

Série de TD N°5 (Corrigé)

Arbres de Décision

Exercice 1 :

Class P: Customer Satisfaction = “yes”

Class N: Customer Satisfaction = “no”

$$I(p, n) = I(9, 6) = -9/15 \log(9/15) - 6/15 \log(6/15) = 0.971$$

Attribut A1= “**Battery Life**”:

S1 = “high”: $p_1=5$ $n_1=0$ $I(p_1, n_1)=0$

S2 = “medium”: $p_2=2$ $n_2=3$ $I(p_2, n_2) = 0.9710$

S3 = “low”: $p_3=2$ $n_3=3$ $I(p_3, n_3) = 0.9710$

$E(A1) = 5/15 * I(p_1, n_1) + 5/15 * I(p_2, n_2) + 5/15 * I(p_3, n_3) = 0.6473$

$Gain(A1) = I(p, n) - E(A1) = 0.3237$

Attribut A2= “**Memory**”:

S1 = “≤4”: $p_1=3$ $n_1=4$ $I(p_1, n_1) = 0.9852$

S2 = “>4”: $p_2=6$ $n_2=2$ $I(p_2, n_2) = 0.8113$

$E(A2) = 7/15 * I(p_1, n_1) + 8/15 * I(p_2, n_2) = 0.8925$

$Gain(A2) = I(p, n) - E(A2) = 0.0785$

Attribut A3= “**Price**”:

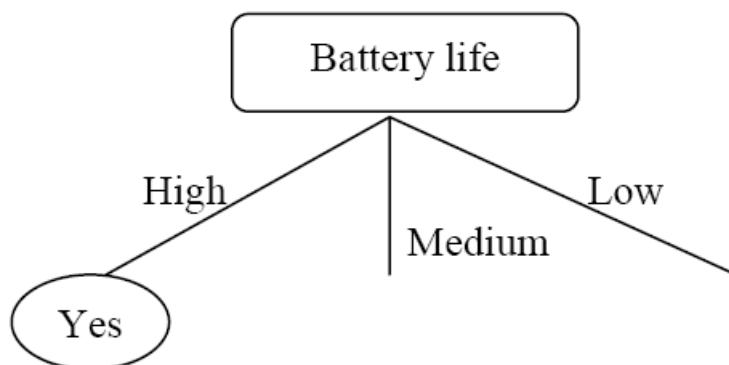
S1 = “≤150”: $p_1=3$ $n_1=4$ $I(p_1, n_1) = 0.9852$

S2 = “>150”: $p_2=6$ $n_2=2$ $I(p_2, n_2) = 0.8113$

$E(A3) = 7/15 * I(p_1, n_1) + 8/15 * I(p_2, n_2) = 0.8925$

$Gain(A3) = I(p, n) - E(A3) = 0.0785$

Puisque A1 = “**Battery Life**” possède le meilleur gain, il est sélectionné comme racine de l’arbre. Pour la valeur “high” tous les exemples appartiennent à la même classe (“yes”). Nous avons:



Pour la valeur “Medium”:

$$I(p, n) = I(2, 3) = 0.9710$$

Attribut A2= **Memory**:

S1 = " ≤ 4 ": $p1=1$ $n1=1$ $I(p1, n1)= 1$

S2 = " >4 ": $p2=1$ $n2=2$ $I(p2, n2)= 0.9183$

$E(A2)=2/5 * I(p1, n1)+ 3/5 * I(p2, n2) =0.9510$

$Gain(A2)= I(p, n)-E(A2)= 0.0200$

Attribut A3= **Price**:

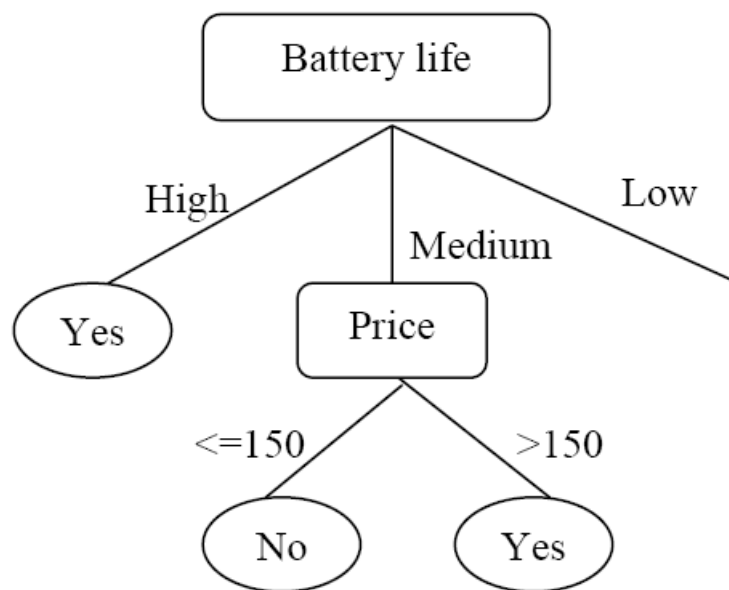
S1 = " ≤ 150 ": $p1=0$ $n1=3$ $I(p1, n1)= 0$

S2 = " >150 ": $p2=2$ $n2=0$ $I(p2, n2)= 0$

$E(A3)=3/5 * I(p1, n1)+ 2/5 * I(p2, n2) =0$

$Gain(A3)=I(p, n)-E(A3)= 0.9710$

$A3 > A2$, donc A3= **Price** est sélectionné. Pour " ≤ 150 " tous les exemples appartiennent à la même classe "no". Pour " >150 " tous les exemples appartiennent à la classe "yes". Nous avons:



Pour la valeur **"Low"**:

$I(p, n)=I(2,3)= 0.9710$

Attribut A2= **Memory**:

S1 = " ≤ 4 ": $p1=0$ $n1=3$ $I(p1, n1)= 0$

S2 = " >4 ": $p2=2$ $n2=0$ $I(p2, n2)= 0.9183$

$E(A2)=3/5 * I(p1, n1)+ 2/5 * I(p2, n2) =0$

$Gain(A2)= I(p, n)-E(A2)= 0.9710$

Attribut A3= **Price**:

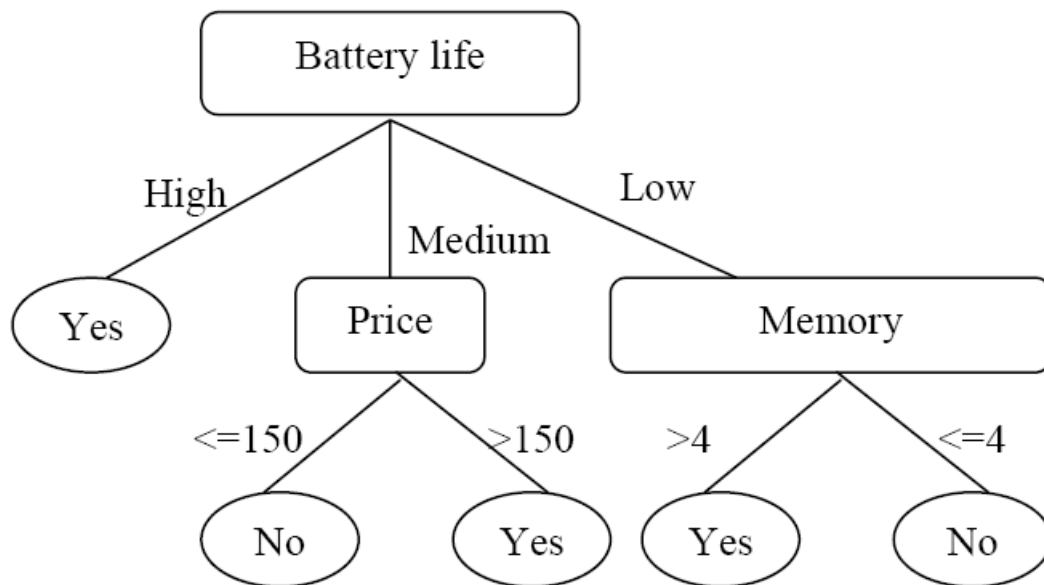
S1 = " ≤ 150 ": $p1=1$ $n1=1$ $I(p1, n1)= 1$

S2 = " >150 ": $p2=1$ $n2=2$ $I(p2, n2)= 0.9183$

$E(A3)=2/5 * I(p1, n1)+ 3/5 * I(p2, n2) =0.9510$

$Gain(A3)=I(p, n)-E(A3)= 0.0200$

$A2 > A3$, donc A2= **Memory** est sélectionné. Pour " >4 " tous les exemples appartiennent à la même classe "yes". Pour " ≤ 4 " tous les exemples appartiennent à la classe "no". Nous avons l'arbre final:



Exercise 2 :

C1 : Internet=OUI

C2 : Internet=NON

Entropie = $-p(c1) \log(p(c1)) - p(c2) \log(p(c2))$

Gain = Entropie - $n1/n [p(c1) \log(p(c1))] - n2/n [p(c2) \log(p(c2))]$

n= 8 Entropie= 0.954434003

C1= 3
C2= 5

Montant	Internet=Oui	Internet=Non	Somme	ni/n	Entropie	Gain
P1 (élevé)	0	2	2	0.25	0	0.26571213
P2 (moyen)	2	1	3	0.38	0.34436094	
P3 (faible)	1	2	3	0.38	0.34436094	

Age	Internet=Oui	Internet=Non	Somme	ni/n	Entropie	Gain
P1 (jeune)	1	0	1	0.13	0	0.454434
P2 (moyen)	2	2	4	0.50	0.5	
P3 (âgé)	0	3	3	0.38	0	

Résidence	Internet=Oui	Internet=Non	Somme	ni/n	Entropie	Gain
P1 (ville)	1	2	3	0.38	0.34436094	0.01571213
P2 (village)	1	1	2	0.25	0.25	
P3 (bourg)	2	1	3	0.38	0.34436094	

Etudes	Internet=Oui	Internet=Non	Somme	ni/n	Entropie	Gain
P1 (Oui)	3	2	5	0.63	0.60684412	0.34758988
P2 (Non)	0	3	3	0.38	0	

Attribut choisi comme racine: **Age**

Client	Age	Montant	Résidence	Etudes	Internet
3	âgé	faible	bourg	non	non
6	âgé	élevé	ville	oui	non
7	âgé	moyen	ville	oui	non
5	jeune	moyen	ville	oui	oui
1	moyen	moyen	village	oui	oui
2	moyen	élevé	bourg	non	non
4	moyen	faible	bourg	oui	oui
8	moyen	faible	village	non	non

Branche: Age= âgé	Classe=	Internet=NON
Branche: Age= jeune	Classe=	Internet=OUI
Branche: Age=moyen	Classe=	Internet=???

Client	Montant	Résidence	Etudes	Internet	n=	
1	moyen	village	oui	oui	C1=	4
2	élevé	bourg	non	non	C2=	2
4	faible	bourg	oui	oui	E=	2
8	faible	village	non	non		1

Montant	Internet=Oui	Internet=Non	Somme	ni/n	Entropie	Gain
P1 (élevé)	0	1	1	0.25	0	0.5
P2 (moyen)	1	0	1	0.25	0	
P3 (faible)	1	1	2	0.50	0.5	

Résidence	Internet=Oui	Internet=Non	Somme	ni/n	Entropie	Gain
P1 (ville)	0	0	0	0.00	0	0
P2 (village)	1	1	2	0.50	0.5	
P3 (bourg)	1	1	2	0.50	0.5	

Etudes	Internet=Oui	Internet=Non	Somme	ni/n	Entropie	Gain
P1 (Oui)	2	0	2	0.50	0	1
P2 (Non)	0	2	2	0.50	0	

Attribut choisi pour le branchement: **Etudes**

Branche: Etudes= oui	Classe=	Internet=OUI
Branche: Etudes= non	Classe=	Internet=NON

