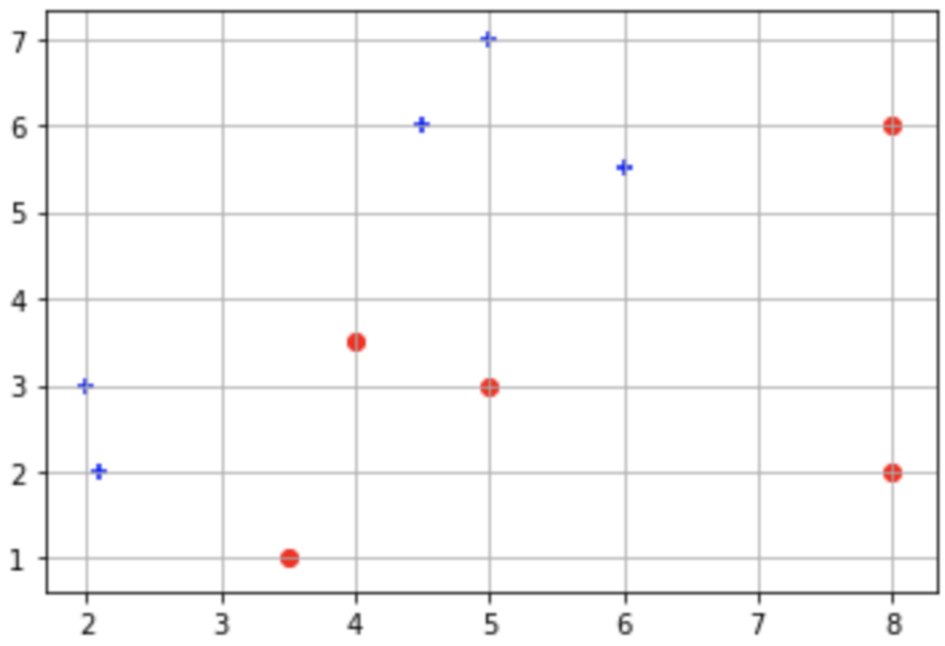
EXERCICE 1 Decision Tree

Consider a training dataset with two classes:

C_1: $\left\{x^{(1)} = \begin{pmatrix}2\\3\end{pmatrix}, x^{(2)} = \begin{pmatrix}2.1\\2\end{pmatrix}, x^{(3)} = \begin{pmatrix}4.5\\6\end{pmatrix}, x^{(6)} = \begin{pmatrix}5\\7\end{pmatrix}, x^{(8)} = \begin{pmatrix}6\\5.5\end{pmatrix}\right\}$ with label 1

C_2: $\left\{x^{(4)} = \begin{pmatrix}4\\3.5\end{pmatrix}, x^{(5)} = \begin{pmatrix}3.5\\1\end{pmatrix}, x^{(7)} = \begin{pmatrix}5\\3\end{pmatrix}, x^{(9)} = \begin{pmatrix}8\\6\end{pmatrix}, x^{(10)} = \begin{pmatrix}8\\2\end{pmatrix}\right\}$ with label -1

(a) Plot the points. Is it feasible to find a linear "decision border" to classify both classes?



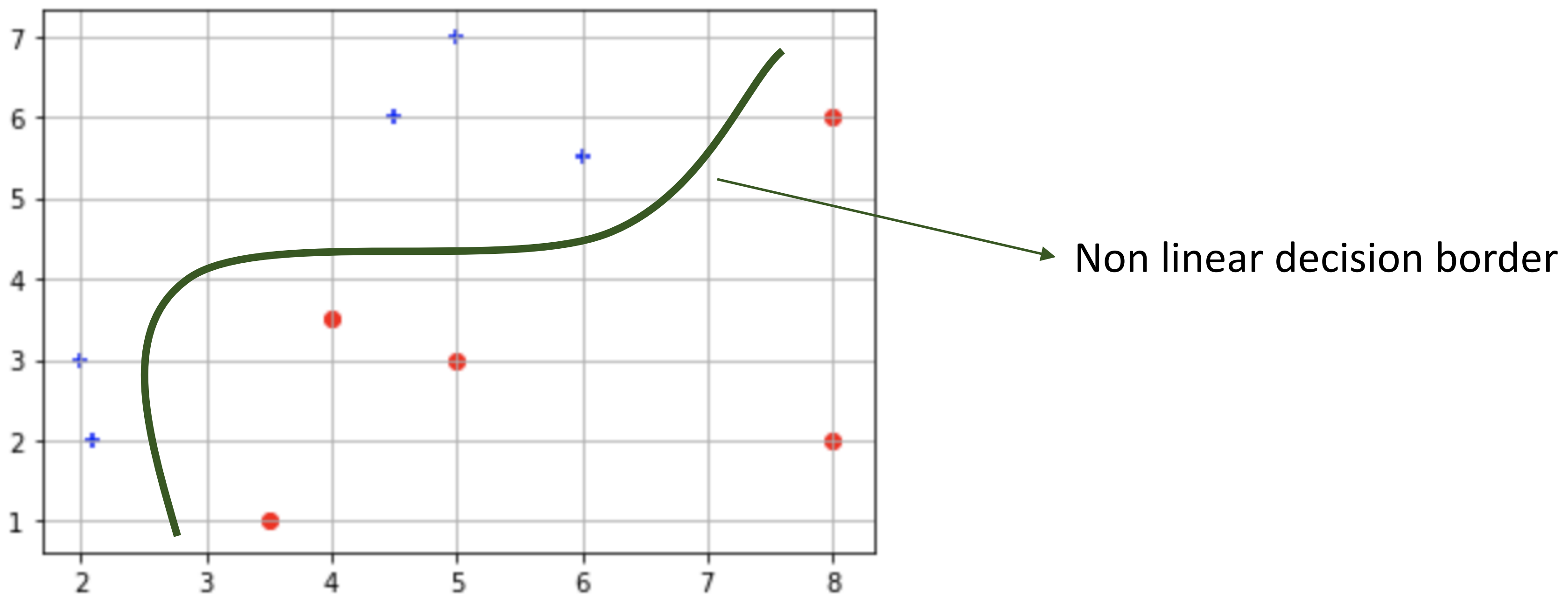
EXERCICE 1 Decision Tree

Consider a training dataset with two classes:

$$C_1: \left\{ x^{(1)} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, x^{(2)} = \begin{pmatrix} 2.1 \\ 2 \end{pmatrix}, x^{(3)} = \begin{pmatrix} 4.5 \\ 6 \end{pmatrix}, x^{(6)} = \begin{pmatrix} 5 \\ 7 \end{pmatrix}, x^{(8)} = \begin{pmatrix} 6 \\ 5.5 \end{pmatrix} \right\} \text{ with label } 1$$

$$C_2: \left\{ x^{(4)} = \begin{pmatrix} 4 \\ 3.5 \end{pmatrix}, x^{(5)} = \begin{pmatrix} 3.5 \\ 1 \end{pmatrix}, x^{(7)} = \begin{pmatrix} 5 \\ 3 \end{pmatrix}, x^{(9)} = \begin{pmatrix} 8 \\ 6 \end{pmatrix}, x^{(10)} = \begin{pmatrix} 8 \\ 2 \end{pmatrix} \right\} \text{ with label } -1$$

(a) Plot the points. Is it feasible to find a linear "decision border" to classify both classes?



EXERCICE 1 Decision Tree

$C_1: \left\{x^{(1)} = \begin{pmatrix}2\\3\end{pmatrix}, x^{(2)} = \begin{pmatrix}2.1\\2\end{pmatrix}, x^{(3)} = \begin{pmatrix}4.5\\6\end{pmatrix}, x^{(6)} = \begin{pmatrix}5\\7\end{pmatrix}, x^{(8)} = \begin{pmatrix}6\\5.5\end{pmatrix}\right\}$ with label 1

$C_2: \left\{x^{(4)} = \begin{pmatrix}4\\3.5\end{pmatrix}, x^{(5)} = \begin{pmatrix}3.5\\1\end{pmatrix}, x^{(7)} = \begin{pmatrix}5\\3\end{pmatrix}, x^{(9)} = \begin{pmatrix}8\\6\end{pmatrix}, x^{(10)} = \begin{pmatrix}8\\2\end{pmatrix}\right\}$ with label -1

(b) Draw the node splits decision (or decision stumps) a decision tree in the scatter plot with the following hyperparameters:

- Minimum samples for a node split = 2

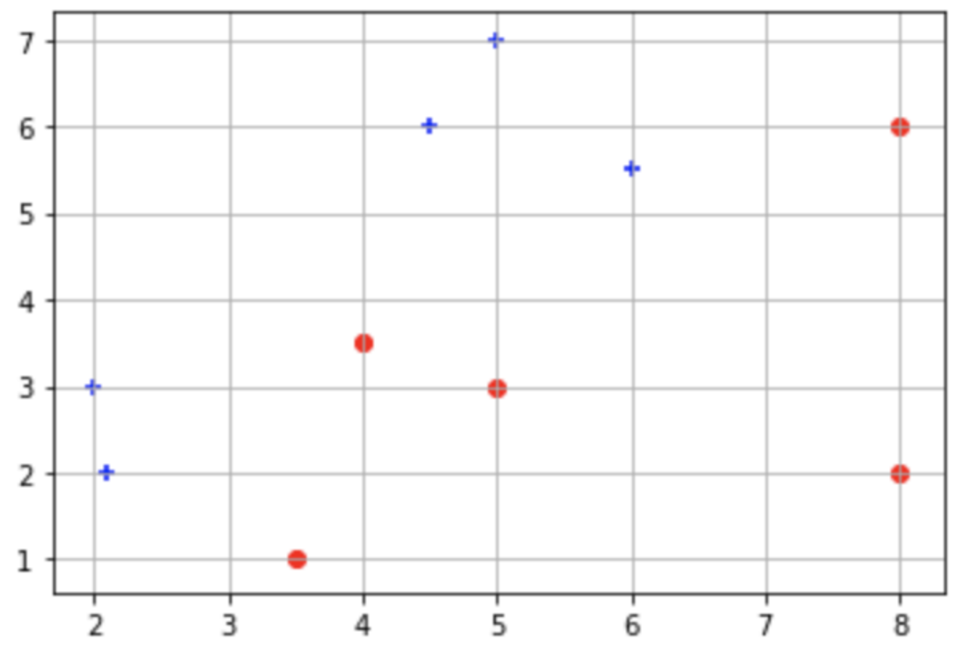
  At least 2 points for a node to be split.

- Minimum samples for a terminal node or leaf = 2

  At least 2 points in a leaf.

- Maximum depth of tree (vertical depth) = 3

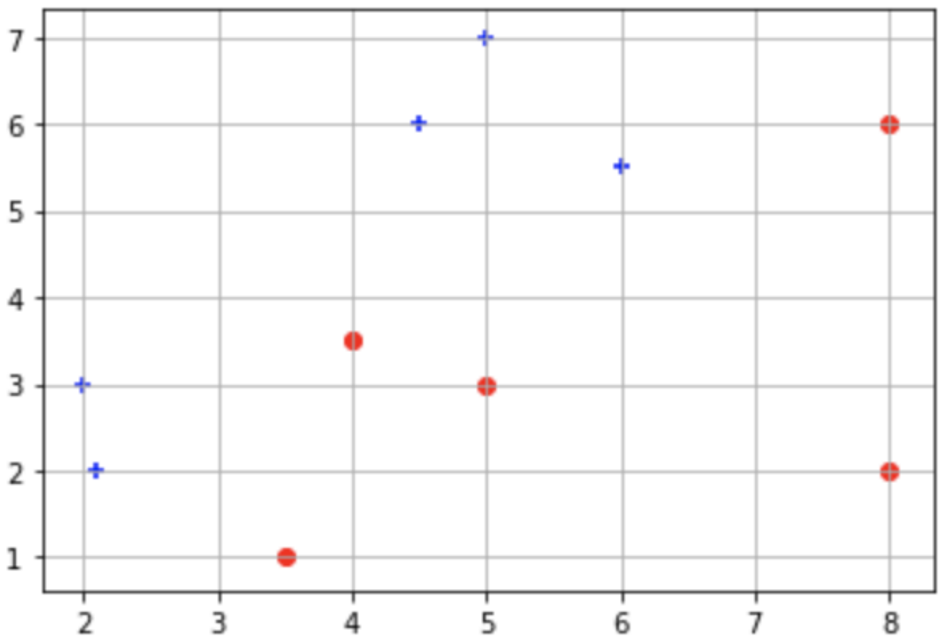
  At most 3 "rounds" of splits.

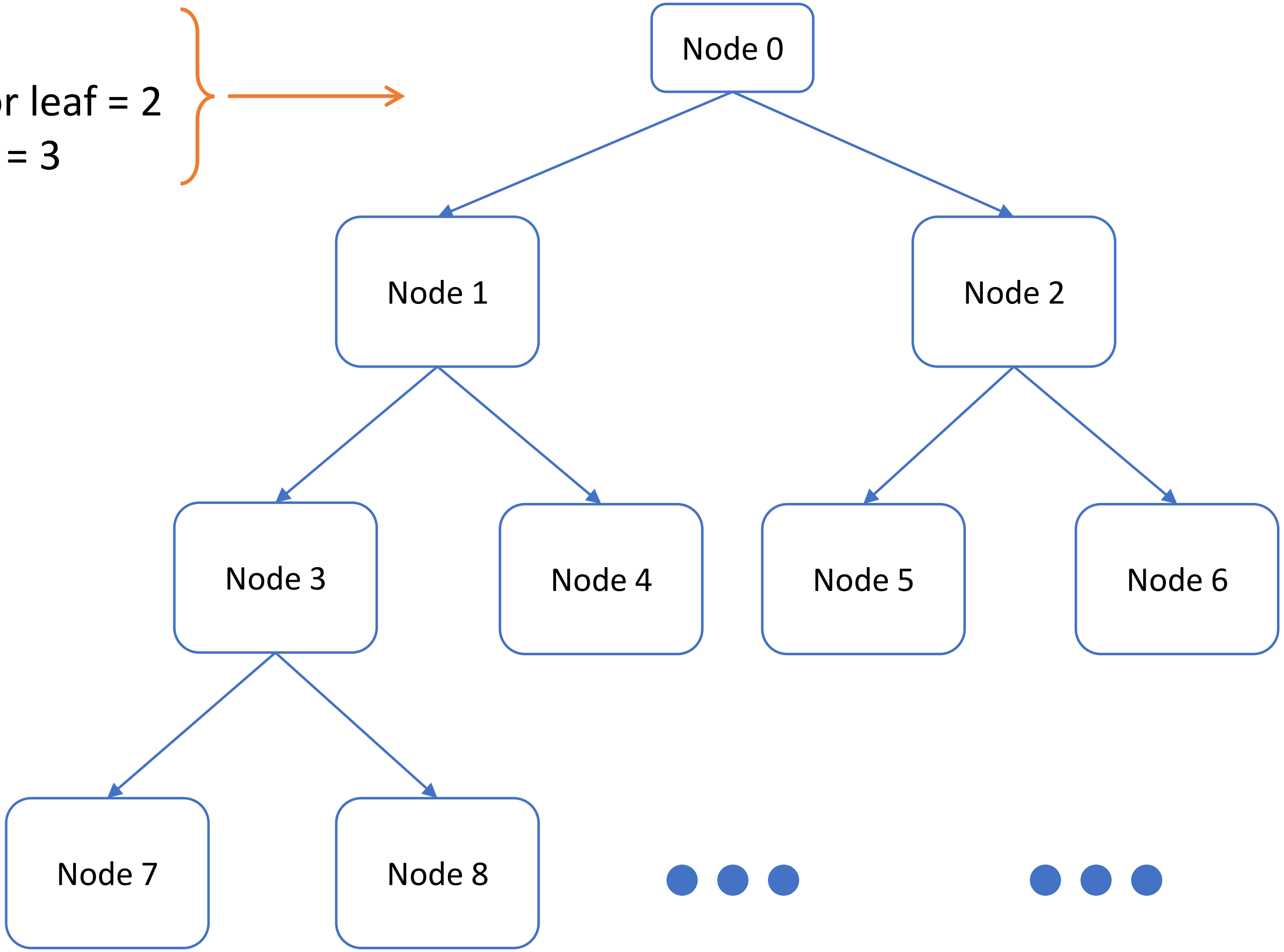


EXERCICE 1 Decision Tree

(b) Draw the node splits decision (or decision stumps) a decision tree in the scatter plot with the following hyperparameters:

- Minimum samples for a node split = 2
- Minimum samples for a terminal node or leaf = 2
- Maximum depth of tree (vertical depth) = 3

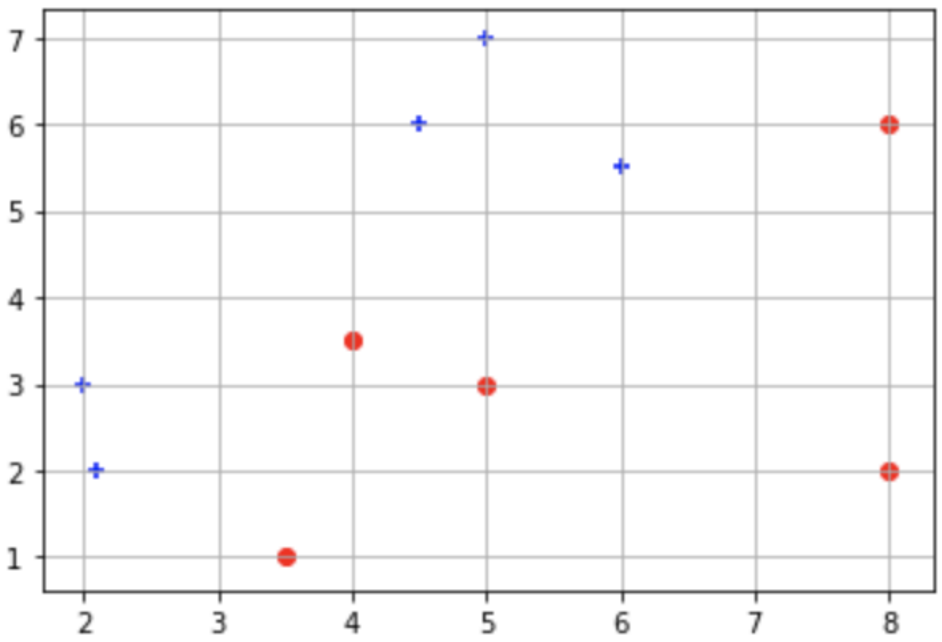


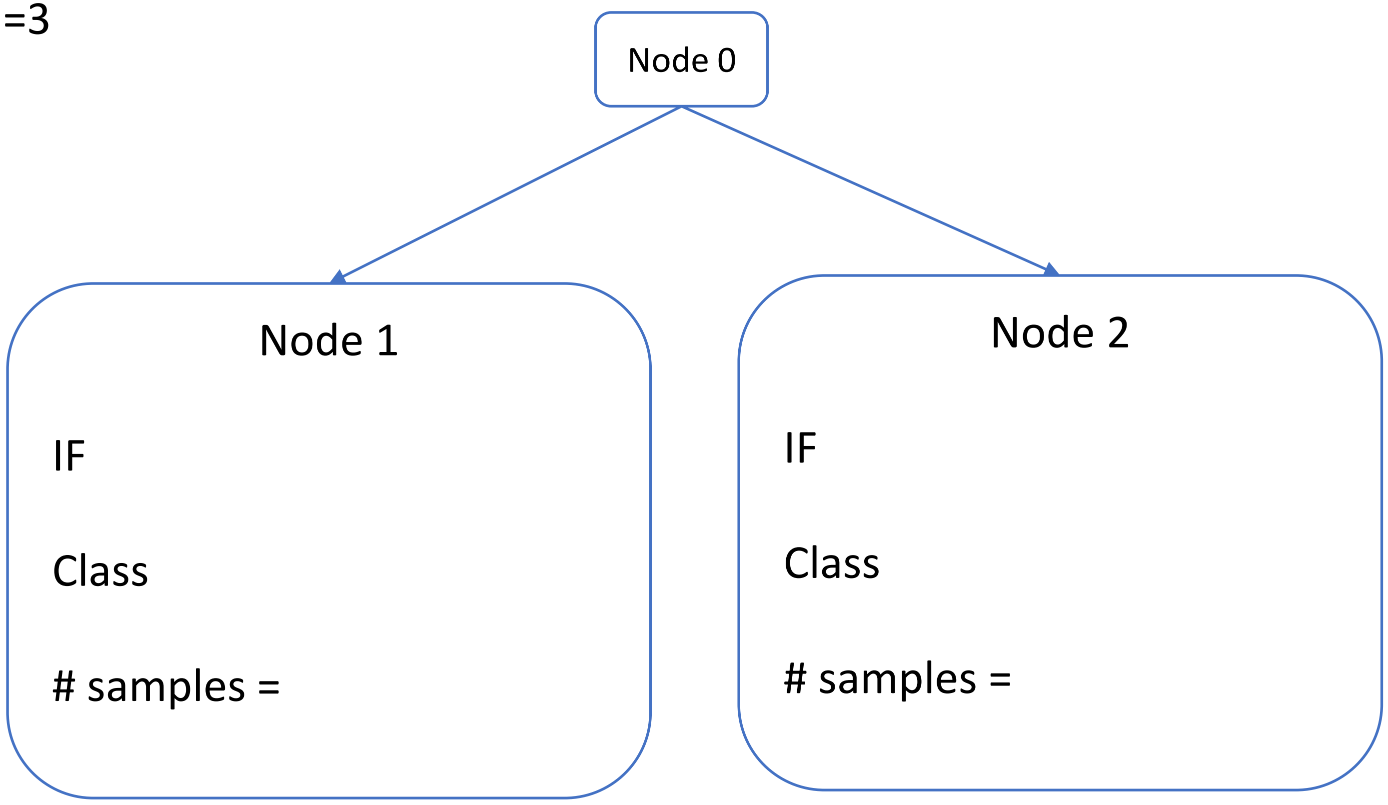


EXERCICE 1

(b) Draw the node splits decision (or decision stumps) a decision tree in the scatter plot with the following hyperparameters:

- Minimum samples for a node split= 2
- Minimum samples for a terminal node or leaf=2
- Maximum depth of tree (vertical depth)=3

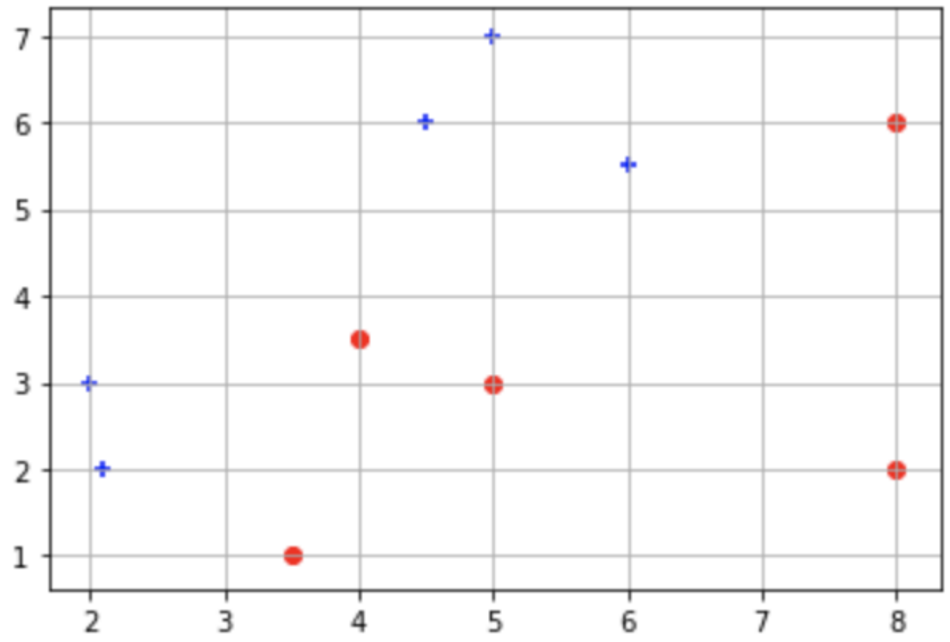


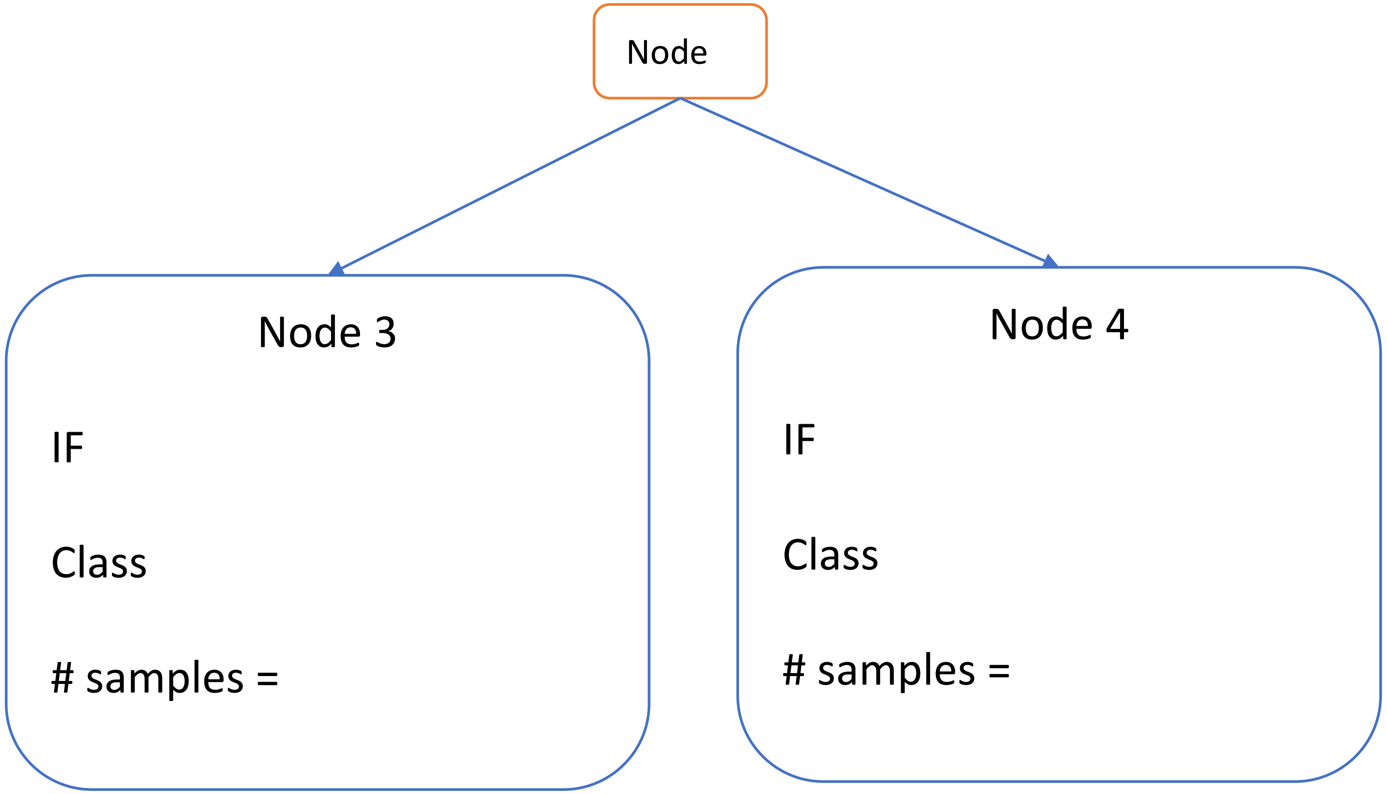


EXERCICE 1

(b) Draw the node splits decision (or decision stumps) a decision tree in the scatter plot with the following hyperparameters:

- Minimum samples for a node split= 2
- Minimum samples for a terminal node or leaf=2
- Maximum depth of tree (vertical depth)=3

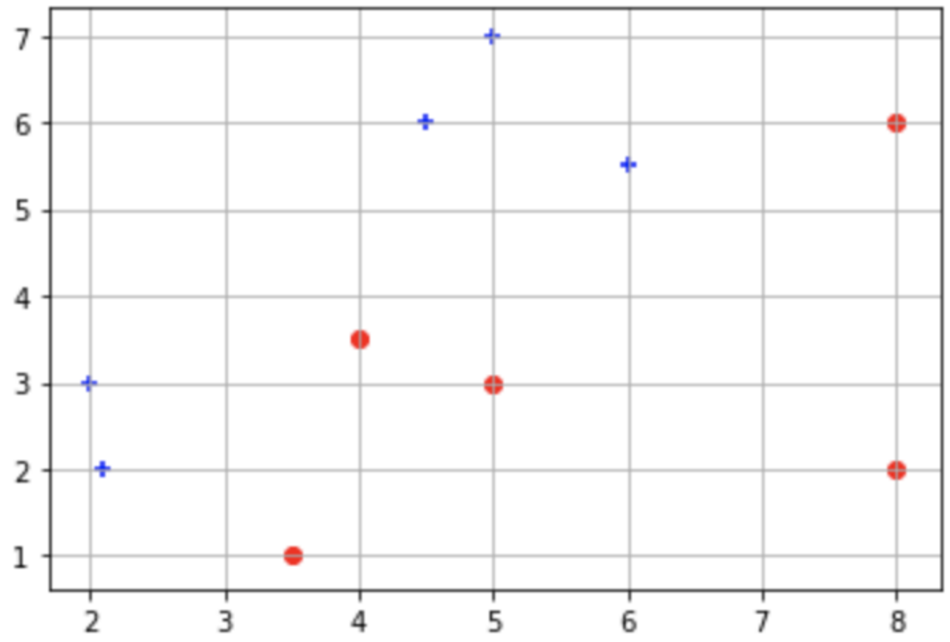


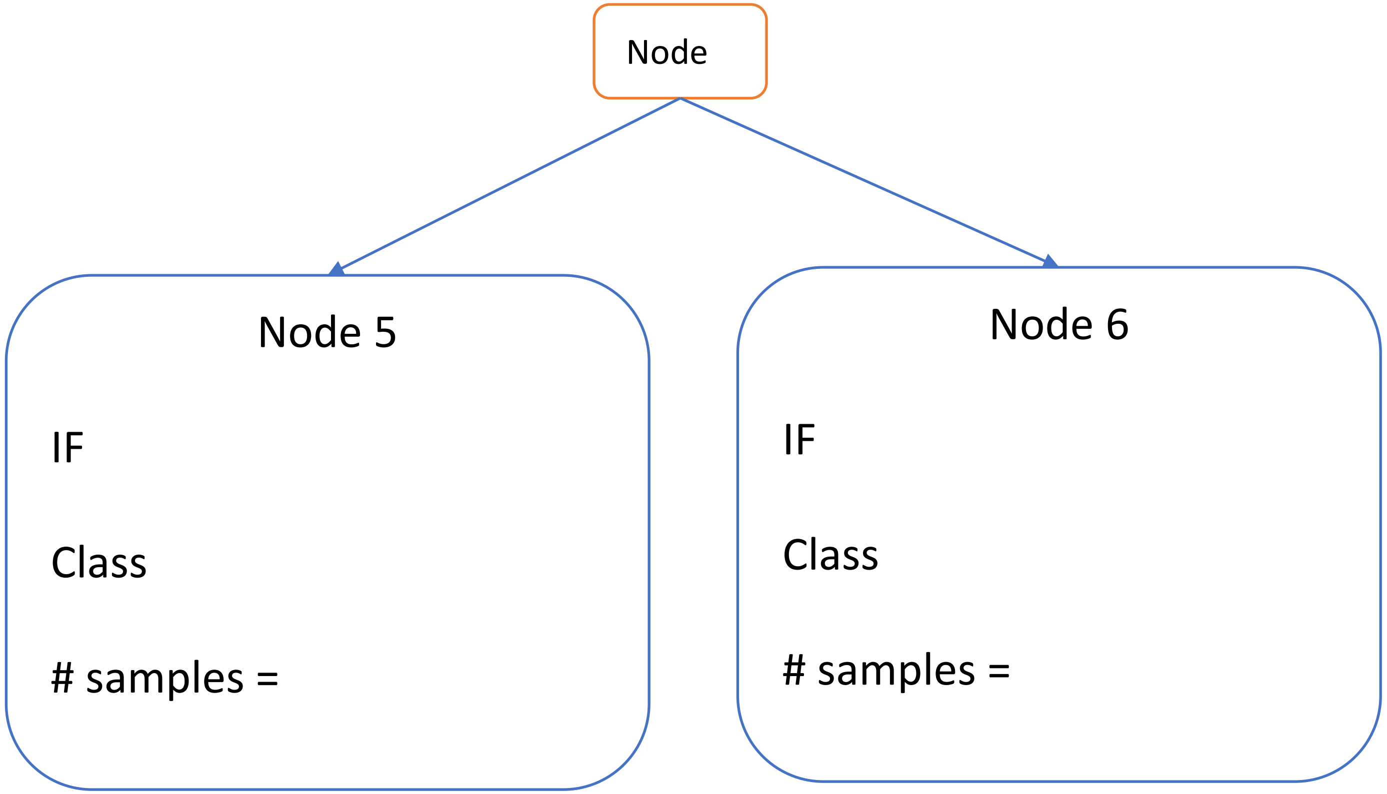


EXERCICE 1

(b) Draw the node splits decision (or decision stumps) a decision tree in the scatter plot with the following hyperparameters:

- Minimum samples for a node split= 2
- Minimum samples for a terminal node or leaf=2
- Maximum depth of tree (vertical depth)=3





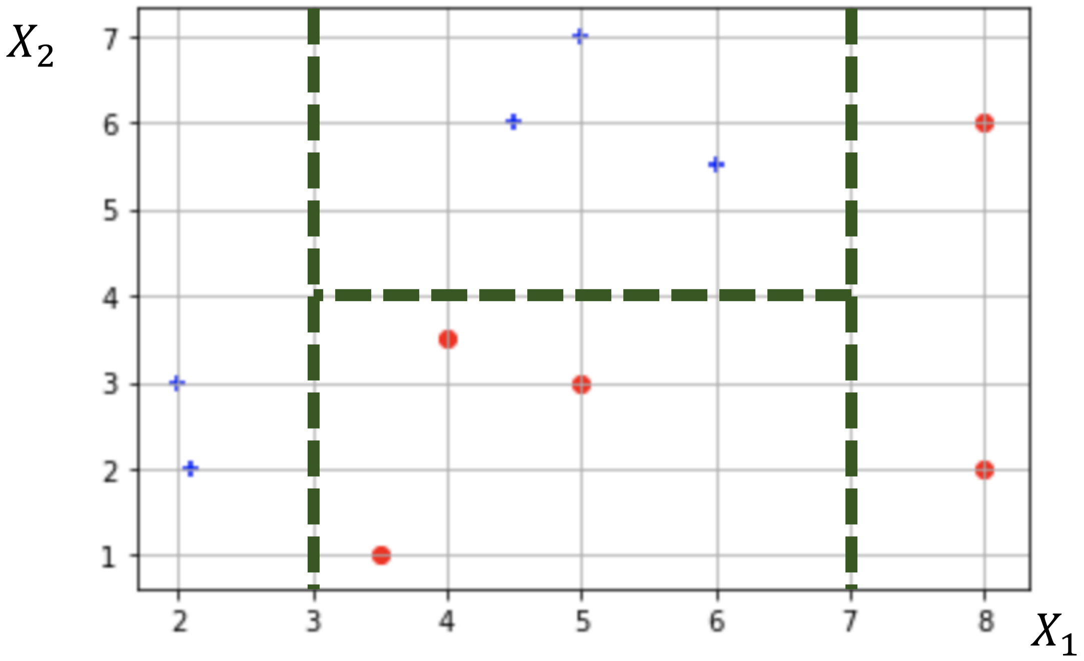
EXERCICE 1 Decision Tree

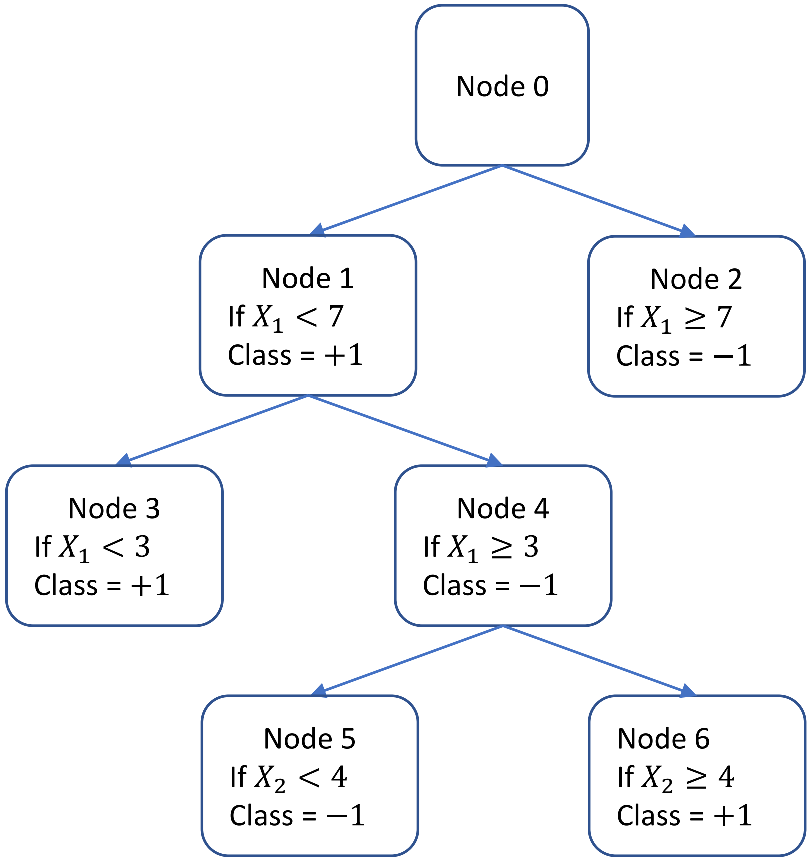
$C_1: \left\{ x^{(1)} = \binom{2}{3}, x^{(2)} = \binom{2.1}{2}, x^{(3)} = \binom{4.5}{6}, x^{(6)} = \binom{5}{7}, x^{(8)} = \binom{6}{5.5} \right\}$ with label 1

$C_2: \left\{ x^{(4)} = \binom{4}{3.5}, x^{(5)} = \binom{3.5}{1}, x^{(7)} = \binom{5}{3}, x^{(9)} = \binom{8}{6}, x^{(10)} = \binom{8}{2} \right\}$ with label -1

(b) <u>Draw</u> the node splits decision (or decision stumps) a decision tree in the scatter plot with the following hyperparameters:

- Minimum samples for a node split = 2
- Minimum samples for a terminal node or leaf = 2
- Maximum depth of tree (vertical depth) = 3





EXERCICE 1 Adaboost

$$C_1: \left\{x^{(1)} = \begin{pmatrix}2\\3\end{pmatrix}, x^{(2)} = \begin{pmatrix}2.1\\2\end{pmatrix}, x^{(3)} = \begin{pmatrix}4.5\\6\end{pmatrix}, x^{(6)} = \begin{pmatrix}5\\7\end{pmatrix}, x^{(8)} = \begin{pmatrix}6\\5.5\end{pmatrix}\right\} \text{ with label } 1$$

$$C_2: \left\{x^{(4)} = \begin{pmatrix}4\\3.5\end{pmatrix}, x^{(5)} = \begin{pmatrix}3.5\\1\end{pmatrix}, x^{(7)} = \begin{pmatrix}5\\3\end{pmatrix}, x^{(9)} = \begin{pmatrix}8\\6\end{pmatrix}, x^{(10)} = \begin{pmatrix}8\\2\end{pmatrix}\right\} \text{ with label } -1$$

(c) Draw an Adaboost ensembling architecture with 4 estimators (i.e. 4 weak learners) considering the following pseudo-code blocks:

h_i	**Decision** stump of i-th weak learner.
ϵ_i	Prediction **error** of i-th weak learner.
α_i	**Coefficient** of i-th weak learner.
D_i	Point **weights** for i-th weak learner.
$H(\cdot)$	**Strong learner**: combination of all trained weak learners.

EXERCICE 1 Adaboost

(c) Draw a 4 estimators (i.e. 4 weak learners) Adaboost ensembling architecture considering the following pseudo-code blocks:

REMINDER: Adaboost: combine weak learners to form a strong learner.

Weak learners are added one by one, looking at each iteration (weak learner) for the best possible pair (coefficient + weak learner) to add to the current ensemble model.

Each new weak learner is trained so it rectifies the errors (misclassified points) of the previous weak learners (which have been already trained).

HOW? Give more weight to the misclassified points so the penalization of misclassifying them is higher.

EXERCICE 1 Adaboost

(c) Draw a 4 estimators (i.e. 4 weak learners) Adaboost ensembling architecture considering the following pseudo-code blocks:

REMINDER: Adaboost

Weak learners are added one by one, looking at each iteration (weak learner) for the best possible pair (coefficient + weak learner) to add to the current ensemble model.

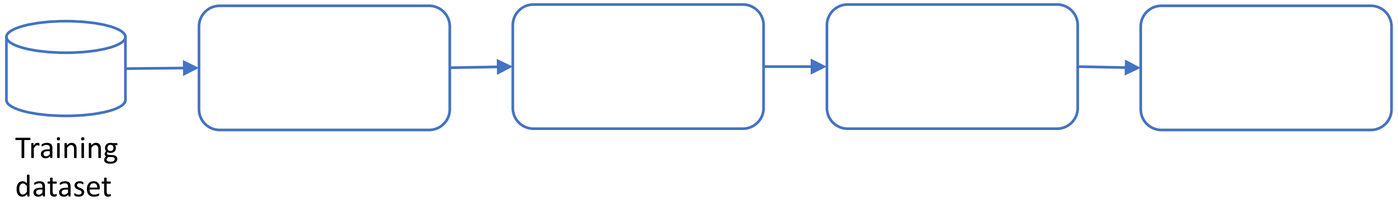
- Dataset: $\{\mathbf{x}^{(i)}, y^{(i)}\}, \mathrm{i} = 1, \dots, N,\ y^{(i)} \in \{+1, -1\}$
- Initialize weights for each observation $D_1^{(i)} = \frac{1}{N}$ for all i
- For $t = 1, \dots, T$ (for all the weak learners)
  - Train a weak classifier h_t that minimizes the (misclassification) error for the current weights $D_t^{(i)}$.
  - Compute (misclassification) error: $error_t = \sum_{i=1}^{N} D_t^{(i)} \left[y^{(i)} \neq h_t\left(\mathbf{x}^{(i)}\right) \right]$
  - Compute classifier coefficient: $\boldsymbol{\alpha}_t = \frac{1}{2} \ln\left(\frac{1 - error_t}{error_t}\right)$
  - Update weights: $D_{t+1}^{(i)} = D_t^{(i)} \cdot e^{-\boldsymbol{\alpha}_t y^{(i)} h_t\left(\mathbf{x}^{(i)}\right)} \implies (normalize) \quad D_{t+1}^{(i)} = \frac{D_{t+1}^{(i)}}{\sum_{i=1}^{N} D_{t+1}^{(i)}}$ (so they sum 1)

EXERCICE 1 Adaboost

(c) Draw a 4 estimators (i.e. 4 weak learners) Adaboost ensembling architecture considering the following pseudo-code blocks:

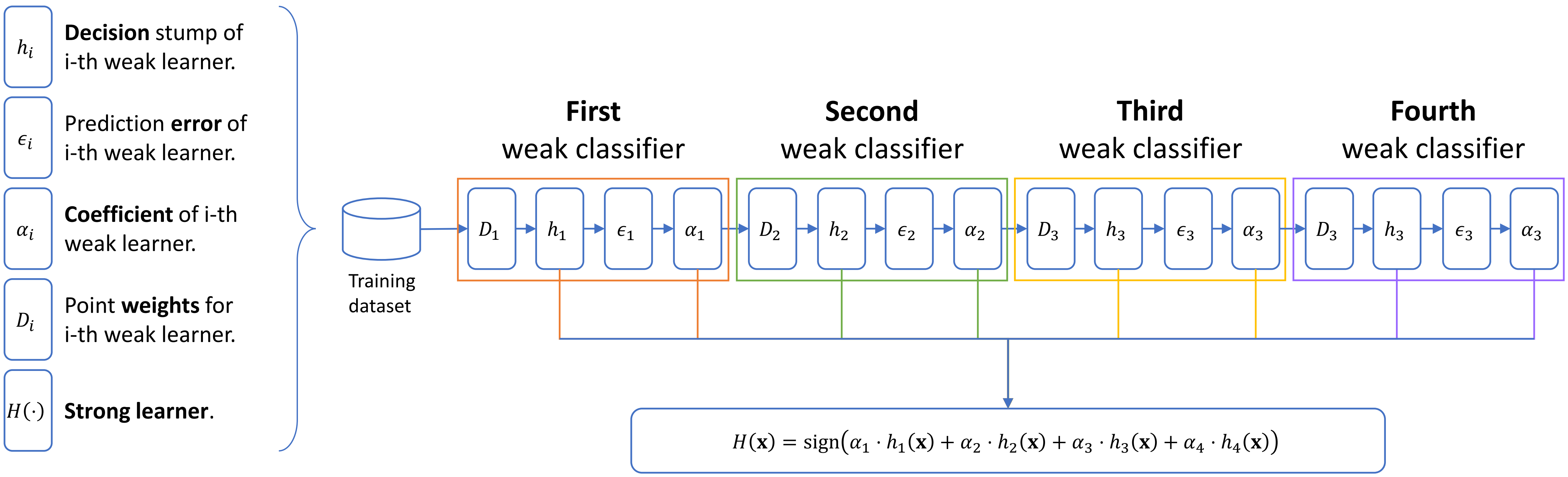
- h_i — **Decision** stump of i-th weak learner.
- ϵ_i — Prediction **error** of i-th weak learner.
- α_i — **Coefficient** of i-th weak learner.
- D_i — Point **weights** for i-th weak learner.
- $H(\cdot)$ — **Strong learner**.

Steps to train the first weak classifier



EXERCICE 1 Adaboost

(c) Draw a 4 estimators (i.e. 4 weak learners) Adaboost ensembling architecture considering the following pseudo-code blocks:



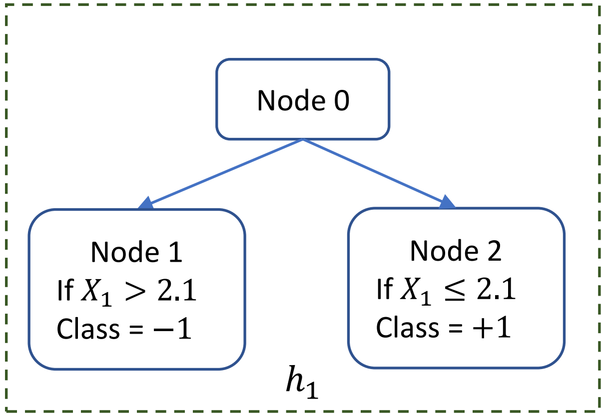
EXERCICE 1 Adaboost

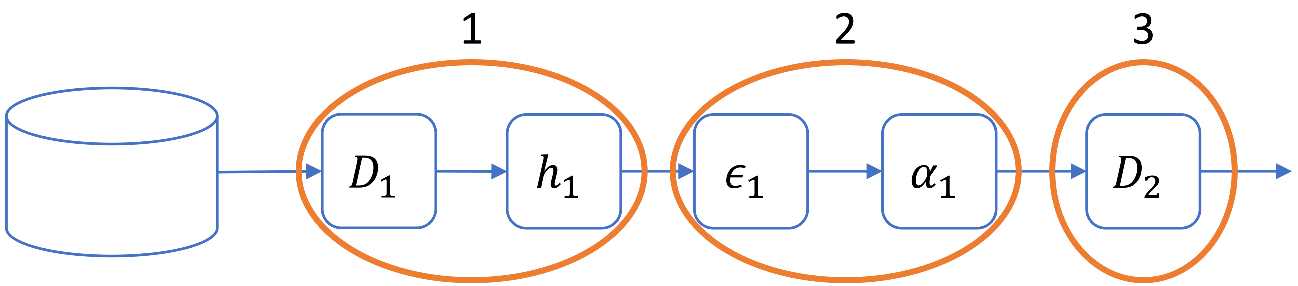
$C_1: \left\{ x^{(1)} = \binom{2}{3}, x^{(2)} = \binom{2.1}{2}, x^{(3)} = \binom{4.5}{6}, x^{(6)} = \binom{5}{7}, x^{(8)} = \binom{6}{5.5} \right\}$ with label 1

$C_2: \left\{ x^{(4)} = \binom{4}{3.5}, x^{(5)} = \binom{3.5}{1}, x^{(7)} = \binom{5}{3}, x^{(9)} = \binom{8}{6}, x^{(10)} = \binom{8}{2} \right\}$ with label -1

(d) First round:

1. Consider the decision stump of figure 1, build a table with the actual class, the weight, prediction, loss and weight*loss for every datapoint
2. Calculate the ϵ_1, α_1.
3. Update the weights and normalized weights for every datapoint for next round.





EXERCICE 1 Adaboost

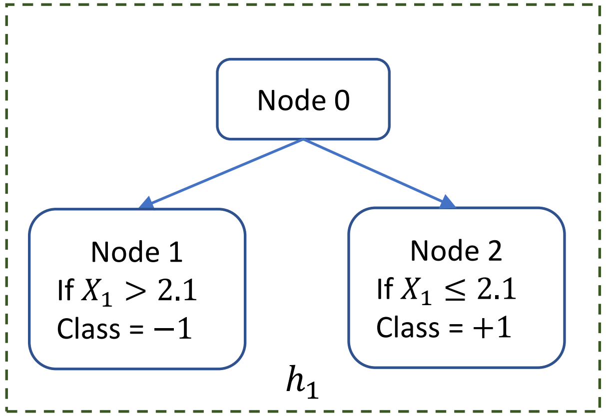
(d) First round:

1. Consider the decision stump of figure 1, build a table with the actual class, the weight, prediction, loss and weight*loss for every datapoint.

Apply the decision stump to every datapoint

Initial $D_1^{(i)} = \frac{1}{N}$ *for all* i

$[y^{(i)} \neq h_1(\mathbf{x}^{(i)})]$



	x_1	x_2	Y=Actual class	D_1	pred	Loss	D_1*loss
$\mathbf{x}^{(1)}$	2	3	1				
$\mathbf{x}^{(2)}$	2.1	2	1				
$\mathbf{x}^{(3)}$	4.5	6	1				
$\mathbf{x}^{(4)}$	4	3.5	-1				
$\mathbf{x}^{(5)}$	3.5	1	-1				
$\mathbf{x}^{(6)}$	5	7	1				
$\mathbf{x}^{(7)}$	5	3	-1				
$\mathbf{x}^{(8)}$	6	5.5	1				
$\mathbf{x}^{(9)}$	8	6	-1				
$\mathbf{x}^{(10)}$	8	2	-1				

EXERCICE 1 Adaboost

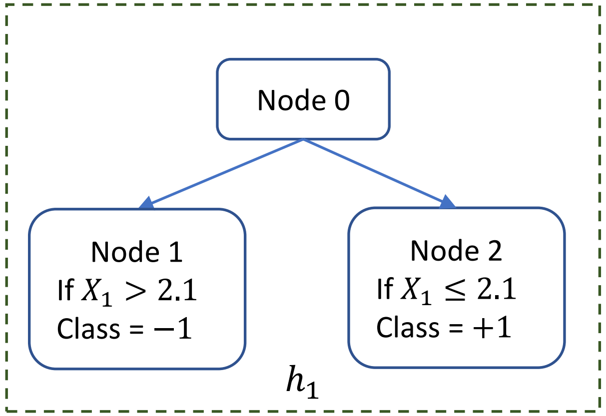
(d) First round:

1. Consider the decision stump of figure 1, build a table with the actual class, the weight, prediction, loss and weight*loss for every datapoint.

Apply the decision stump to every datapoint

Initial $D_1^{(i)} = \frac{1}{N}$ *for all* i

$[y^{(i)} \neq h_1(\mathbf{x}^{(i)})]$



	x_1	x_2	Y=Actual class	D_1	pred	Loss	D_1*loss
$\mathbf{x}^{(1)}$	2	3	1	0.1	1	0	0
$\mathbf{x}^{(2)}$	2.1	2	1	0.1	1	0	0
$\mathbf{x}^{(3)}$	4.5	6	1	0.1	-1	1	0.1
$\mathbf{x}^{(4)}$	4	3.5	-1	0.1	-1	0	0
$\mathbf{x}^{(5)}$	3.5	1	-1	0.1	-1	0	0
$\mathbf{x}^{(6)}$	5	7	1	0.1	-1	1	0.1
$\mathbf{x}^{(7)}$	5	3	-1	0.1	-1	0	0
$\mathbf{x}^{(8)}$	6	5.5	1	0.1	-1	1	0.1
$\mathbf{x}^{(9)}$	8	6	-1	0.1	-1	0	0
$\mathbf{x}^{(10)}$	8	2	-1	0.1	-1	0	0

EXERCICE 1 Adaboost

(d) First round:

1. Consider the decision stump of figure 1, build a table with the actual class, the weight, prediction, loss and weight*loss for every datapoint.
2. Calculate the ϵ_1, α_1.

	x_1	x_2	D_1	D_1*loss
$\mathbf{x}^{(1)}$	2	3	0.1	0
$\mathbf{x}^{(2)}$	2.1	2	0.1	0
$\mathbf{x}^{(3)}$	4.5	6	0.1	0.1
$\mathbf{x}^{(4)}$	4	3.5	0.1	0
$\mathbf{x}^{(5)}$	3.5	1	0.1	0
$\mathbf{x}^{(6)}$	5	7	0.1	0.1
$\mathbf{x}^{(7)}$	5	3	0.1	0
$\mathbf{x}^{(8)}$	6	5.5	0.1	0.1
$\mathbf{x}^{(9)}$	8	6	0.1	0
$\mathbf{x}^{(10)}$	8	2	0.1	0

Σ

$$\epsilon_1 = \sum_{i=1}^{10} D_1(i) * loss\left(\mathbf{x}^{(i)}\right) = 0.3$$

$$\alpha_1 = \frac{1}{2}\ln\left(\frac{1-\epsilon_1}{\epsilon_1}\right) = \frac{1}{2}\ln\left(\frac{1-0.3}{0.3}\right) = 0.42$$

EXERCICE 1 Adaboost

(d) First round:

1. Consider the decision stump of figure 1, build a table with the actual class, the weight, prediction, loss and weight*loss for every datapoint.
2. Calculate the ϵ_1, α_1.
3. Update the weights and normalized weights for every datapoint for next round.

	x_1	x_2	Actual class	pred	D_1	D_2	Norm_D_2
$\mathbf{x}^{(1)}$	2	3	1	1	0.1		
$\mathbf{x}^{(2)}$	2.1	2	1	1	0.1		
$\mathbf{x}^{(3)}$	4.5	6	1	-1	0.1		
$\mathbf{x}^{(4)}$	4	3.5	-1	-1	0.1		
$\mathbf{x}^{(5)}$	3.5	1	-1	-1	0.1		
$\mathbf{x}^{(6)}$	5	7	1	-1	0.1		
$\mathbf{x}^{(7)}$	5	3	-1	-1	0.1		
$\mathbf{x}^{(8)}$	6	5.5	1	-1	0.1		
$\mathbf{x}^{(9)}$	8	6	-1	-1	0.1		
$\mathbf{x}^{(10)}$	8	2	-1	-1	0.1		

$$\boxed{D_2^{(i)} = D_1^{(i)} \cdot e^{-\alpha_1 y^{(i)} h_1(\mathbf{x}^{(i)})}}$$

$$\boxed{\text{norm\_}D_2^{(i)} = \frac{D_2^{(i)}}{\sum_{j=1}^{N} D_2^{(j)}}}$$

EXERCICE 1 Adaboost

(d) First round:

1. Consider the decision stump of figure 1, build a table with the actual class, the weight, prediction, loss and weight*loss for every datapoint.
2. Calculate the ϵ_1, α_1.
3. Update the weights and normalized weights for every datapoint for next round.

	x_1	x_2	Actual class	pred	D_1	D_2	Norm_D_2
$\mathbf{x}^{(1)}$	2	3	1	1	0.1	0.065	0.071
$\mathbf{x}^{(2)}$	2.1	2	1	1	0.1	0.065	0.071
$\mathbf{x}^{(3)}$	4.5	6	1	-1	0.1	0.153	0.167
$\mathbf{x}^{(4)}$	4	3.5	-1	-1	0.1	0.065	0.071
$\mathbf{x}^{(5)}$	3.5	1	-1	-1	0.1	0.065	0.071
$\mathbf{x}^{(6)}$	5	7	1	-1	0.1	0.153	0.167
$\mathbf{x}^{(7)}$	5	3	-1	-1	0.1	0.065	0.071
$\mathbf{x}^{(8)}$	6	5.5	1	-1	0.1	0.153	0.167
$\mathbf{x}^{(9)}$	8	6	-1	-1	0.1	0.065	0.071
$\mathbf{x}^{(10)}$	8	2	-1	-1	0.1	0.065	0.071

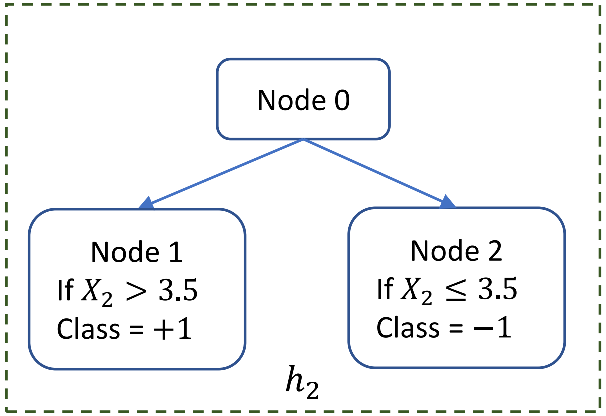
$$D_2^{(1)} = D_1^{(1)} \cdot e^{-\alpha_1 y^{(1)} h_1(\mathbf{x}^{(1)})}$$
$$= 0.1 \cdot e^{-0.42 \cdot 1 \cdot 1} = 0.065$$

$$\text{norm\_}D_2^{(1)} = \frac{D_2^{(1)}}{\sum_{j=1}^{N} D_2^{(j)}} = \frac{0.065}{0.914} = 0.071$$

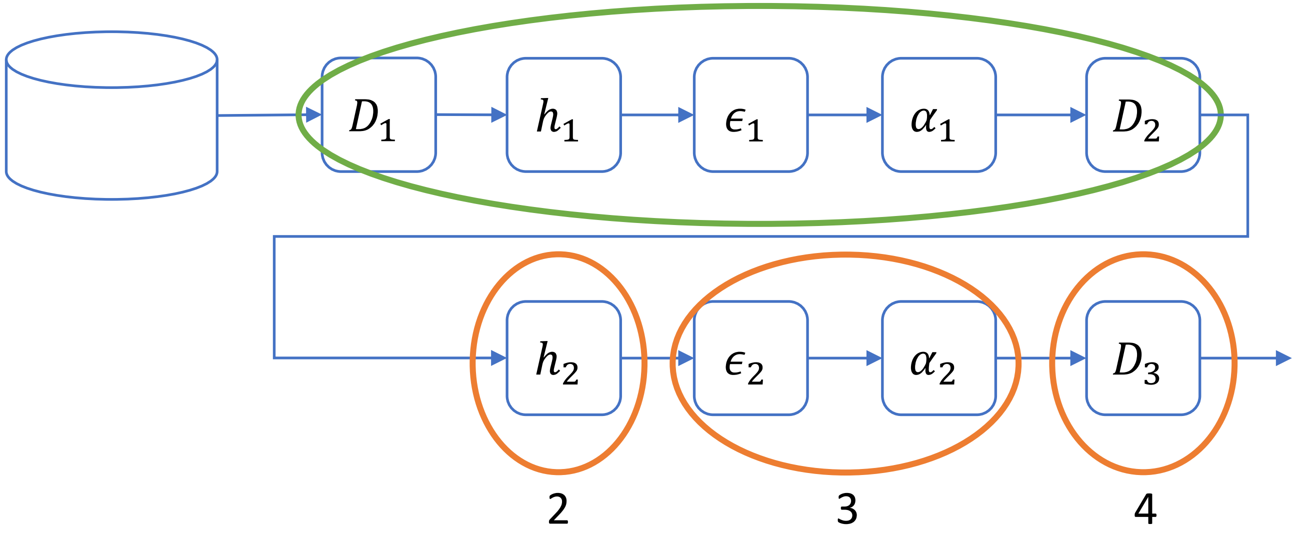
EXERCICE 1 Adaboost

(e) Second round:

1. Plot the points which sizes should be aligned with Norm_$\boldsymbol{D_2}$(i) value
2. Consider the decision stump of figure 2, build a table with the actual class, the weight, prediction, loss and weight*loss for every datapoint
3. Calculate the ϵ_2, α_2
4. Update the weights and normalized weights for every datapoint for next round



(d)



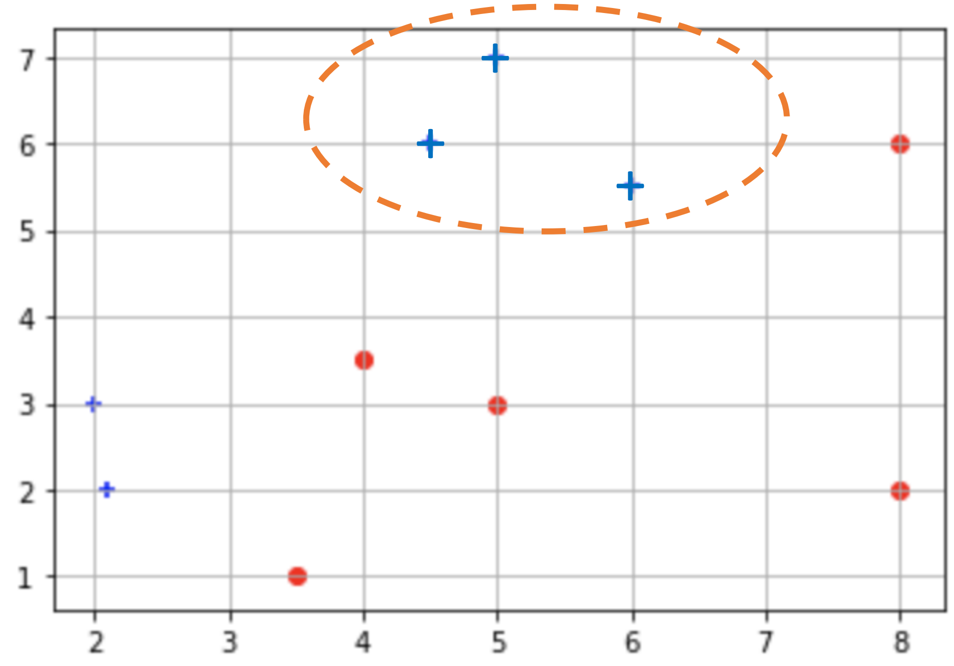
EXERCICE 1 Adaboost

(e) Second round:

1. Plot the points which sizes should be aligned with Norm_$\boldsymbol{D_2}$(i) value.

- These points were **misclassified** by the first classifier.
- These points are “more important” in the 2nd round (second weak classifier).
- The next classifier will be trained to focus in classifying well this three points.

Norm_$\boldsymbol{D_2} = 0.167$



Norm_$\boldsymbol{D_2} = 0.071$

EXERCICE 1 Adaboost

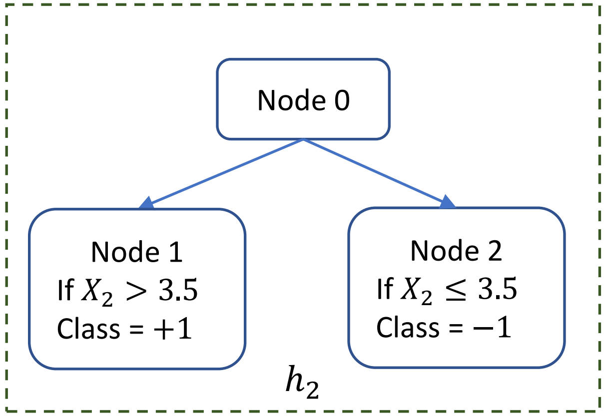
(e) Second round:

2. Consider the decision stump of figure 2, build a table with the actual class, the weight, prediction, loss and weight*loss for every datapoint.

Apply the decision stump to every datapoint

From in the first round. Updated from h_1.

$[y^{(i)} \neq h_2(\mathbf{x}^{(i)})]$



	x_1	x_2	Y=Actual class	Norm_D_2	pred	Loss	Norm_D_2 *loss
$\mathbf{x}^{(1)}$	2	3	1	0.071			
$\mathbf{x}^{(2)}$	2.1	2	1	0.071			
$\mathbf{x}^{(3)}$	4.5	6	1	0.167			
$\mathbf{x}^{(4)}$	4	3.5	-1	0.071			
$\mathbf{x}^{(5)}$	3.5	1	-1	0.071			
$\mathbf{x}^{(6)}$	5	7	1	0.167			
$\mathbf{x}^{(7)}$	5	3	-1	0.071			
$\mathbf{x}^{(8)}$	6	5.5	1	0.167			
$\mathbf{x}^{(9)}$	8	6	-1	0.071			
$\mathbf{x}^{(10)}$	8	2	-1	0.071			

EXERCICE 1 Adaboost

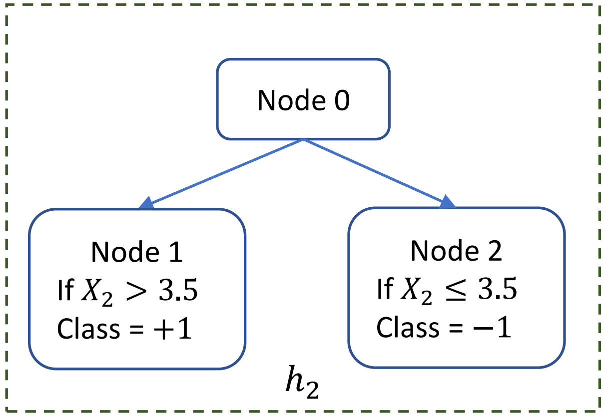
(e) Second round:

2. Consider the decision stump of figure 2, build a table with the actual class, the weight, prediction, loss and weight*loss for every datapoint.

Apply the decision stump to every datapoint

From in the first round. Updated from h_1.

$[y^{(i)} \neq h_2(\mathbf{x}^{(i)})]$



	x_1	x_2	Y=Actual class	Norm_D_2	pred	Loss	Norm_D_2 *loss
$\mathbf{x}^{(1)}$	2	3	1	0.071	-1	1	0.071
$\mathbf{x}^{(2)}$	2.1	2	1	0.071	-1	1	0.071
$\mathbf{x}^{(3)}$	4.5	6	1	0.167	1	0	0
$\mathbf{x}^{(4)}$	4	3.5	-1	0.071	-1	0	0
$\mathbf{x}^{(5)}$	3.5	1	-1	0.071	-1	0	0
$\mathbf{x}^{(6)}$	5	7	1	0.167	1	0	0
$\mathbf{x}^{(7)}$	5	3	-1	0.071	-1	0	0
$\mathbf{x}^{(8)}$	6	5.5	1	0.167	1	0	0
$\mathbf{x}^{(9)}$	8	6	-1	0.071	1	1	0.071
$\mathbf{x}^{(10)}$	8	2	-1	0.071	-1	0	0

EXERCICE 1 Adaboost

(e) Second round:

1. Plot the points which sizes should be aligned with Norm_$\boldsymbol{D_2}$(i) value
2. Consider the decision stump of figure 2, build a table with the actual class, the weight, prediction, loss and weight*loss for every datapoint
3. Calculate the ϵ_2, α_2

	x_1	x_2	Norm_$\boldsymbol{D_2}$	Norm_$\boldsymbol{D_2}$*loss
$\mathbf{x}^{(1)}$	2	3	0.071	0.071
$\mathbf{x}^{(2)}$	2.1	2	0.071	0.071
$\mathbf{x}^{(3)}$	4.5	6	0.167	0
$\mathbf{x}^{(4)}$	4	3.5	0.071	0
$\mathbf{x}^{(5)}$	3.5	1	0.071	0
$\mathbf{x}^{(6)}$	5	7	0.167	0
$\mathbf{x}^{(7)}$	5	3	0.071	0
$\mathbf{x}^{(8)}$	6	5.5	0.167	0
$\mathbf{x}^{(9)}$	8	6	0.071	0.071
$\mathbf{x}^{(10)}$	8	2	0.071	0

Σ

$$\epsilon_2 = \sum_{i=1}^{10} \text{norm\_}D_2(i) * loss(i) = 0.21$$

$$\alpha_2 = \frac{1}{2}\ln\left(\frac{1-\epsilon_2}{\epsilon_2}\right) = \frac{1}{2}\ln\left(\frac{1-0.21}{0.21}\right) = 0.65$$

EXERCICE 1 Adaboost

(e) Second round:

4. Update the weights and normalized weights for every datapoint for next round

	x_1	x_2	Actual class	pred	Norm_D_2	D_3	Norm_D_3
$\mathbf{x}^{(1)}$	2	3	1	-1	0.071		
$\mathbf{x}^{(2)}$	2.1	2	1	-1	0.071		
$\mathbf{x}^{(3)}$	4.5	6	1	1	0.167		
$\mathbf{x}^{(4)}$	4	3.5	-1	-1	0.071		
$\mathbf{x}^{(5)}$	3.5	1	-1	-1	0.071		
$\mathbf{x}^{(6)}$	5	7	1	1	0.167		
$\mathbf{x}^{(7)}$	5	3	-1	-1	0.071		
$\mathbf{x}^{(8)}$	6	5.5	1	1	0.167		
$\mathbf{x}^{(9)}$	8	6	-1	1	0.071		
$\mathbf{x}^{(10)}$	8	2	-1	-1	0.071		

$$D_3^{(i)} = \mathrm{norm}\_D_2^{(i)} \cdot e^{-\alpha_2 y^{(i)} h_2(\mathbf{x}^{(i)})}$$

$$\mathrm{norm}\_D_3^{(i)} = \frac{D_3^{(i)}}{\sum_{j=1}^{N} D_3^{(j)}}$$

EXERCICE 1 Adaboost

(e) Second round:

4. Update the weights and normalized weights for every datapoint for next round

	x_1	x_2	Actual class	pred	Norm_D_2	D_3	Norm_D_3
$\mathbf{x}^{(1)}$	2	3	1	-1	0.071	0.136	0.167
$\mathbf{x}^{(2)}$	2.1	2	1	-1	0.071	0.136	0.167
$\mathbf{x}^{(3)}$	4.5	6	1	1	0.167	0.087	0.106
$\mathbf{x}^{(4)}$	4	3.5	-1	-1	0.071	0.037	0.045
$\mathbf{x}^{(5)}$	3.5	1	-1	-1	0.071	0.037	0.045
$\mathbf{x}^{(6)}$	5	7	1	1	0.167	0.087	0.106
$\mathbf{x}^{(7)}$	5	3	-1	-1	0.071	0.037	0.045
$\mathbf{x}^{(8)}$	6	5.5	1	1	0.167	0.087	0.106
$\mathbf{x}^{(9)}$	8	6	-1	1	0.071	0.137	0.167
$\mathbf{x}^{(10)}$	8	2	-1	-1	0.071	0.037	0.045

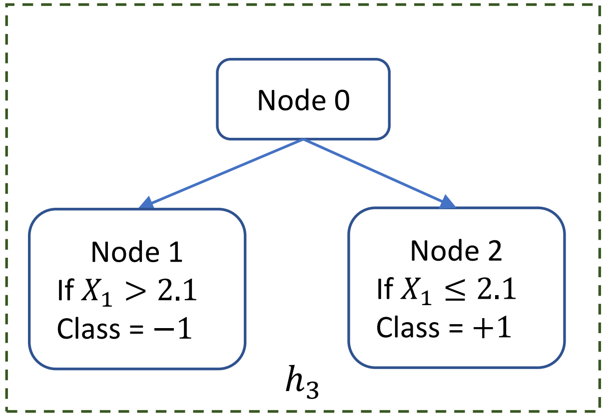
$$D_3^{(1)} = \text{norm\_}D_2^{(1)} \cdot e^{-\alpha_2 y^{(1)} h_2(\mathbf{x}^{(1)})} = 0.071 \cdot e^{-0.65 \cdot 1 \cdot (-1)} = 0.136$$

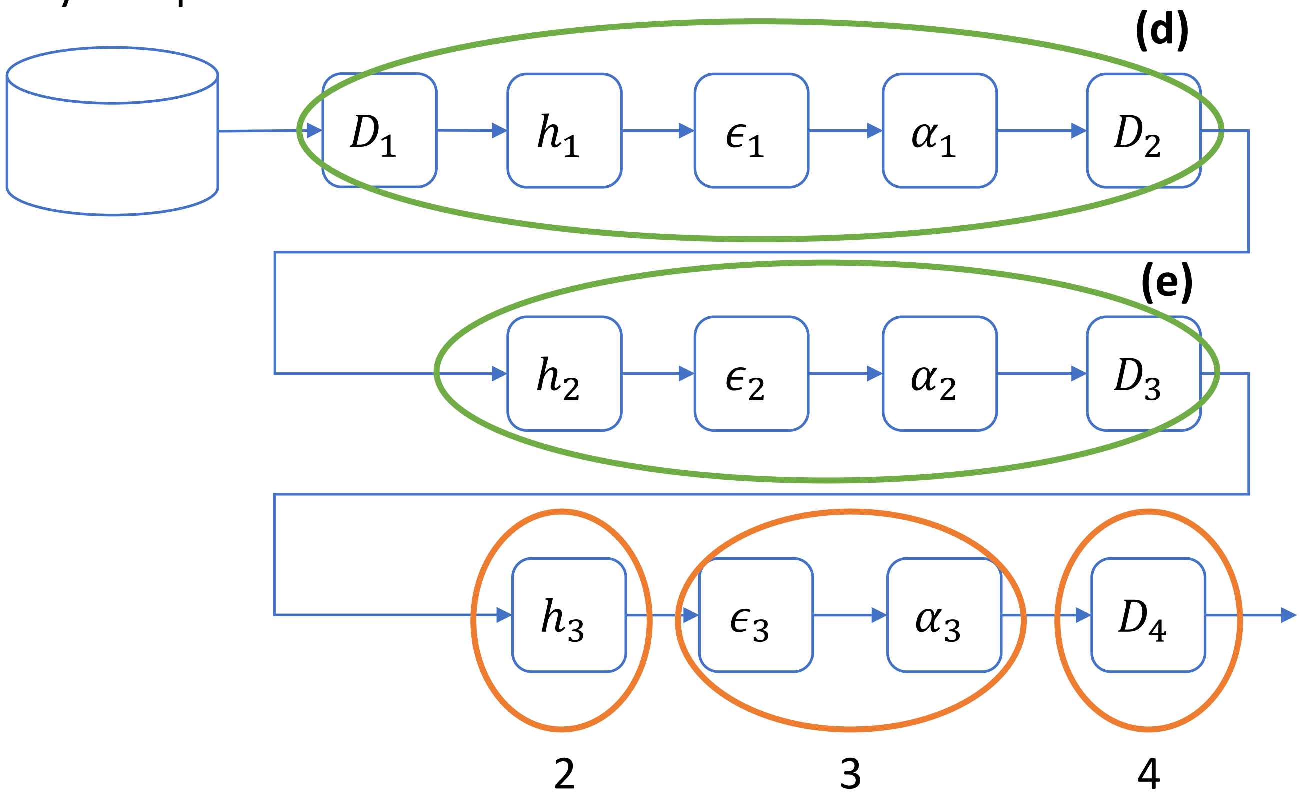
$$\text{norm\_}D_3^{(1)} = \frac{D_3^{(1)}}{\sum_{j=1}^{N} D_3^{(j)}} = \frac{0.136}{0.818} = 0.167$$

EXERCICE 1 Adaboost

(f) Third round:

1. Plot the points which sizes should be aligned with Norm_$\boldsymbol{D}_3$(i) value
2. Consider the decision stump of figure 3, build a table with the actual class, the weight, prediction, loss and weight*loss for every datapoint
3. Calculate the ϵ_3, α_3
4. Update the weights and normalized weights for every datapoint for next round



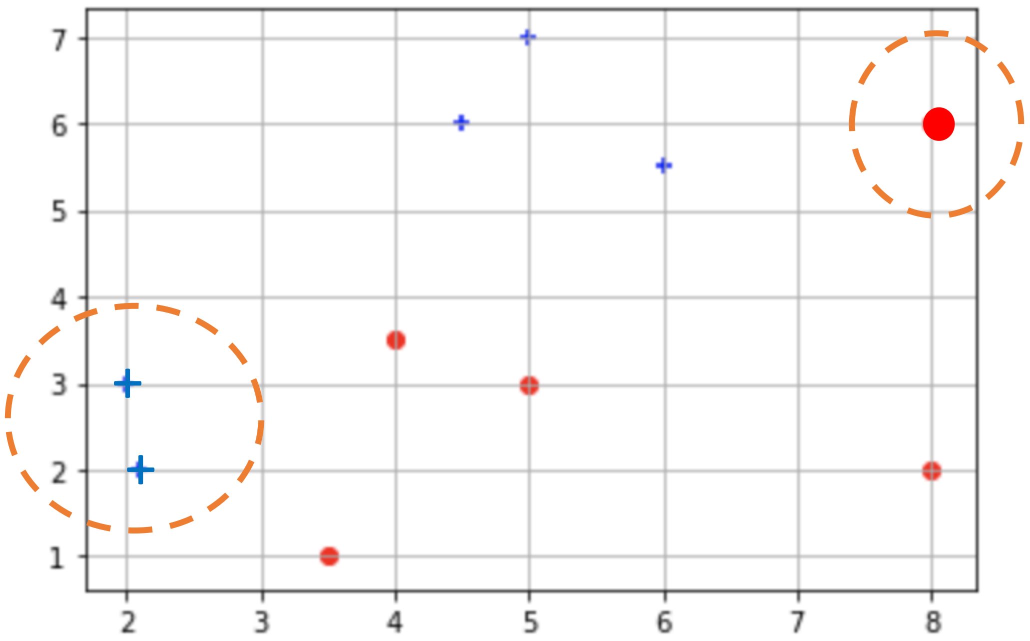


EXERCICE 1 Adaboost

(f) Third round:

1. Plot the points which sizes should be aligned with Norm_$\boldsymbol{D}_3$(i) value

- These points were **misclassified** by the second classifier.
- These points are "more important" in the 3rd round (third weak classifier).
- The next classifier will be trained to focus in classifying well this three points.



EXERCICE 1 Adaboost

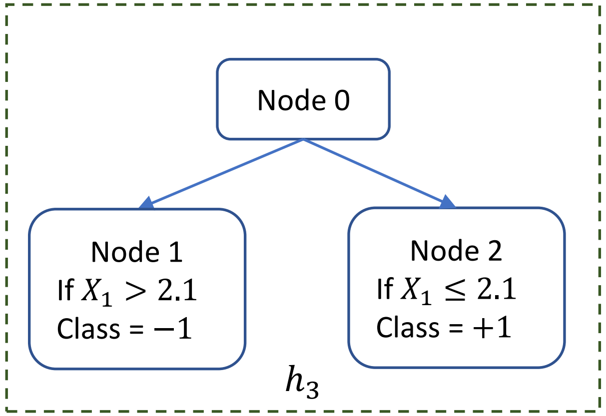
(f) Third round:

2. Consider the decision stump of figure 3, build a table with the actual class, the weight, prediction, loss and weight*loss for every datapoint

Apply the decision stump to every datapoint

From in the second round. Updated from h_2.

$[y^{(i)} \neq h_3(\mathbf{x}^{(i)})]$



	x_1	x_2	Y=Actual class	Norm_D_3	pred	Loss	Norm_D_3 *loss
$\mathbf{x}^{(1)}$	2	3	1	0.167	1	0	0
$\mathbf{x}^{(2)}$	2.1	2	1	0.167	1	0	0
$\mathbf{x}^{(3)}$	4.5	6	1	0.106	-1	1	0.106
$\mathbf{x}^{(4)}$	4	3.5	-1	0.045	-1	0	0
$\mathbf{x}^{(5)}$	3.5	1	-1	0.045	-1	0	0
$\mathbf{x}^{(6)}$	5	7	1	0.106	-1	1	0.106
$\mathbf{x}^{(7)}$	5	3	-1	0.045	-1	0	0
$\mathbf{x}^{(8)}$	6	5.5	1	0.106	-1	1	0.106
$\mathbf{x}^{(9)}$	8	6	-1	0.167	-1	0	0
$\mathbf{x}^{(10)}$	8	2	-1	0.045	-1	0	0

EXERCICE 1 Adaboost

(f) Third round:

1. Plot the points which sizes should be aligned with Norm_$\boldsymbol{D}_3$(i) value
2. Consider the decision stump of figure 3, build a table with the actual class, the weight, prediction, loss and weight*loss for every datapoint
3. Calculate the ϵ_3, α_3

	x_1	x_2	Norm_D_3	Norm_D_3*loss
$\mathbf{x}^{(1)}$	2	3	0.167	0
$\mathbf{x}^{(2)}$	2.1	2	0.167	0
$\mathbf{x}^{(3)}$	4.5	6	0.106	0.106
$\mathbf{x}^{(4)}$	4	3.5	0.045	0
$\mathbf{x}^{(5)}$	3.5	1	0.045	0
$\mathbf{x}^{(6)}$	5	7	0.106	0.106
$\mathbf{x}^{(7)}$	5	3	0.045	0
$\mathbf{x}^{(8)}$	6	5.5	0.106	0.106
$\mathbf{x}^{(9)}$	8	6	0.167	0
$\mathbf{x}^{(10)}$	8	2	0.045	0

Σ

$$\epsilon_3 = \sum_{i=1}^{10} \text{norm\_}D_3(i) * loss(i) = 0.31$$

$$\alpha_3 = \frac{1}{2}\ln\left(\frac{1-\epsilon_3}{\epsilon_3}\right) = \frac{1}{2}\ln\left(\frac{1-0.31}{0.31}\right) = 0.38$$

EXERCICE 1 Adaboost

(f) Third round:

4. Update the weights and normalized weights for every datapoint for next round

	x_1	x_2	Actual class	pred	Norm_D_3	D_4	Norm_D_4
$\mathbf{x}^{(1)}$	2	3	1	1	0.167	0.136	0.122
$\mathbf{x}^{(2)}$	2.1	2	1	1	0.167	0.136	0.122
$\mathbf{x}^{(3)}$	4.5	6	1	-1	0.106	0.087	0.167
$\mathbf{x}^{(4)}$	4	3.5	-1	-1	0.045	0.037	0.033
$\mathbf{x}^{(5)}$	3.5	1	-1	-1	0.045	0.037	0.033
$\mathbf{x}^{(6)}$	5	7	1	-1	0.106	0.087	0.167
$\mathbf{x}^{(7)}$	5	3	-1	-1	0.045	0.037	0.033
$\mathbf{x}^{(8)}$	6	5.5	1	-1	0.106	0.087	0.167
$\mathbf{x}^{(9)}$	8	6	-1	-1	0.167	0.137	0.122
$\mathbf{x}^{(10)}$	8	2	-1	-1	0.045	0.037	0.033

$$D_4^{(1)} = \text{norm\_}D_3^{(1)} \cdot e^{-\alpha_3 y^{(1)} h_3(\mathbf{x}^{(1)})}$$
$$= 0.167 \cdot e^{-0.38 \cdot 1 \cdot 1} = 0.136$$

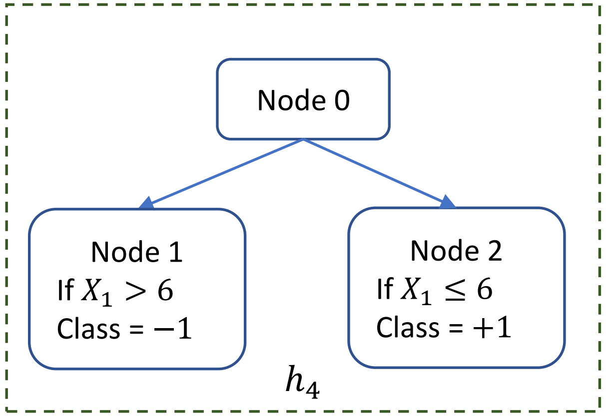
$$\text{norm\_}D_4^{(1)} = \frac{D_4^{(1)}}{\sum_{j=1}^{N} D_4^{(j)}} = 0.167$$

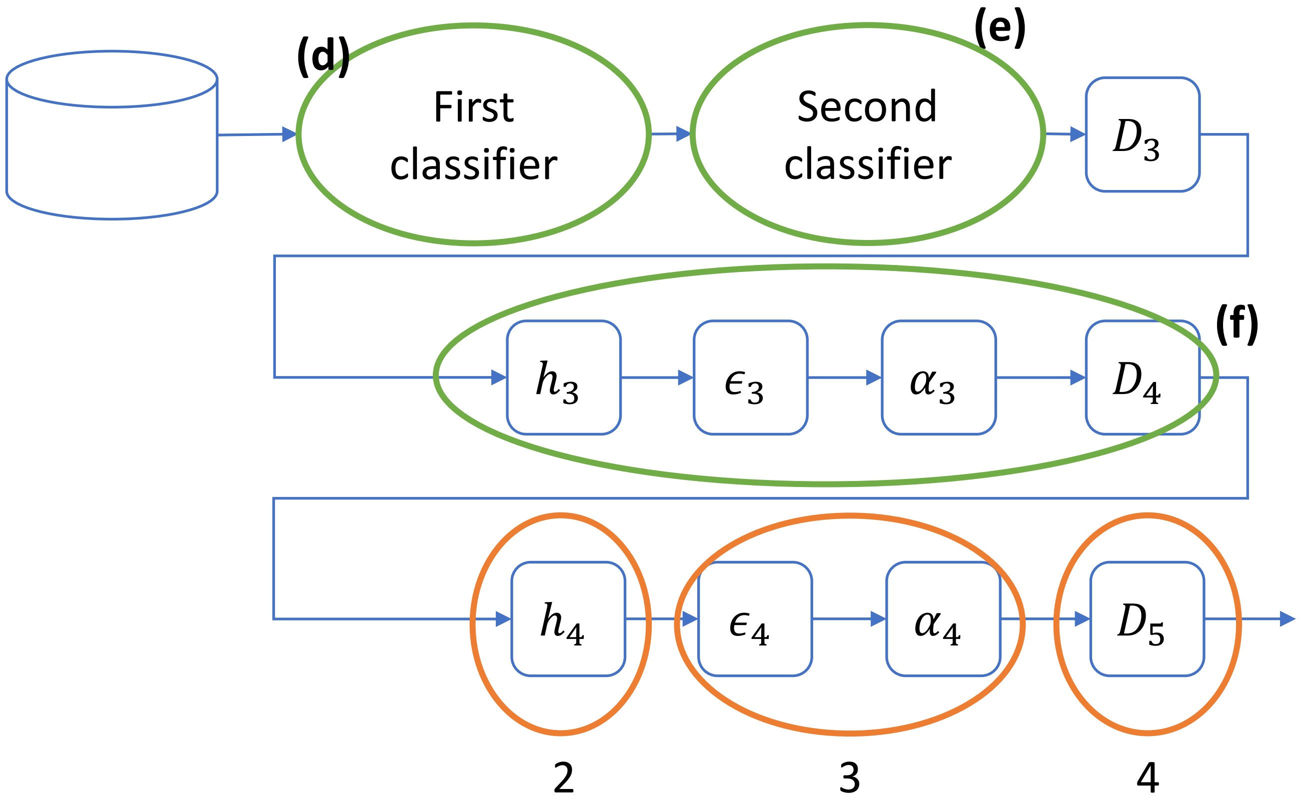
EXERCICE 1 Adaboost

(g) Fourth round:

1. Plot the points which sizes should be aligned with Norm_$\boldsymbol{D_4}$(i) value
2. Consider the decision stump of figure 4, build a table with the actual class, the weight, prediction, loss and weight*loss for every datapoint
3. Calculate the ϵ_3, α_3
4. Update the weights and normalized weights for every datapoint for next round

Not needed, since we only have 4 weak classifiers.



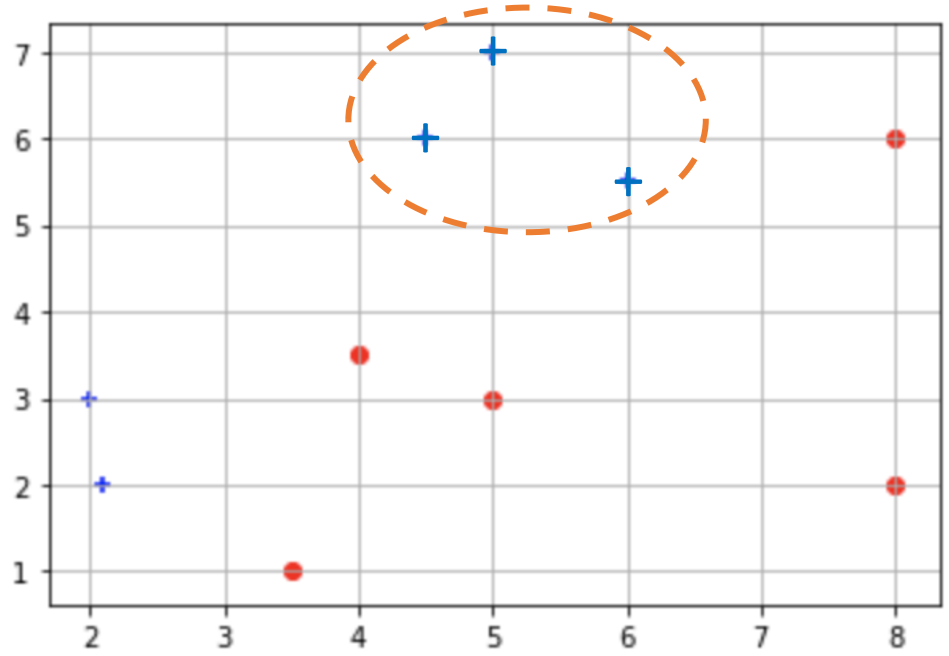


EXERCICE 1 Adaboost

(g) Fourth round:

1. Plot the points which sizes should be aligned with Norm_$\boldsymbol{D_4}$(i) value

- These points were **misclassified** by the third classifier.
- These points are "more important" in the 4th round (fourth weak classifier).
- The next classifier will be trained to focus in classifying well this three points.



EXERCICE 1 Adaboost

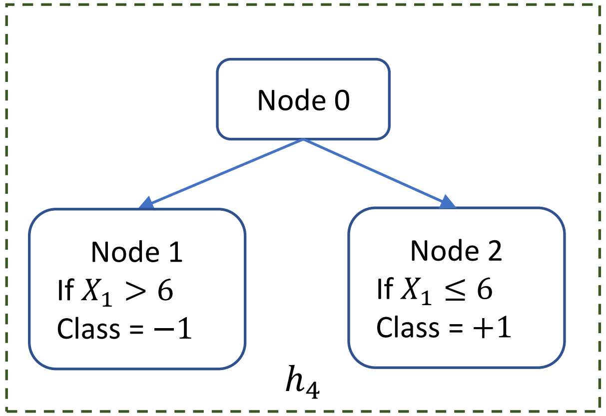
(g) Fourth round:

2. Consider the decision stump of figure 4, build a table with the actual class, the weight, prediction, loss and weight*loss for every datapoint

Apply the decision stump to every datapoint

From in the third round. Updated from h_3.

$[y^{(i)} \neq h_4(\mathbf{x}^{(i)})]$



	x_1	x_2	Y=Actual class	Norm_D_4	pred	Loss	Norm_D_4 *loss
$\mathbf{x}^{(1)}$	2	3	1	0.122	1	0	0
$\mathbf{x}^{(2)}$	2.1	2	1	0.122	1	0	0
$\mathbf{x}^{(3)}$	4.5	6	1	0.167	1	0	0
$\mathbf{x}^{(4)}$	4	3.5	-1	0.033	1	1	0.033
$\mathbf{x}^{(5)}$	3.5	1	-1	0.033	1	1	0.033
$\mathbf{x}^{(6)}$	5	7	1	0.167	1	0	0
$\mathbf{x}^{(7)}$	5	3	-1	0.033	1	1	0.033
$\mathbf{x}^{(8)}$	6	5.5	1	0.167	1	0	0
$\mathbf{x}^{(9)}$	8	6	-1	0.122	-1	0	0
$\mathbf{x}^{(10)}$	8	2	-1	0.033	-1	0	0

EXERCICE 1 Adaboost

(g) Fourth round:

1. Plot the points which sizes should be aligned with Norm_$\boldsymbol{D_4}$(i) value
2. Consider the decision stump of figure 4, build a table with the actual class, the weight, prediction, loss and weight*loss for every datapoint
3. Calculate the ϵ_4, α_4

	x_1	x_2	Norm_D_4	Norm_D_4*loss
$\mathbf{x}^{(1)}$	2	3	0.122	0
$\mathbf{x}^{(2)}$	2.1	2	0.122	0
$\mathbf{x}^{(3)}$	4.5	6	0.167	0
$\mathbf{x}^{(4)}$	4	3.5	0.033	0.033
$\mathbf{x}^{(5)}$	3.5	1	0.033	0.033
$\mathbf{x}^{(6)}$	5	7	0.167	0
$\mathbf{x}^{(7)}$	5	3	0.033	0.033
$\mathbf{x}^{(8)}$	6	5.5	0.167	0
$\mathbf{x}^{(9)}$	8	6	0.122	0
$\mathbf{x}^{(10)}$	8	2	0.033	0

Σ

$$\epsilon_4 = \sum_{i=1}^{10} \text{norm\_}D_3(i) * loss(i) = 0.10$$

$$\alpha_4 = \frac{1}{2}\ln\left(\frac{1-\epsilon_4}{\epsilon_4}\right) = \frac{1}{2}\ln\left(\frac{1-0.10}{0.10}\right) = 1.10$$

EXERCICE 1 Adaboost

(h) Calculate the prediction for $\mathbf{x} = \binom{2}{3}$.

First, let's write the expression of the adaboost classifier.

$$\left.\begin{aligned} \alpha_1 &= 0.42 \\ \alpha_2 &= 0.65 \\ \alpha_3 &= 0.38 \\ \alpha_4 &= 1.10 \end{aligned}\right\} \quad H(\mathbf{x}) = ?$$

EXERCICE 1 Adaboost

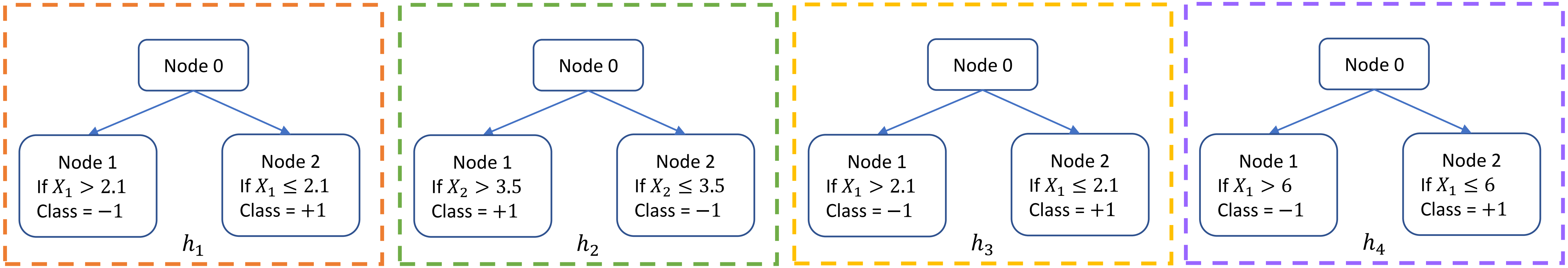
(h) Calculate the prediction for $\mathbf{x} = \binom{2}{3}$.

First, let's write the expression of the adaboost classifier.

$$\left.\begin{aligned} \alpha_1 &= 0.42 \\ \alpha_2 &= 0.65 \\ \alpha_3 &= 0.38 \\ \alpha_4 &= 1.10 \end{aligned}\right\} \quad H(\mathbf{x}) = \text{sign}(0.42 \cdot h_1(\mathbf{x}) + 0.65 \cdot h_2(\mathbf{x}) + 0.38 \cdot h_3(\mathbf{x}) + 1.10 \cdot h_4(\mathbf{x}))$$

EXERCICE 1 Adaboost

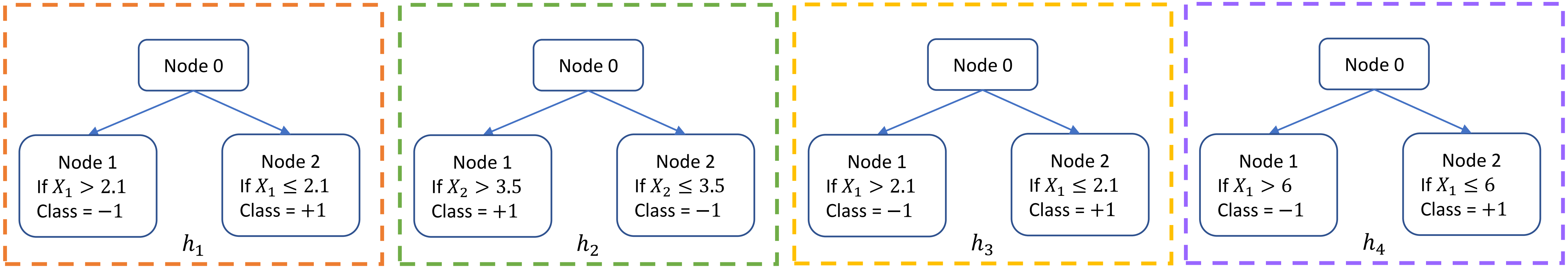
(h) Calculate the prediction for $\mathbf{x} = \binom{2}{3}$.



$$H(\mathbf{x}) = \operatorname{sign}(0.42 \cdot h_1(\mathbf{x}) + 0.65 \cdot h_2(\mathbf{x}) + 0.38 \cdot h_3(\mathbf{x}) + 1.10 \cdot h_4(\mathbf{x}))$$

EXERCICE 1 Adaboost

(h) Calculate the prediction for $\mathbf{x} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$.



$$H(\mathbf{x}) = \text{sign}(0.42 \cdot h_1(\mathbf{x}) + 0.65 \cdot h_2(\mathbf{x}) + 0.38 \cdot h_3(\mathbf{x}) + 1.10 \cdot h_4(\mathbf{x}))$$

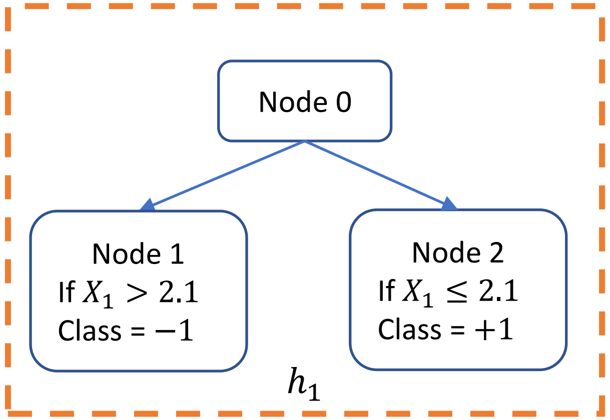
$$H(\mathbf{x}) = \text{sign}(0.42 \cdot 1 + 0.65 \cdot (-1) + 0.38 \cdot 1 + 1.10 \cdot 1)$$

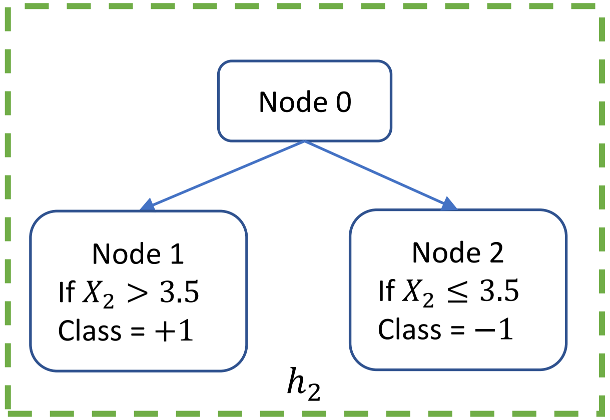
$$H(\mathbf{x}) = \text{sign}(1.25) = 1$$

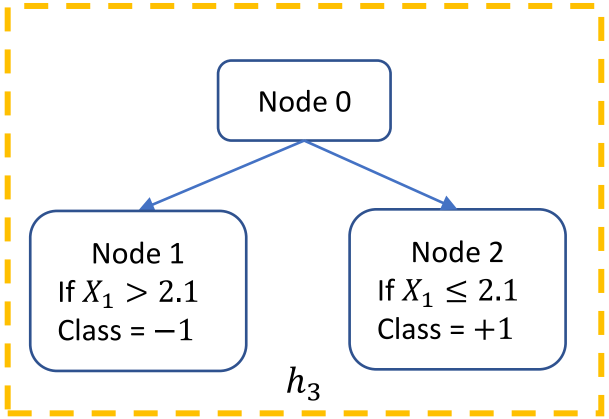
$$\boxed{\textit{The prediction for } \mathbf{x} = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \textit{ is class 1}}$$

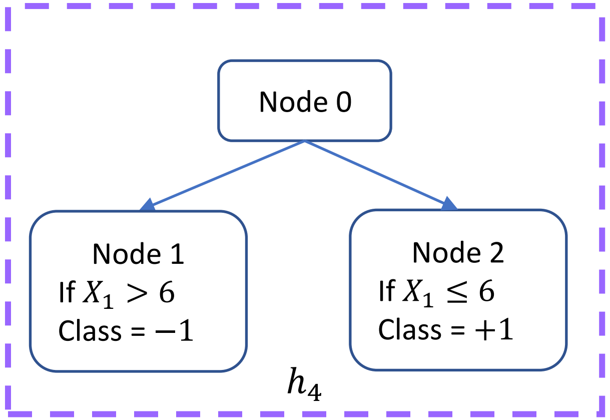
EXERCICE 1 Adaboost

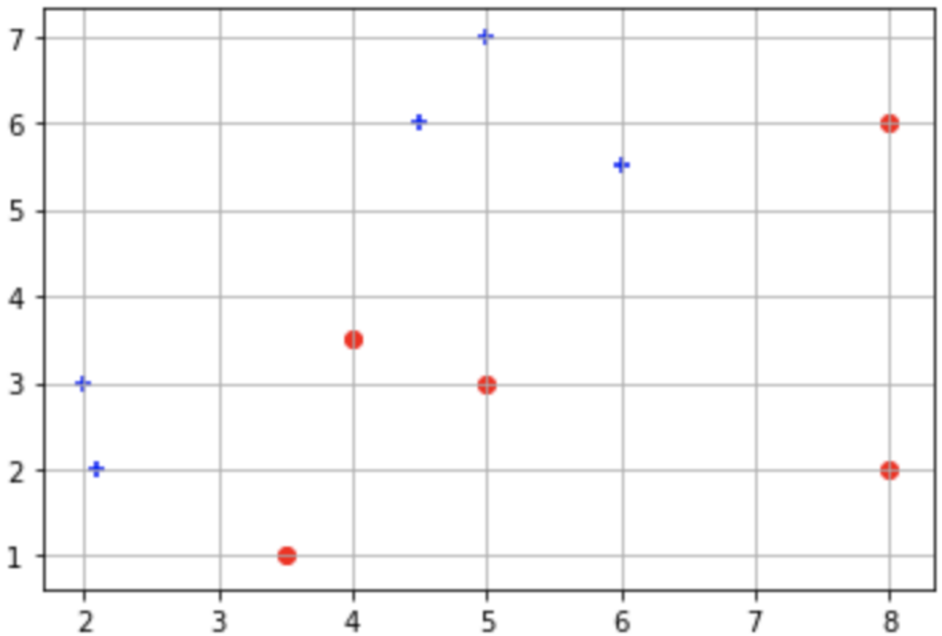
(i) Draw the decision areas in the Adaboost classifiers.











EXERCICE 1

(j) Draw the decision areas in the Adaboost classifier