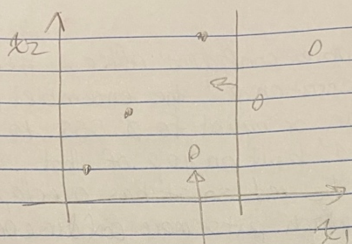
Q4.

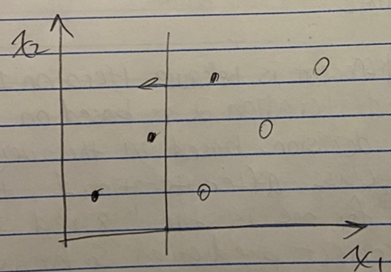
a).




~~this point will~~ gain weight for the next iteration

Because is a negative point classified as a positive point, this will create error rate and that the mis-classified point will gain weight for the next iteration.

b).




c). $\alpha_1$ will have higher coefficient in ensemble because as the first iteration, it classifies all positive points correctly with 1 misclassification on a negative point. at this point, all weights are equal. It is until the second iteration, that previous misclassified point gained more weights than all the other points, but the classification still got 1 positive point wrong. thus the error rate is higher and that $\alpha_2$ is smaller.

Q2

a). The ensemble is more confident after 2 iterations because on the 'Iteration ∅' the ensemble predicts points that are greater than or equal to ∅.558 to be red points and 'Iteration 1' built on top of that, 'Iteration 2' predict all points to be red so it has overlapping region with 'Iteration ∅' which puts more confidence on those "double confirmed" points, also because based on the accuracy of 'Iteration ∅' is higher than 'Iteration 1' therefore the the ensemble is more confident in the overlapping region. (those 8 red samples)

Same reason as why ensemble not as confident about any of the other samples. because ensemble chose 'Iteration∅' prediction over 'iteration 1' due to higher accuracy. and smaller error rate.

b). The prediction is different is because Iteration 1 is based on Iteration ∅ and iteration 2 is based on iteration 1. because it's making decisions based on the weighted sum of samples in each class. after iteration 1, there are 3 weighted sum, for one is the "8 red sampes" that have consistent correct prediction so far which has the smallest weight, then there are the other 2 parts predicted incorrectly once individually so that each of the 2 parts will have a different weighted sum as well.

On Iteration ∅, there were no weights on the blue points, but after iteration 1, the blue points gain weights, therefore, in iteration 2, in consistent with the same cut of point from iteration 1, but classify the blue points (because large weights and small 8 red samples weights) into a blue classification region.

j) $1 = w_\phi + (-0.2(-2) + 0.6(4))$

$1 = w_\phi + (0.4 + 2.4)$

$1 = w_\phi + 2.8$

$w_\phi = -1.8$

k). Yes they are the same.

for 3(e) the equation is

$\phi = \frac{1}{3}x_1 - x_2 + 3$

which has an ratio $\frac{1}{3} : 1 : 3$

for the other equation is

$\phi = -0.2x_1 + 0.6x_2 - 1.8$

which has the same ratio $\frac{1}{3} : 1 : 3$

g). $\alpha_1 + \alpha_2 - 0.5\Big[(\alpha_1\alpha_1(1)(1)(20)) + (\alpha_1\alpha_2(1)(-1)6)$
$+ (\alpha_2\alpha_1(-1)(1)(6) + (\alpha_2\alpha_2(-1)(-1)(2)\Big]$

$= \alpha_1 + \alpha_2 - 0.5[20\alpha_1^2 - 6\alpha_1\alpha_2 - 6\alpha_2\alpha_1 +$
$2\alpha_2^2]$

$= \alpha_1 + \alpha_2 - 0.5[20\alpha_1^2 - 12\alpha_1\alpha_2 + 2\alpha_2^2]$

$= \alpha_1 + \alpha_2 - 10\alpha_1^2 + 6\alpha_1\alpha_2 - \alpha_2^2$

h). $\alpha + \alpha - 10\alpha^2 + 6\alpha^2 - \alpha^2$
$= 2\alpha - 4\alpha^2 - \alpha^2$
$= 2\alpha - 5\alpha^2$

i). $\frac{d}{d\alpha} = 2 - 5(2)\alpha = 2 - 10\alpha$

$2 - 10\alpha = \emptyset$
$-10\alpha = -2$
$\alpha = \frac{2}{10} = \frac{1}{5} = 0.2$

$W_1 = 0.2 \cdot 1 \cdot (-2) + 0.2(-1)(-1)$
$= -0.4 + 0.2$
$= -0.2$

$W_2 = 0.2 \cdot 1 \cdot (4) + 0.2(-1)(1)$
$= 0.8 - 0.2$
$= 0.6$

d). the separating hyperplane should pass through the midpoint because it should create the shortest distance between the SV over the margin from the 2 classes.

e). $\frac{y_2-y_1}{x_2-x_1} = \frac{1-4}{-1+2} = \frac{-3}{1} = -3$ , $(-2,4)$ & $(-1,1)$

$4 = -2(3)+b$

$4 = -6+b$

$10 = b$

line $l = 3x+10$

midpoint $= \left(\frac{-2+(-1)}{2}, \frac{4+1}{2}\right) = \left(\overset{x_1}{-\frac{3}{2}}, \overset{y_1}{\frac{5}{2}}\right)$

Separating hyperplane =

$\frac{5}{2} = \frac{1}{3}\left(-\frac{3}{2}\right)+b$

$\frac{5}{2} = -\frac{1}{2}+b$

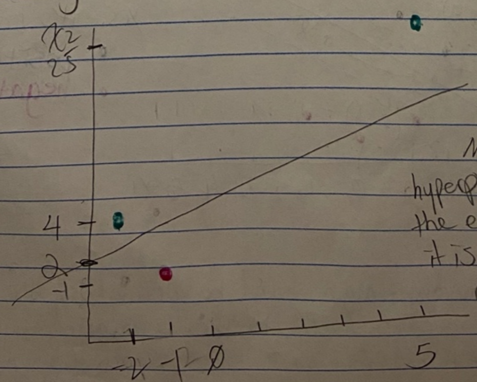
$\frac{5}{2}+\frac{1}{2} = b = \frac{6}{2} = 3$

$\Rightarrow y = \frac{1}{3}x+3$

$\Downarrow$

$\emptyset = \frac{1}{3}x_1 - x_2 + 3$

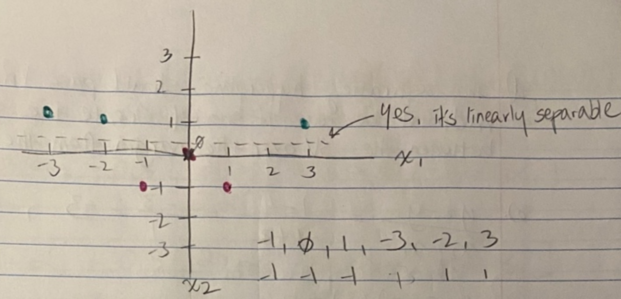
f). $(x,y) = (5,1) \in R^1 \rightarrow ((5,25),1) \in R^2$




No, it will not change the hyperplane. It's above the the edge of the margin and it is considered a solid class '1' point.

Q3.

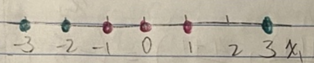
1).




-1, ∅, 1, -3, -2, 3
-1 -1 -1 1 1 1

Q3

a).



• positive $x_2$ (1)
• negative $x_2$ (-1)

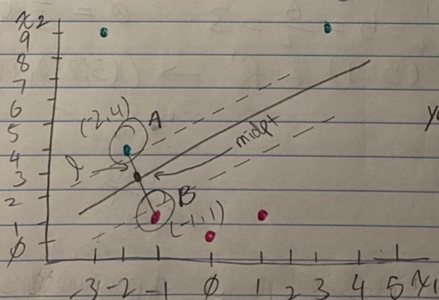
No, it's not linearly separable.

b). $\phi(x) = (x, x^2)$

(-1, -1), (∅, -1), (1, -1), (-3, 1), (-2, 1), (3, 1)

↓

= ((-1,1), -1), ((∅,∅), -1), ((1,1), -1), ((-3,9), 1), ((-2,4), 1), ((3,9), 1), $\in R^2$




yes, it's linearly separable
• positive class
• negative class

c). because maximal margin should be defined by the points which have the shortest distance to the separating hyperplane. And the point A and point B I choose, each from a different class, satisfy the shortest distance requirement, they are the SV over the margin.

Q1. a).

Not Play Tennis: 1 2 3 4 5 6 7
Play Tennis: 8 9 10 11 12 13 14

Entropy(S)
9, 5, |S| = 14
$-\frac{9}{14}\log_2(\frac{9}{14}) - \frac{5}{14}\log_2(\frac{5}{14}) = 0.94$

Outlook → Sunny: 1 2 8 9 11

Entropy(S|sunny)
3, 2, 5
$-\frac{3}{5}\log_2(\frac{3}{5}) - \frac{2}{5}\log_2(\frac{2}{5})$
$= 0.97$

Outlook → Overcast: 3 7 12 13

Entropy(S|Overcast) 0, 4, 4

Outlook → Rain: 4 5 6 10 14

Entropy(S|Rain)
2, 3, 5
$-\frac{2}{5}\log_2(\frac{2}{5}) - \frac{3}{5}\log_2(\frac{3}{5})$
$= 0.97$

Humidity: high → 1 2 8; normal → 9 11

wind: Strong → 6 14; weak → 4 5 10

Suny Temperature: hot → 1 2; mild → 8 11; Cool → 9

2, 0, 2    1 1 2    1 0 1

Entropy (sunny | hot) = $\phi$    (sunny | mild) = 1    (sunny | cool) = $\phi$

Entropy (Sunny | temperature) = 0.4

Gain (sunny, temperature) = 0.97 − 0.4 = 0.93

Sunny humidity: high → 1 2 8; normal → 9 11

3, 0, 3    2, 0, 2

(sunny | high) = $\phi$    (sunny | normal) = $\phi$

Entropy (Sunny | humidity) = $\phi$

Gain (sunny, humidity) = 0.97 − $\phi$ = 0.97

Sunny wind: Strong → 2 11; weak → 1 8 9

1 1 2    2 1 3

(Sunny | Strong) = 1

(sunny | weak) $= -\frac{2}{3}\log_2(\frac{2}{3}) - \frac{1}{3}\log_2(\frac{1}{3}) = 0.92$

Entropy (Sunny | wind) $= \frac{2}{5} + 0.92(\frac{3}{5}) = 0.952$

Gain (sunny, wind) = 0.97 − 0.952 = −0.018

Rain − temperature: hot → ; mild → 4 10 14; cool → 5 6

1 2 3    1 1 2

(rain | hot) $\phi$    (rain | mild) 0.92    (rain | cool) 1

Entropy (Rain | temperature) $= \phi + \frac{3}{5}(0.92) + (\frac{2}{5}) = 0.952$

G(rain, temperature) = 0.97 − 0.952 = 0.018

Rain − humidity: high → 4 14; normal → 10 5 6

1 1 2    1 2 3

(Rain | high) = 1

(rain | normal) $= -\frac{2}{3}\log_2(\frac{2}{3}) - \frac{1}{3}\log_2(\frac{1}{3}) = 0.92$

Entropy (rain | humidity) $= \frac{2}{5} + 0.92(\frac{3}{5}) = 0.952$

G(rain, humidity) = 0.97 − 0.952 = 0.018

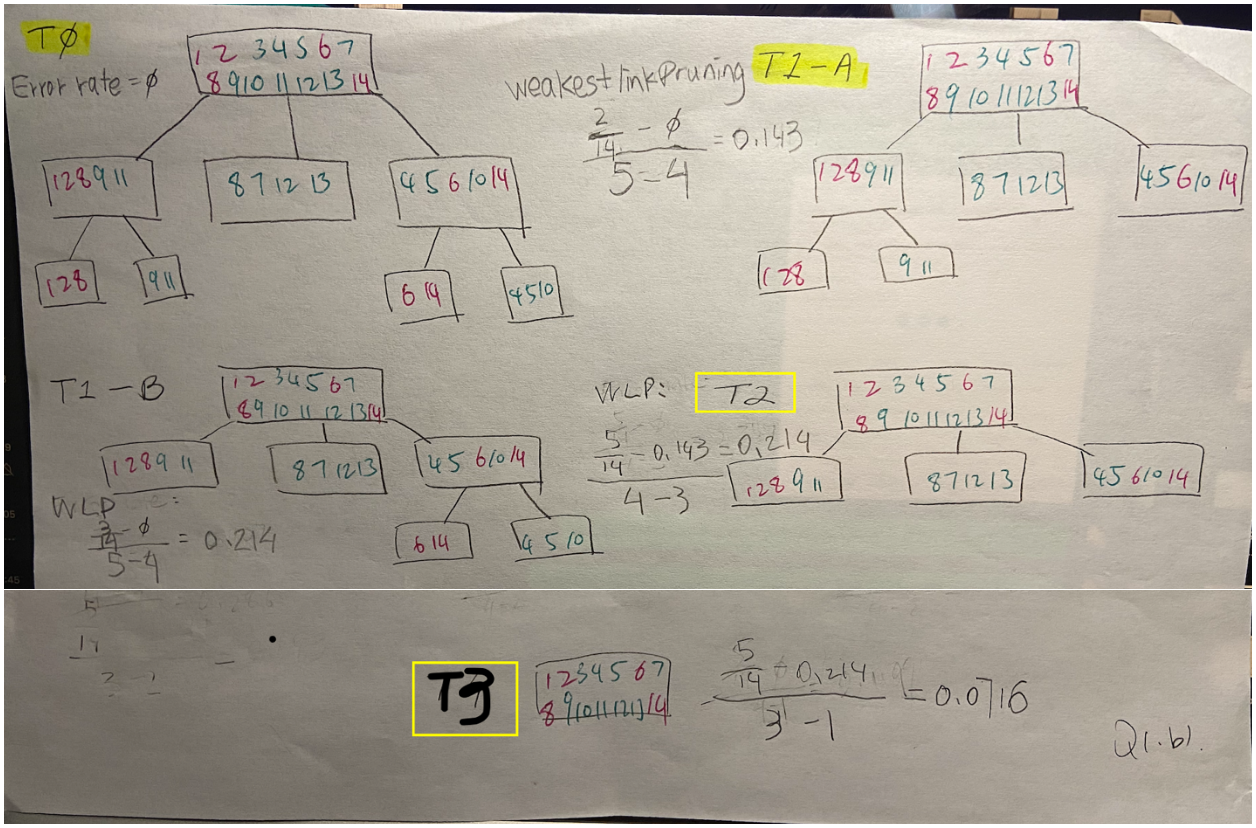
Rain − wind: strong → 6 14; weak → 10 4 5

3 0 3    0 3 3

(rain | Strong) = $\phi$    (rain | weak) = $\phi$

Entropy (rain | wind) = $\phi$

Gain (rain, wind) = 0.97 − $\phi$ = 0.97




The yellow squared tree is my selected tree.
The red color number is misclassifications and the green color is correct classification.