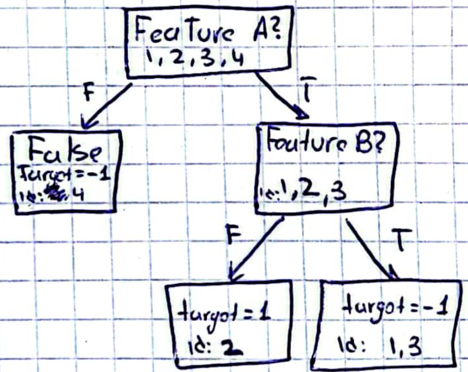
1.1)

	Attribute	$\frac{\lvert V_{a=T}\rvert}{\lvert V\rvert}$	$\frac{\lvert V_{a=F}\rvert}{\lvert V\rvert}$	$H(V_{a=T})$	$H(V_{a=F})$	$IG(v,a)-H(v)$
~~Feature A~~	Feature A	3/4	1/4	$H(2/3)$	$H(0/1)$	$-\frac{3}{4}H(\frac{2}{3})-\frac{1}{4}H(\frac{0}{1})=-0.689$
~~Feature B~~	Feature B	2/4	2/4	$H(1/2)$	$H(1/2)$	$-\frac{2}{4}H(\frac{1}{2})-\frac{2}{4}H(\frac{1}{2})=-1$
~~Feature C~~	Feature C	2/4	2/4	$H(1/2)$	$H(1/2)$	$-\frac{2}{4}H(\frac{1}{2})-\frac{2}{4}H(\frac{1}{2})=-1$

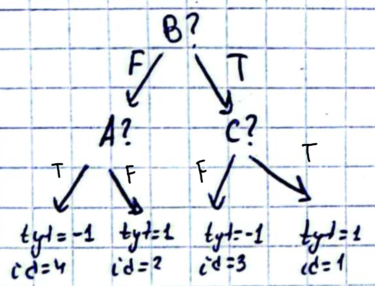
"T" = 1
"F" = 0



	Attribute	$\frac{\lvert V_{a=T}\rvert}{\lvert V\rvert}$	$\frac{V_{a=F}}{\lvert V\rvert}$	$H(V_{a=T})$	$H(V_{a=F})$	$IG(v,a)-H(V)$
T	Feature B	2/3	1/3	$H(\frac{1}{2})$	$H(\frac{1}{1})$	$-\frac{2}{3}(H(\frac{1}{2}))-\frac{1}{3}H(\frac{1}{1})=-\frac{2}{3}$
T	Feature C	1/3	2/3	$H(1/1)$	$H(1/2)$	$-\frac{1}{3}H(1)-\frac{2}{3}H(\frac{1}{2})=-\frac{2}{3}$
F	~~Feature B~~	~~0/1~~	~~1/1~~	~~H(0)~~	~~H~~	
F	~~Feature C~~	~~1/1~~	~~0/1~~			

נשים לב כי רק id=1 מסווג לא נכון ולכן יש שגיאה של 25% $(\frac{1}{4})$

1.2)

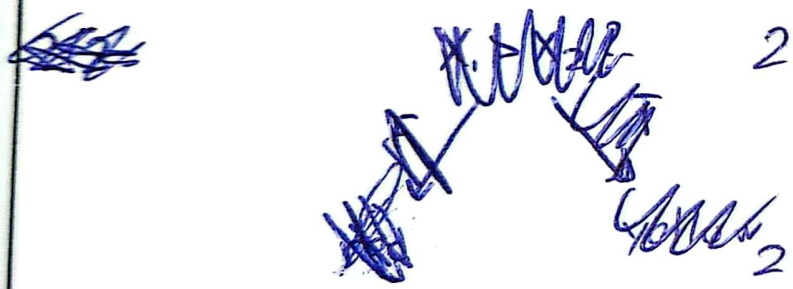


2.1

$$h(x_1, x_2) = \text{Sign}(x_1 - x_2) = \text{Sign}\left((1, -1)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + 0\right)$$

red = -1

yellow = 1

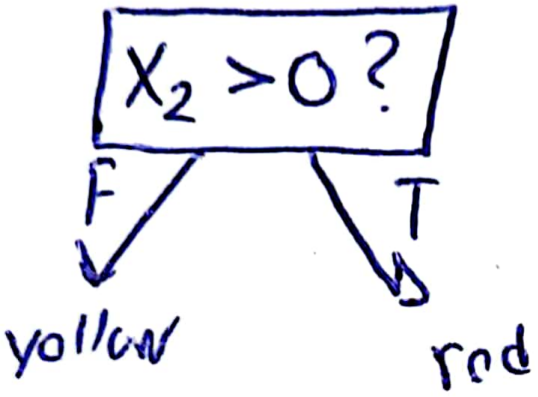


2.2) there's no way to ~~beat~~ solve the problem with a depth 1 decision tree because we can only draw lines that are parallel to the x axis or y axis (one line)

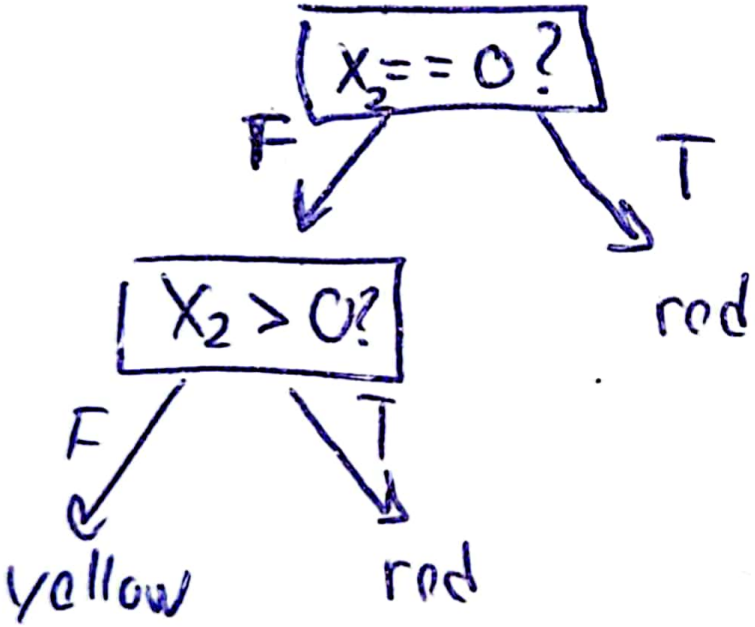
2.3) ~~same reason~~ same answer with the same reason except this time we have a depth of 2 which still isn't enough (2 lines)

2.4.1. $h(x_1, x_2) = Sign(-x_2) = Sign\left((0,-1)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + 0\right)$

2.4.2.



2.4.3.



3.1. the problem with the proof is that it didn't contradict anything: having both Entropy(D) and Entropy(D_v) be non-negative <u>does NOT</u> contradict the assumption that $IG(A) < 0$ because

$\underbrace{\text{Entropy}(D)}_{I} < \underbrace{\text{Sum}\left(\left[\frac{|D_v|}{|D|}\text{Entropy}(D_v)\right]\right)}_{II}$ and so the subtraction of II from I will be negative ($IG(A) < 0$ still).

3.2. $IG(V,a) = H(V) - \frac{|V_a=T|}{|V|}H(V_a=T) - \frac{|V_a=F|}{|V|}H(V_a=F) =$

$= H(V) - \underbrace{\frac{|V_a=T|}{|V|}}_{\alpha} H\left(\underbrace{\frac{|\{(x,y)\in V_a=T \mid y=1\}|}{|V_a=T|}}_{\beta_1}\right) - \underbrace{\frac{|V_a=F|}{|V|}}_{1-\alpha} H\left(\underbrace{\frac{|\{(x,y)\in V_a=F \mid y=1\}|}{|V_a=F|}}_{\beta_2}\right) =$

$\geq H(V) - \alpha H(\beta_1) - (1-\alpha) H(\beta_2) \geq H(V) - H(\alpha\beta_1 + (1-\alpha)\beta_2) =$ (Jensen property)

$= H(V) - H\left(\frac{|V_a=T|}{|V|}\frac{|\{(x,y)\in V_a=T \mid y=1\}|}{|V_a=T|} + \frac{|V_a=F|}{|V|}\cdot\frac{|\{(x,y)\in V_a=F \mid y=1\}|}{|V_a=F|}\right) =$

$= H(V) - H\left(\frac{|\{(x,y)\in V_a=T \mid y=1\}| + |\{(x,y)\in V_a=F \mid y=1\}|}{|V|}\right) = H(V) - H(V) = 0$

In conclusion; $IG(V,a) \geq 0$

4.1 Both ~~w~~ 'w' and 'b' can be scaled by the same positive constant λ without affecting the classifiers decision rule, thus having ~~a~~ the ~~bigger~~ same degree of "freedom" ~~as~~ a homogenous classifier.

4.2 A is b_1; B is b_4; C is b_2; D is b_3

5.1

M/D	A	B
i	No, because the (0,0) dot will be classified as blue instead of red	No, because the (0,0) dot will be classified as red instead of blue
ii	No, will ~~always~~ classify all dots as blue ~~because~~ since the majority (2 out of 3) are always blue	yes
iii	yes	No, because the (0,0) will be classified as red instead of blue

5.2

M/D	A	B
i	unchanged because rotation doesn't change the values of the 2nd norm U	U
ii	U	U
iii	rotate by 45° and the answer changes s.t the dot in the middle will be classified as blue instead of red	might change for a 45°, everything including the middle dot will be classified correctly