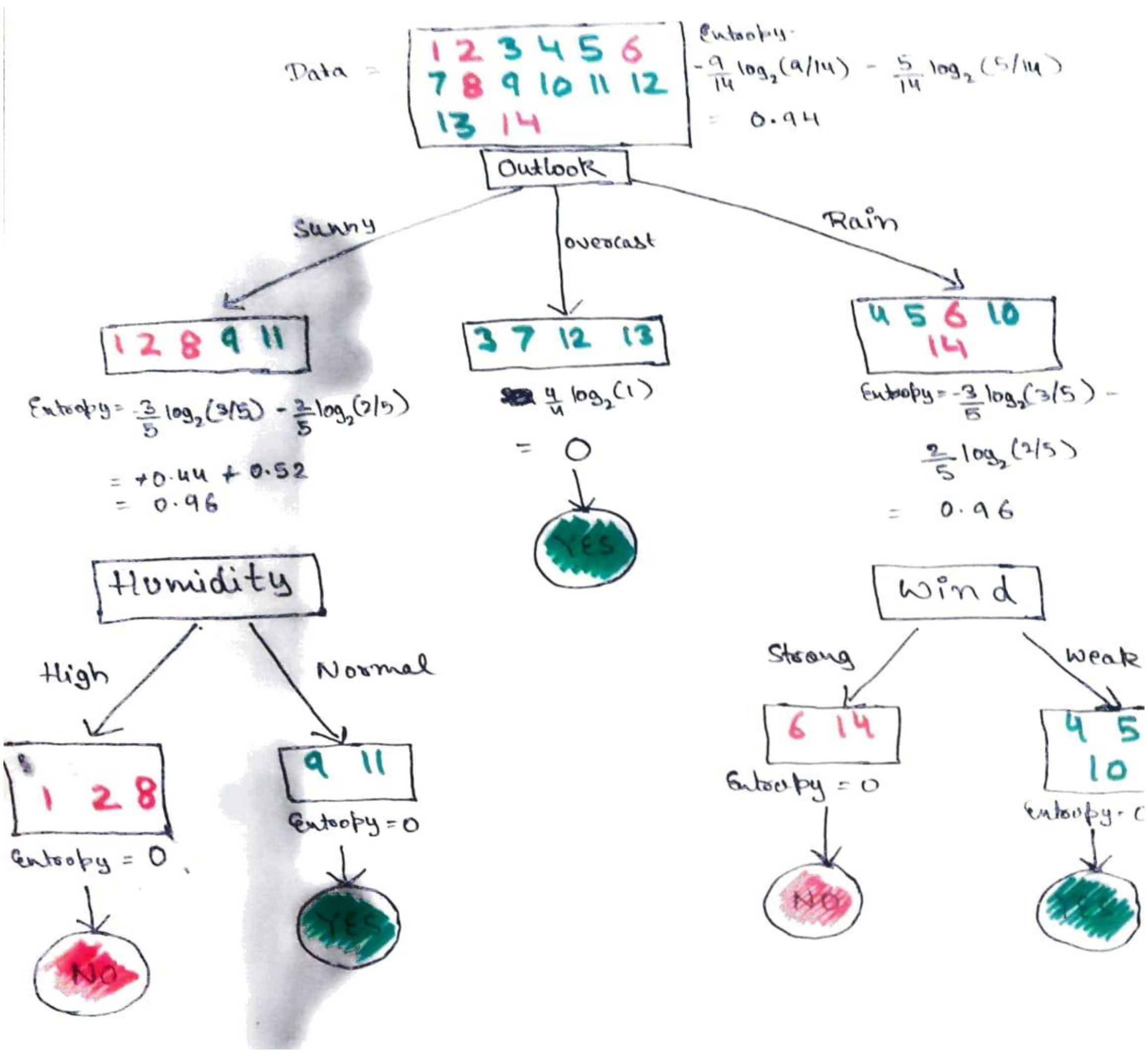
Introduction to Machine Learning
Problem Set: Decision Trees, Ensembles, Support Vector Classifier

[redacted]

Q1. Decision tree for Tennis dataset:

a) Build Decision tree classifier

Complete Decision tree:



Information Gain:
Entropy (S) = 0.94

Entropy(S, Outlook) = 0.246 [Selected]
Entropy(S, Wind) = 0.151
Entropy(S, Temperature) = 0.048
Entropy(S, Humidity) = 0.026

→ Split on Outlook

Gain($S_{Sunny,}$ Humidity) =
E(S) - E(High humidity) - E(Normal Humidity)
0.94 - ⅗(0) - ⅖(0) = 0.94 [Selected]

Gain(S_{Sunny}, Wind) =
E(S) - E(Weak wind) - E(Strong wind)
0.94 - ⅖(1) - ⅗(0.918) = 0.0108

Gain(S_{sunny}, Temperature) =
E(S) - E(Hot) - E(Mild) - E(Cold)
0.94 - ⅖(0) - ⅖(1) - ⅕(0) = 0.540

→ Split on Outlook_Sunny, Humidity

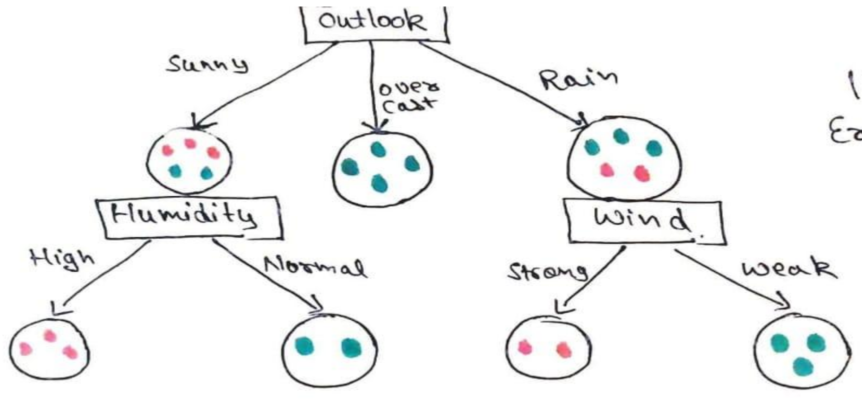
Gain(S_{Rain}, Wind) =
E(S) - E(Weak) - E(Strong)
0.94 - ⅗(0) - ⅖(0) = 0.94 [Selected]

Gain(S_{Rain}, Temperature) =
E(S) - E(Hot) - E(Mild) - E(Cold)
0.94 - 0 - ⅖(1) - ⅗(0.918) = -0.01

→ Split on Outlook_Rain, Wind)

b) Weakest link pruning

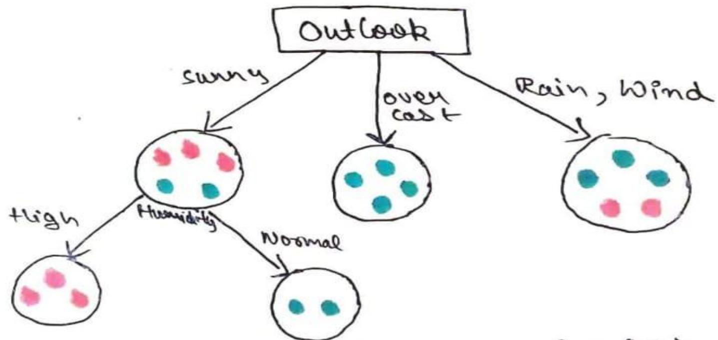
T_0



$|T_0| = 5$

$Err|T_0| = 0$

$T_1(A)$



$Err(T_1) = 2/14 \qquad |T_1| = 4$

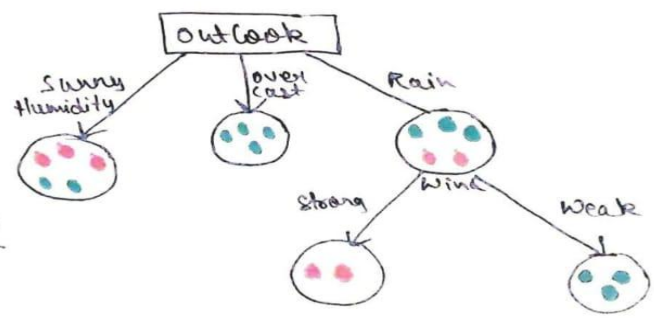
$$\frac{Err(T_1) - Err(T_0)}{|T_0| - |T_1|} = \frac{\frac{2}{14} - 0}{5 - 4} = 0.142$$

$T_1(B)$

$Err(T_1) = 3/14$

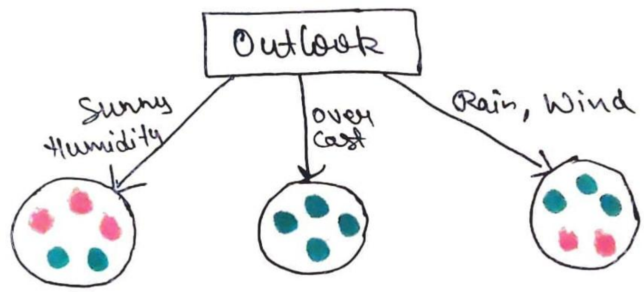
$|T_1| = 4$

$$\frac{Err(T_1) - Err(T_0)}{|T_0| - |T_1|} = \frac{3/14 - 0}{5 - 4} = 0.214$$



We select $T_1(A)$

$T_2 \rightarrow$



$Err(T_2) = \frac{5}{14}$ $\quad |T_2| = 3$

$$\frac{Err(T_2) - Err(T_1 A)}{|T_1 A| - |T_2|} = \frac{\frac{5}{14} - \frac{2}{14}}{4-3} = \frac{3}{14} = 0.214$$

$T_3 \rightarrow$

$Err(T_3) = \frac{5}{14}$ $\quad |T_3| = 1$

$$\frac{Err(T_3) - Err(T_2)}{|T_2| - |T_1|} = \frac{\frac{5}{14} - \frac{5}{14}}{3-1} =$$

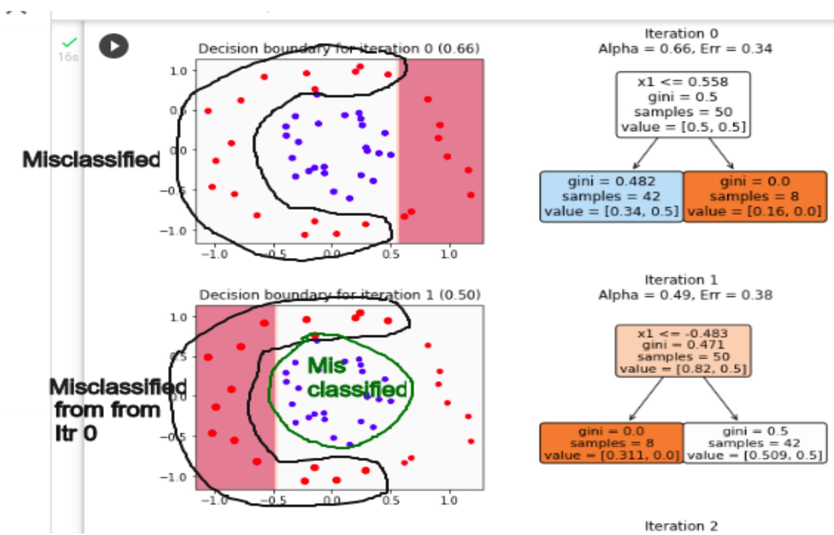
$= 0$

Q2. From demo: AdaBoost Classifier notebook:

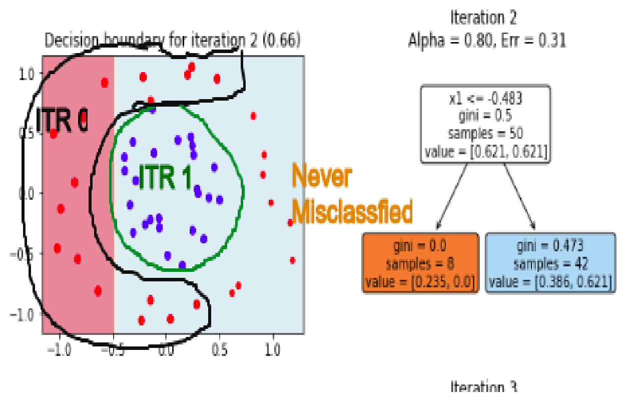
a) The ensemble is more confident about those eight samples after two iterations because these are the samples where both the trees agree i.e these samples are red in both iterations.

In Iteration 0, the left side of the split is classified as blue because of a higher value equal to 0.5, and the right side of the split is predicted red. The left side red samples are misclassified therefore in the next iteration we increase their weight by a defined factor.

In the Iteration 1, we classify all samples as red because red samples on the left have more weight and have a higher value in both classes i.e 0.311, 0.509. This results in misclassification of the blue samples in the middle. Therefore, we are only confident about those 8 samples, other samples were misclassified at least once in both the iterations therefore we can't draw any concrete conclusions out of them.

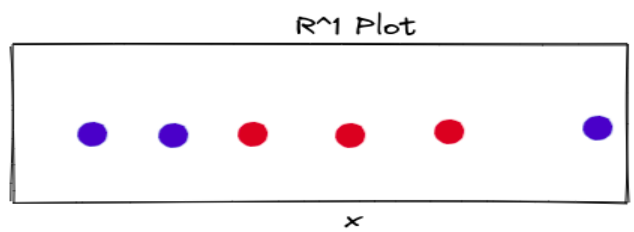


b) What changed between Iteration 1 and 2 is the weight of the samples. In Iteration 2 we have 3 weights: Black region because these samples were misclassified in Iteration 0, Green region because these samples were misclassified in Iteration 1 and the other 8 points were never misclassified therefore they have smaller weight and high confidence. And because of this i.e. Weight difference we get a higher value for the blue class which is 0.62 (The weight gets added to the value) and 0.38 for the red class, therefore we predict the right side of the split as blue in Iteration 2.



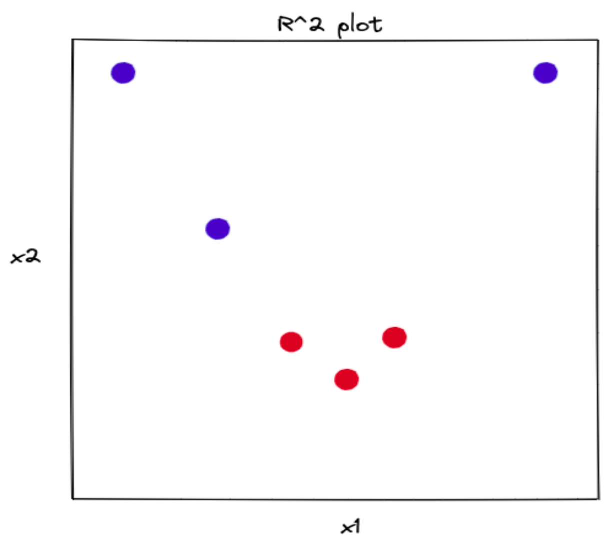
Q3.

a) $R^1 = (-1, -1), (0, -1), (1, -1), (-3, +1), (-2, +1), (3, +1)$



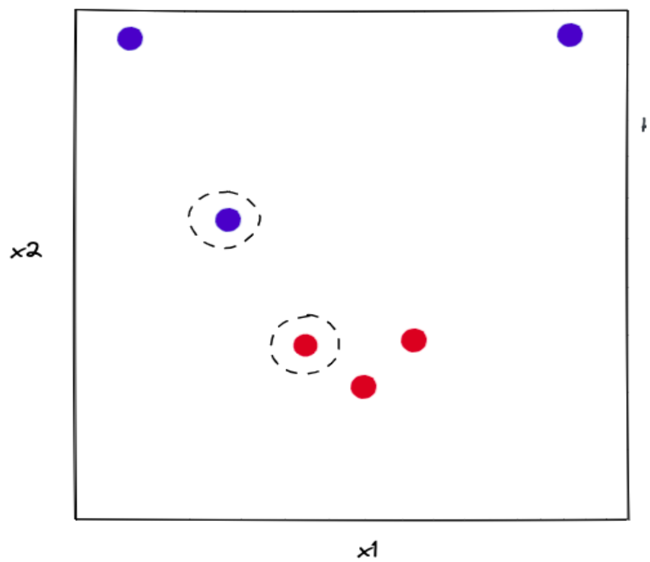
Data is not linearly separable.

b) R^2 = ((-1,1),-1) , ((0,0),-1), ((1,1),-1), ((-3,9), +1), ((-2,4), +1), ((3,9), +1)

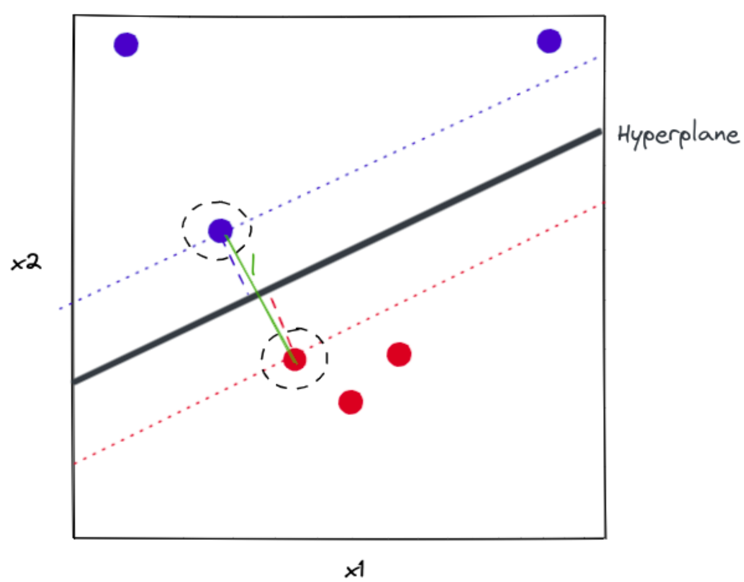


Data is linearly separable.

c) As per my geometric intuition the hyperplane should pass between the points ((-1,1), -1) and ((-2,4), +1) so as to separate both the classes correctly. These are the only points which have max margin between then and no other sample lies between them.
On calculating their perpendicular distance from the hyperplane i.e 0.5, these points are the closest to the hyperplane and can fall on the dotted lines when we draw the Maximum margin classifier. Therefore, I selected these points as support vectors.



d) Maximum Margin Classifier : The midpoint between the two support vectors is at (-1.5, 2.5) on the line l. The separating hyperplane should pass through this point in order to have maximum margin, which is 1 here. This will also help to maintain equal max margin for each class which is 0.5.
Length of line l i.e Max margin = 1



e) The equation is $w_0 + w_1x_1 + w_2x_2 < 0$ if y = -1 then the point lies below the hyperplane i.e. red point. And the equation is $w_0 + w_1x_1 + w_2x_2 > 0$ if y = +1 then the point lies above the hyperplane i.e blue point.

→ From the points we know, our hyperplane has a slope of ⅓ and an intercept of 3. And the midpoint between the two support vectors is at (-1.5, 2.5).Therefore, the hyperplane equation is $W_o - 1.5W_1 + 2.5W_2 = 0$

f) No, the hyperplane won't change. Point x=5 y=+1 when written as per R^2 will be ((5,25),+1) which will lie in the correct region i.e above the hyperplane i.e blue. The margin will also remain the same i.e 0.5.

g) Simplifying for α1 and α2: We have x1 is x+ so x1= [-2,4] x2 is x_ so x2= [-1,1] and y+ = 1, y_ = -1

Given equation.

$\alpha_i, i = 1,2$

$$\sum_{i=1}^{2} \alpha_i - \frac{1}{2} \sum_{i=1}^{2} \left(\sum_{j=1}^{2} \alpha_i \alpha_j y_i y_j \phi(x_i)^T \phi(x_j) \right)$$

$$\alpha_1 + \alpha_2 - \frac{1}{2} \sum_{i=1}^{2} \left(\alpha_i \alpha_1 y_i y_1 \phi(x_i)^T \phi(x_1) + \alpha_i \alpha_2 y_i y_2 \phi(x_i)^T \phi(x_2) \right)$$

$$\alpha_1 + \alpha_2 - \frac{1}{2} \left(\alpha_1 \alpha_1 y_1 y_1 \phi(x_1)^T \phi(x_1) + \alpha_1 \alpha_2 y_1 y_2 \phi(x_1)^T \phi(x_2) + \alpha_2 \alpha_1 y_2 y_1 \phi(x_2)^T \phi(x_1) + \alpha_2 \alpha_2 y_2 y_2 \phi(x_2)^T \phi(x_2) \right)$$

Substituting values

y_1 is y_- i.e -1

y_2 is y_+ i.e +1

$$\phi(x_1)^T \phi(x_1) = [-2,4] \cdot [-2,4]$$
$$= 20$$

$$\phi(x_1)^T \phi(x_2) = [-2,4] \cdot [-1,1]$$
$$= 6$$

$$\phi(x_2)^T \phi(x_2) = [-1,1] \cdot [1,1]$$
$$= 2$$

Putting these values in equation

$$\alpha_1 + \alpha_2 - \frac{1}{2}\left(\alpha_1^2 (1)(20) + \alpha_1\alpha_2(-1)(6) \right.$$
$$\left. \alpha_2\alpha_1(-1)(6) + \alpha_2^2 (1)(2) \right)$$

$$\alpha_1 + \alpha_2 - \frac{1}{2}\left(20\alpha_1^2 + 2\alpha_2^2 - 12\alpha_1\alpha_2 \right)$$

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

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

h) Simplifying for $\alpha 1 = \alpha 2 = \alpha$

h) From g) we have,

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

A.T.Q $\alpha_1 = \alpha_2 = \alpha$

Substituting values

$$-10\alpha^2 - \alpha^2 + 6\alpha^2 + \alpha + \alpha$$

$$-5\alpha^2 + 2\alpha$$

$$\alpha(-5\alpha + 2)$$

i) Derivative of the function w.r.t α which gives
$\alpha = 1/5$, $W_1 = -1/5$, $W_2 = 3/5$

i) From h) we have

$$-5\alpha^2 + 2\alpha$$

$$\frac{d}{d\cancel{x}} (-5\alpha^2 + 2\alpha)$$

$$= -10\alpha + 2$$

A.T.Q, $-10\alpha + 2 = 0$

$$-10\alpha = -2$$

$$\boxed{\alpha = \frac{1}{5}}$$

$$w_j = \sum_{i \in S} \alpha_i y_i x_{ij} \quad , j = 1, 2$$

$$w_1 = \sum_{i=1}^{2} \alpha_i y_i x_{i1}$$

$$= \alpha_1 y_1 x_{11} + \alpha_2 y_2 x_{21}$$

$$= \frac{1}{5}(1)(-2) + \frac{1}{5}(-1)(-1)$$

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

$$\boxed{w_1 = -\frac{1}{5}}$$

$$w_2 = \sum_{i=1}^{2} \alpha_i y_i x_{i2}$$

$$= \alpha_1 y_1 x_{12} + \alpha_2 y_2 x_{22}$$

$$= \frac{1}{5}(1)(4) + \frac{1}{5}(-1)(1)$$

$$= \frac{4}{5} + -\frac{1}{5} = \frac{3}{5}.$$

$$\boxed{w_2 = \frac{3}{5}}$$

j) Calculating W_o

Ans: $W_o = -9/5$

j) $$y_i = w_0 + \sum_{j=1}^{p} w_j x_{ij}$$

$$y_1 = w_0 + \sum_{j=1}^{p} w_j x_{1j}$$

$$y_1 = w_0 + w_1 x_{11} + w_2 x_{12}$$

$$1 = w_0 + \frac{-1}{5}(-2) + \frac{3}{5}(4)$$

$$1 = w_0 + \frac{2}{5} + \frac{12}{5}$$

$$1 = w_0 + \frac{14}{5}$$

$$1 - \frac{14}{5} = w_0$$

$$\therefore \boxed{w_0 = -\frac{9}{5}}$$

k) Verifying the hyperplane equation calculated in (e) by putting up values.
From (e) we know, (-1.5, 2.5) is the midpoint between the two support vectors and it passes through the hyperplane so that it can have maximum margin. Therefore,

k) So, we have

$$W_0 = \frac{-9}{5} \qquad W_1 = \frac{-1}{5} \qquad W_2 = \frac{3}{5}$$

Putting in equation

$$W_0 + W_1 x_1 + W_2 x_2 = 0$$

$$\frac{-9}{5} - \frac{1}{5} x_1 + \frac{3}{5} x_2 = 0$$

Point on the hyperplane is $(-1.5, 2.5)$

$$\frac{-9}{5} - \frac{1}{5}(-1.5) + \frac{3}{5}(2.5)$$

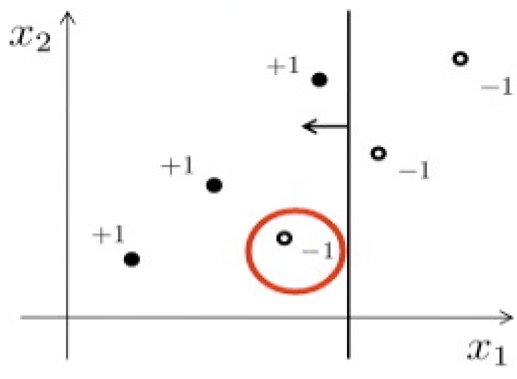
$$-\frac{9}{5} + \frac{1.5}{5} + \frac{7.5}{5}$$

$$\Rightarrow \quad -\frac{9}{5} + \frac{9}{5} = 0$$

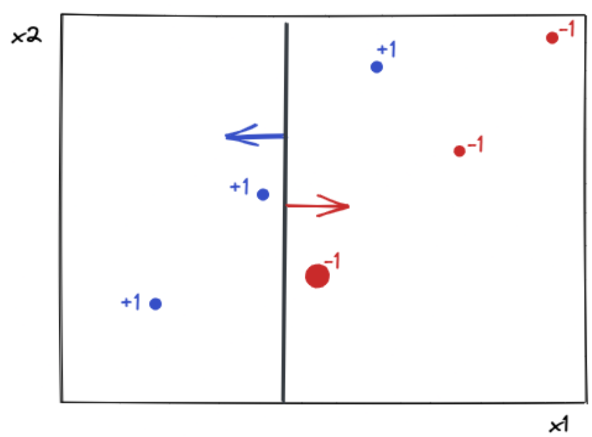
$\therefore$ The hyperplane predicted in (e) is correct

Q4.

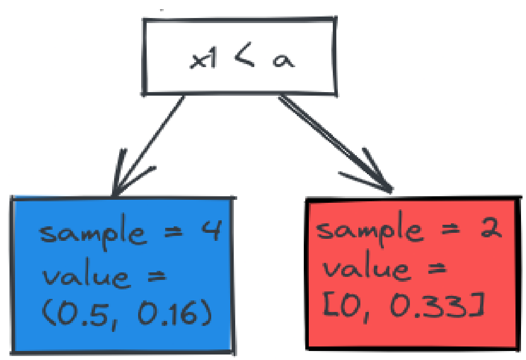
a) Point -1 in the positive region will have higher weight in the next boosting iteration because this point is classified wrong.



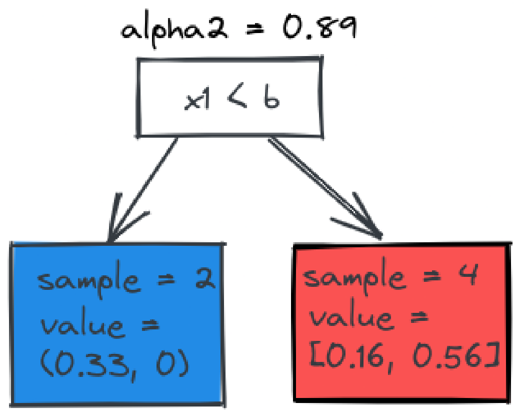
b) Iteration 1:



c) For Iteration 0:



For Iteration 0, $Err\_m_o = 1/6 = 0.166$
$Alpha\_1 = \log((1 - err_m)/ err_m = 0.69$



We know α depends on the weight of any misclassified sample and the weight is increased by a factor $e^{(\alpha_m)}$. In iteration 1, we have 1 misclassified sample from Iteration 0, therefore when we calculate corresponding values, the weight also gets added on.

For Iteration 1, $Err\_m_1 = 1/6 = 0.0166$ but these get added weight from the misclassified weights, which gets added up to the alpha. Therefore, we get $\alpha 2 > \alpha 1$.

For Red class in Iteration 1 we have:
Positive value = 1/6 = 0.16
Negative Value = weight * 3/6 = 0.56

For Blue class in Iteration 1 we have:
Positive class = 2/6 = 0.33
Negative class = 0

Alpha2 = 0.89 alpha1 = 0.69