Esercizio 1

Generare l'albero di decisoone per il seguente dataset (la classificazione c(x) è rappresentata dal valore dell'ultima colonna)

Memory	Battery life	Price	Customer satisfaction
<=4	long	<=150	yes
>4	long	>150	yes
>4	long	<=150	yes
<=4	long	>150	yes
>4	long	>150	yes
>4	low	>150	yes
<=4	low	>150	no
<=4	low	>150	no
>4	low	<=150	yes
<=4	low	<=150	no
<=4	medium	<=150	no
>4	medium	<=150	no
<=4	medium	>150	yes
>4	medium	>150	yes
>4	medium	<=150	no

Esercizio 2

Classificate con Naive Bayes l'esempio <single,light,one>
Dato il seguente dataset (healthy e Virulent sono i due valori di classificazione):

Healthy:
- <single,dark,one>
- <single,light,two>
- <double,light,one>

Virulent:
- <single,dark,two>
- <double,dark,one>
- <double,light,two>

Soluzioni

Esercizio 1

I=-9/15log(9/15) -6/15lg(6/15) = 0.971

Analisi del guadagno informativo dei vari attributi (pi e ni indicano il numero di esempi positivi e negativi)

Attribute A_1= "Battery life":

S_1 ="high":
p_1=5 n_1=0
$I(p_1, n_1)$=0

S_2 ="medium":
p_2=2 n_2=3
$I(p_2, n_2)$= 0.9710

S_3 ="low":
p_3=2 n_3=3
$I(p_3, n_3)$= 0.9710

$E(A_1)$=5/15 *$I(p_1, n_1)$+ 5/15*$I(p_2, n_2)$+ 5/15 *$I(p_3, n_3)$= 0.6473

$Gain(A_1)$= I(p, n)-$E(A_1)$= 0.3237

Attribute A_2= "**Memory**":

S_1 ="<=4":
p_1=3 n_1=4
$I(p_1, n_1)$= 0.9852

S_2 =">4":
p_2=6 n_2=2
$I(p_2, n_2)$= 0.8113

$E(A_2)$=7/15 *$I(p_1, n_1)$+ 8/15*$I(p_2, n_2)$ =0.8925

$Gain(A_2)$= I(p, n)-$E(A_2)$= 0.0785

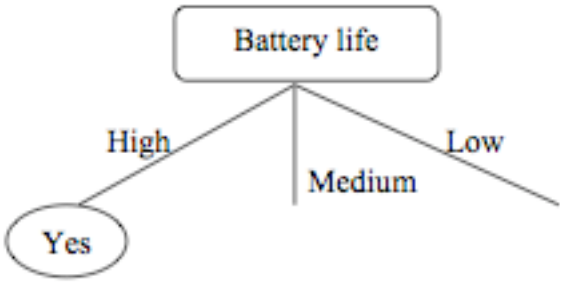
Attribute A_3= "**Price**":

S_1 ="<=150":
p_1=3 n_1=4
$I(p_1, n_1)$= 0.9852

S_2 =">150":
p_2=6 n_2=2
$I(p_2, n_2)$= 0.8113

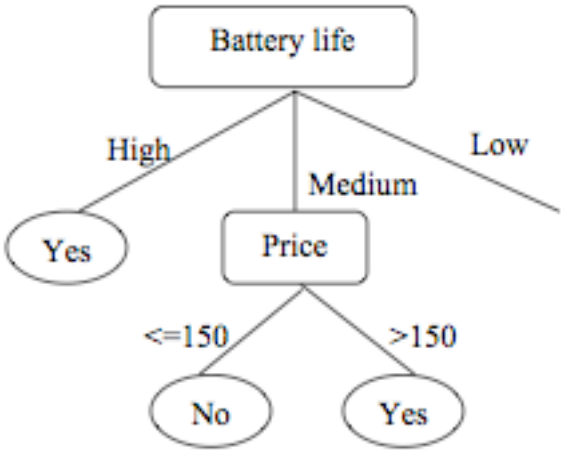
$E(A_3$=7/15 *$I(p_1, n_1)$+ 8/15*$I(p_2, n_2)$ =0.8925

Gain(A_3)=(p, n)-E(A_3)= 0.0785



Seguiamo il ramo Battery_life= Medium

Attribute A_2= "**Memory**": S_1 ="<=4": p_1=1 n_1=1 $I(p_1, n_1)$= 1 S_2 =">4": p_2=1 n_2=2 $I(p_2, n_2)$= 0.9183 $E(A_2)$=2/5 *$I(p_1, n_1)$+ 3/5*$I(p_2, n_2)$ =0.9510 Gain(A_2)= I(p, n)-$E(A_2)$= 0.0200
Attribute A_3= "**Price**": S_1 ="<=150": p_1=0 n_1=3 $I(p_1, n_1)$= 0 S_2 =">150": p_2=2 n_2=0 $I(p_2, n_2)$= 0 $E(A_3)$=3/5 *$I(p_1, n_1)$+ 2/5*$I(p_2, n_2)$ =0 Gain(A_3)=(p, n)-$E(A_3)$= 0.9710



Seguiamo infine il ramo Battery Life = Low

Attribute A_2= "**Memory**":

S_1 ="<=4":
p_1=0 n_1=3
$I(p_1, n_1)$= 0

S_2 =">4":
p_2=2 n_2=0
$I(p_2, n_2)$= 0.9183

$E(A_2)$=3/5 *$I(p_1, n_1)$+ 2/5*$I(p_2, n_2)$ =0

Gain(A_2)= I(p, n)-$E(A_2)$= 0.9710

Attribute A_3= "**Price**":

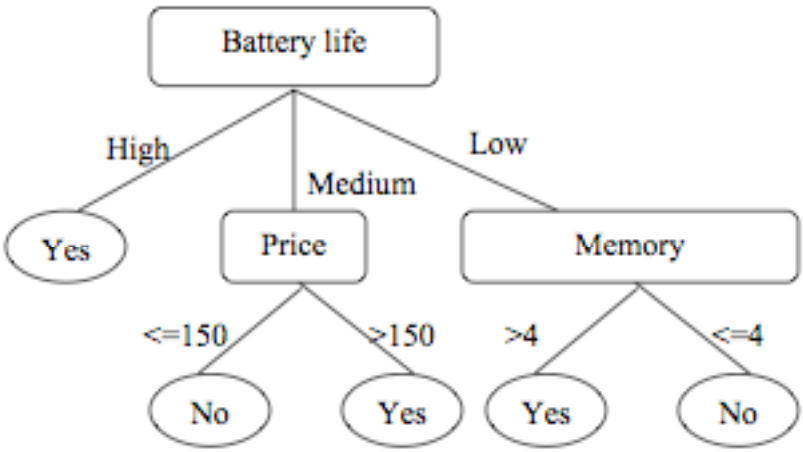
S_1 ="<=150":
p_1=1 n_1=1
$I(p_1, n_1)$= 1

S_2 =">150":
p_2=1 n_2=2
$I(p_2, n_2)$= 0.9183

$E(A_3)$=2/5 *$I(p_1, n_1)$+ 3/5*$I(p_2, n_2)$ =0.9510

Gain(A_3)=(p, n)-$E(A_3)$= 0.0200

L'albero completo è



Esercizio 2

$$\text{argmax}_H P(single|H)P(light|H)P(one|H)P(H)$$

$P(single|Healthy) = 2/3$, $P(double|Healthy) = 1/3$,
$P(dark|Healthy) = 1/3$, $P(light|Healthy) = 2/3$,
$P(one|Healthy) = 2/3$, $P(two|Healthy) = 1/3$
$P(Healthy) = 1/2$

$P(single|Virulent) = 1/3$, $P(double|Virulent) = 2/3$,
$P(dark|Virulent) = 2/3$, $P(light|Virulent) = 1/3$,
$P(one|Virulent) = 1/3$, $P(two|Virulent) = 2/3$
$P(Virulent) = 1/2$

$$\begin{aligned} \text{Likelyhood(Healty)} &= P(\text{single}|\text{Healthy})P(\text{light}|\text{Healthy})P(\text{one}|\text{Healthy})P(\text{Healthy}) \\ &= 2/3 \cdot 2/3 \cdot 2/3 \cdot 1/2 = 4/27 \\ \text{Likelyhood(Virulent)} &= 1/3 \cdot 1/3 \cdot 1/3 \cdot 1/2 = 1/54 \end{aligned}$$

Quindi la classificazione più probabile è Healthy.