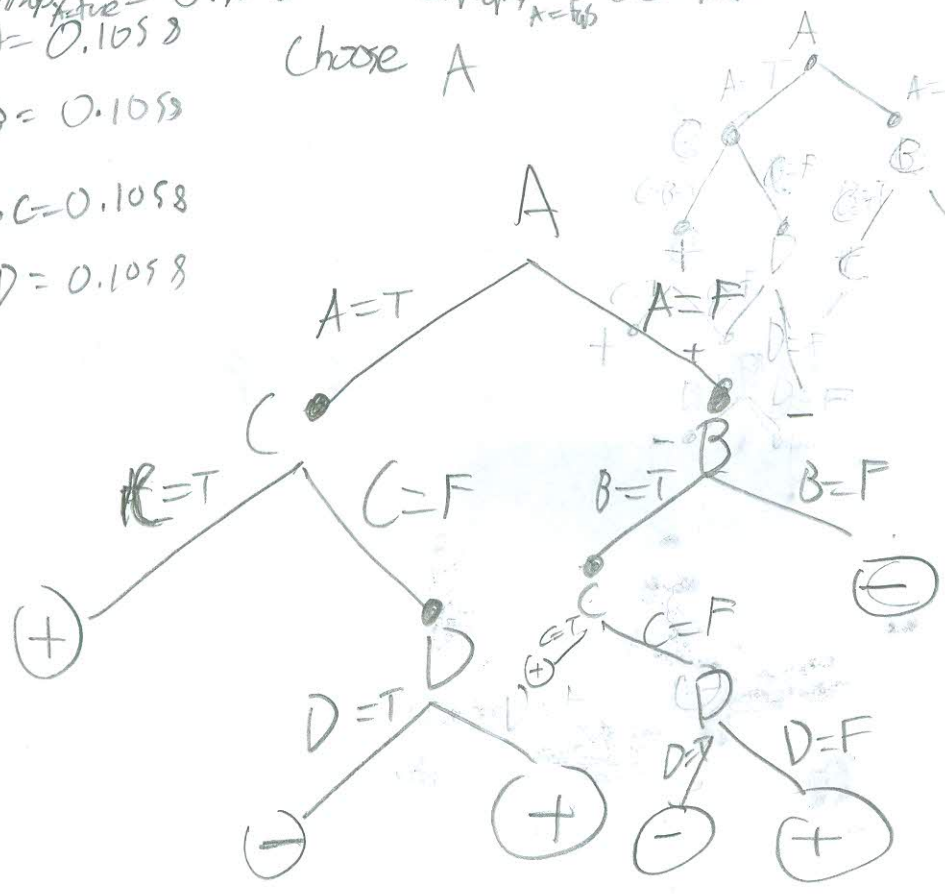
IG.A = 0.1058

IG.B = 0.1058

IG.C = 0.1058

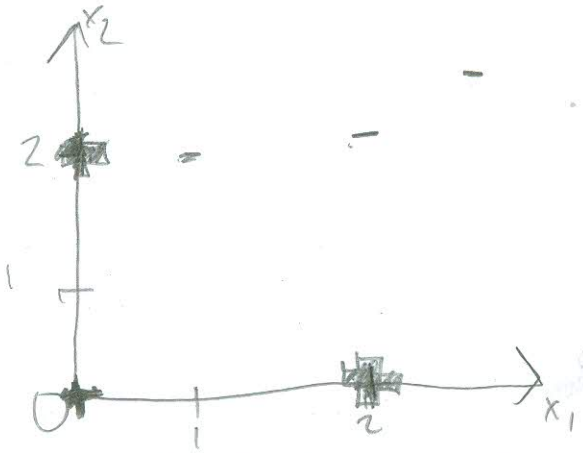
IG.D = 0.1058

Choose A



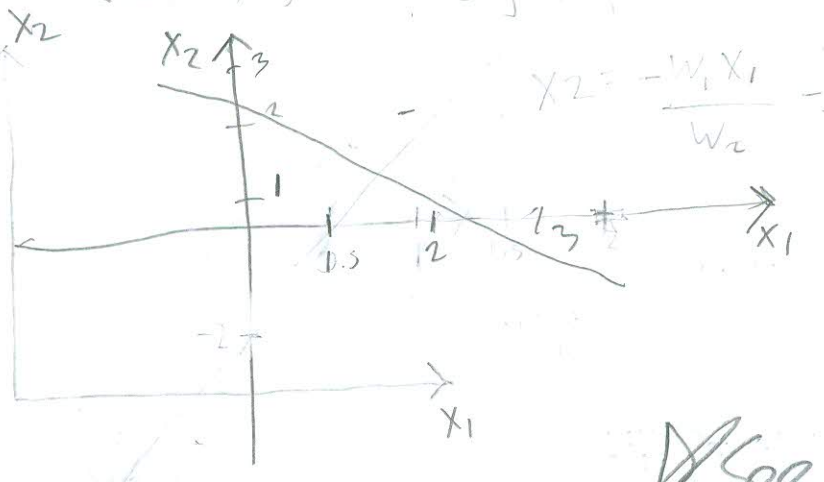
~~See~~ prob 3. coffee

4) a



No Here are not linearly Seperable

b) $w = \langle 0.3606, -0.1626, -0.1580 \rangle$

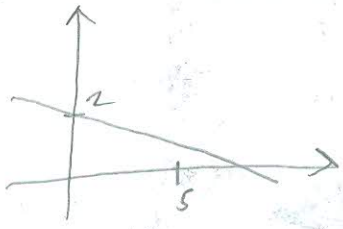


$x_2 = \frac{-w_1 x_1}{w_2} - \frac{w_0}{w_2}$

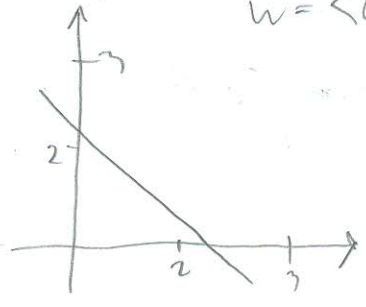
See prob4.coffee

c) Removing $\langle 2, 2, - \rangle$

$w = \langle 0.6513, -0.1083, -0.2945 \rangle$



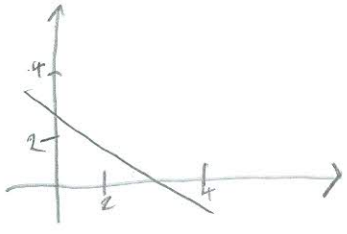
Removing $\langle 2.5, 2.5, - \rangle$ $w = \langle 0.4144, -0.1817, -0.4317 \rangle$



The closest is the one with $\langle 1, 2, - \rangle$ removed

Removing $\langle 1, 2, - \rangle$

$w = \langle 0.3941, -0.1241, -0.1251 \rangle$



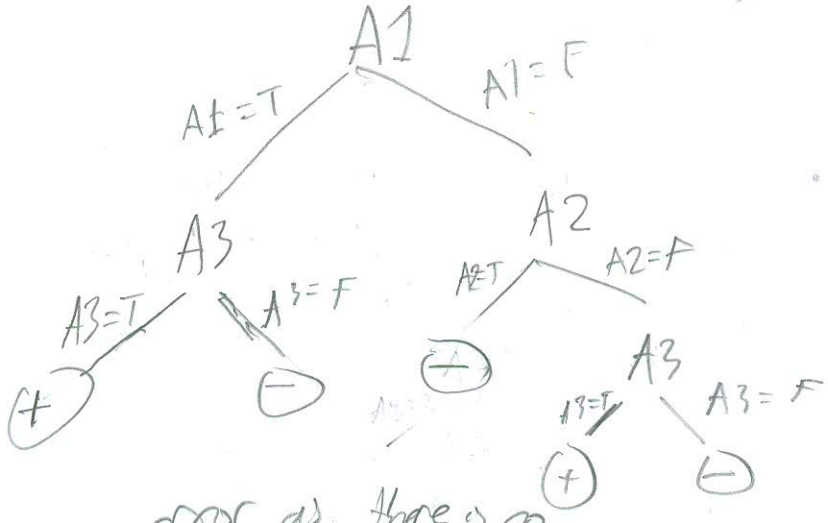
5) total entropy is 1

a) $-\left(\left(\log_2 0.5\right)\cdot 0.5 + \left(\log_2 0.5\right)\cdot 0.5\right) = -(-0.5 - 0.5) = 1$

b) $H(A3 \mid A2=T) = 0.\overline{6}$

See prob5.coffee

c)



d) There is no error as there is no noise data, thus error is not introduced. This can be manually observed

e) A1=T, A2=T, A3=F, Class = −

A1=F A2=F A3=F, Class = −

6) Play Tennis = +

See prob6.coffee

1)

a) $H = \langle \phi, \phi, \phi, \phi \rangle \xrightarrow{3} \langle Overcast, Hot, High, weak \rangle \xrightarrow{\text{example 4}} \langle ?, ?, High, weak \rangle$

examples $\langle ?, ?, ?, weak \rangle \xrightarrow{\text{example}} \langle ?, ?, ?, ? \rangle$ Stop here as it cannot become more general

$H = \langle ?, ?, ?, ? \rangle$

$H(\langle sunny, cool, normal, weak \rangle) = 1$

b) Candidate Elimination Produces an empty G and S set and this does not work for this classification

~~See~~ Rob 7. coffee