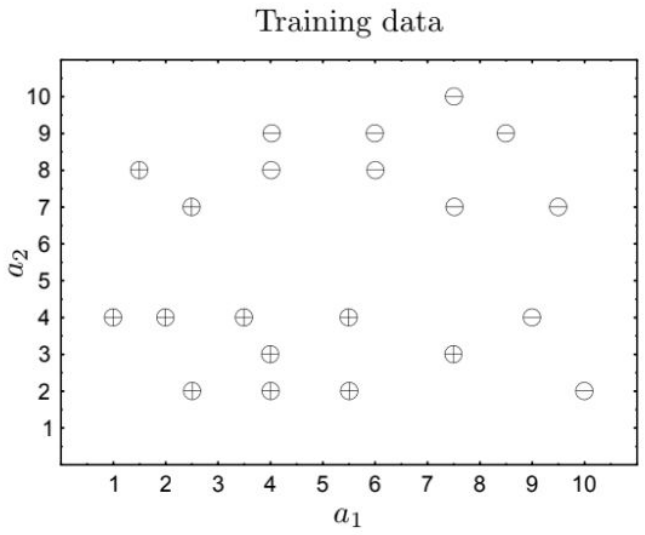
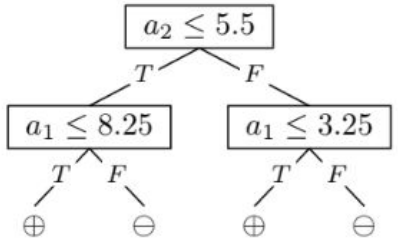
[R] Exercise 1

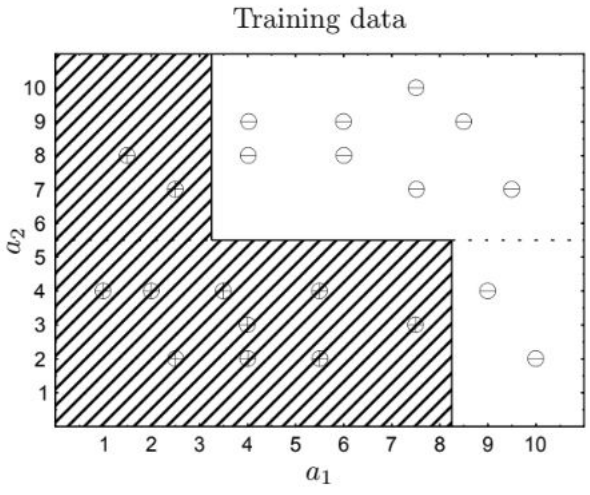
Consider a data set with two numeric attributes a1 and a2 and one nominal target attribute c with two possible values:
⊕ and -
The training examples are shown in the figure below



1. Find a decision tree that classifies all training examples correctly.
2. Draw the decision surface of this tree on the figure.

Solution:





[R] Exercise 2

1. Show a decision tree that could be learned assuming it gets the following examples:

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	sunny	warm	normal	strong	warm	same	yes
2	sunny	warm	high	strong	warm	same	yes
3	rainy	cold	high	strong	warm	change	no
4	sunny	warm	high	strong	cool	change	yes

2. Add the following example and show how a decision tree would be induced using Information Gain for these 5 examples. Continue until all leaf nodes are homogenous

5	sunny	warm	normal	weak	warm	same	no

Starting point:

Attribute	Values	+	−	Entropy	IG
Sky	Sunny	3	1	0.811	0.32
	Rainy	0	1	0.000	
AirTemp	Warm				
	Cold				
Humidity	Normal	1	1	1.000	0.02
	High	2	1	0.918	
Wind	Strong				
	Weak				
Water	Warm				
	Cool				
Forecast	Same				
	Change				

Solution:

1. Sky : sunny → yes ; rainy → no
   Alternative: AirTemp : warm → yes ; cold → no
2. S = [3+,2-], entropy(S) = 0.97

Attribute	Values	+	−	Entropy	IG
Sky	Sunny	3	1	0.811	$0.97 - \frac{4}{5} \cdot 0.811 - \frac{1}{5} \cdot 0 = 0.32$
	Rainy	0	1	0.000	
AirTemp	Warm	3	1	0.811	0.32
	Cold	0	1	0.000	
Humidity	Normal	1	1	1.000	0.02
	High	2	1	0.918	
Wind	Strong	3	1	0.811	0.32
	Weak	0	1	0.000	
Water	Warm	2	2	1.000	0.17
	Cool	1	0	0.000	
Forecast	Same	2	1	0.918	0.02
	Change	1	1	1.000	

Three attributes have equal information gain: Sky, AirTemp, and Wind. Any of them can be selected for the root node of the decision tree.

Assume Sky is selected. The right node corresponding to Sky = rainy is pure, i.e. it contains examples of one class only (−). It becomes a leaf node.

Computing the information gain for the left node (Sky = sunny) yields the following values:

Attribute	IG
$AirTemp$	0.00
Hum	0.31
$Wind$	0.81
$Water$	0.12
$Forecast$	0.12

Wind has the highest IG. Splitting on Wind results in two pure leaves, and the algorithm terminates.