Soluzione Esercizi del Corso di Apprendimento Automatico

Prima Parte (esercizi aggiuntivi)

Esercizio 2

Considerare il problema di apprendere il concetto *EnjoySport* visto a lezione, dove però si considerano i seguenti valori per gli attributi

$$\begin{aligned}
\text{Sky} &\rightarrow \{Sunny, Cloudy\};\\
\text{AirTemp} &\rightarrow \{Hot, Cold\};\\
\text{Humidity} &\rightarrow \{Low, Normal, High\};\\
\text{Wind} &\rightarrow \{Weak, Strong\};\\
\text{Water} &\rightarrow \{Hot, Warm, Cold, VeryCold\};\\
\text{Forecast} &\rightarrow \{Same, Change\};
\end{aligned}$$

a) definire un insieme minimo di esempi che, se presentato all'algoritmo *Candidate-Elimination*, produce il seguente Version Space:

$$\begin{array}{lccccccl}
S:\{< & Sunny & ? & High & ? & Warm & Change & >\}\\
G:\{< & Sunny & ? & ? & ? & ? & ? & >,\\
< & ? & ? & High & ? & ? & ? & >\}
\end{array}$$

Risposta:

Una possibile sequenza di esempi è la seguente:

$$\begin{array}{lcccccccl}
1. & (< S & H & H & W & W & C & >, & Yes)\\
2. & (< S & C & H & S & W & C & >, & Yes)\\
3. & (< C & C & N & S & W & C & >, & No)
\end{array}$$

b) dire quante ipotesi contiene il Version Space di sopra;

Risposta:

12. Si può dare la risposta sia enumerando esplicitamente le ipotesi che appartengono al Version Space, sia osservando che gli attributi che cambiano fra l'ipotesi in S e quelle in G sono 3. Avendo ogni attributo la possibilità di assumere solo 2 valori (un valore specifico, oppure ?), il numero totale di ipotesi nel Version Space è dato dalle ipotesi comprese tra l'ipotesi in S e ognuna delle ipotesi in G. Se si considerano le ipotesi in G separatamente, allora ognuna, in combinazione con l'ipotesi in S, determina $2^3 = 8$ ipotesi. Di queste, tuttavia, una parte è condivisa. Infatti, tutte e due le ipotesi in G hanno in comune, fissato il primo argomento ad "S" ed il terzo ad "H", la variazione sugli ultimi due attibuti ($2^2 = 4$ ipotesi). Quindi, in totale, il Version Space contiene $8 + 8 - 4 = 12$ ipotesi.

c) dare un insieme minimale di esempi addizionali in modo da far convergere il Version Space di sopra al seguente Version Space

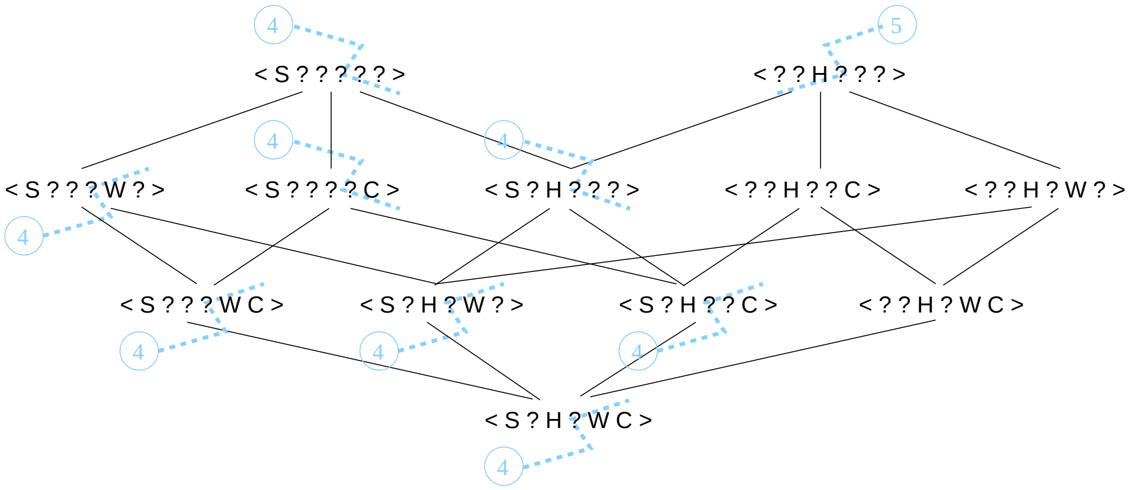
$$\begin{array}{ll} S: \{< & ? \quad ? \quad High \quad ? \quad Warm \quad Change \quad >\} \\ G: \{< & ? \quad ? \quad High \quad ? \quad ? \quad Change \quad >, \\ \phantom{G: \{}< & ? \quad ? \quad High \quad ? \quad Warm \quad ? \quad >\} \end{array}$$

Risposta:

una possibile sequenza di esempi è la seguente:

$$\begin{array}{l} 4.\ (< C \quad H \quad H \quad S \quad W \quad C \quad >, \quad Yes) \\ 5.\ (< C \quad H \quad H \quad S \quad H \quad S \quad >, \quad No) \end{array}$$

Di seguito viene mostrato come il Version Space cambia alla presentazione degli esempi 4 e 5.



Esercizio 3

b) dato il seguente insieme di apprendimento

Esempio	*Target*	A_1 (5 *val.*)	A_2 (3 *val.*)	A_3 (4 *val.*)	A_4 (4 *val.*)
1	$+$	v_1	v_1	v_4	v_1
2	$+$	v_1	v_3	v_2	v_2
3	$-$	v_2	v_1	v_1	v_1
4	$-$	v_2	v_2	v_4	v_4
5	$-$	v_3	v_2	v_4	v_3
6	$+$	v_3	v_1	v_2	v_2
7	$+$	v_4	v_3	v_1	v_1
8	$-$	v_4	v_2	v_4	v_3
9	$+$	v_5	v_1	v_4	v_1
10	$+$	v_5	v_3	v_2	v_2

mostrare come ID3 (con GainRatio(S,A)) costruisce l'albero di decisione corrispondente. Per ogni attributo calcolare $GainRatio(S, A)$ e giustificare la scelta dell'attributo utilizzato ad ogni nodo interno.

Risposta:

Entropia: 0.970951

Attributo A_1

valore 1:[+,-] 2 0, valore 2:[+,-] 0 2, valore 3:[+,-] 1 1, valore 4:[+,-] 1 1,

valore 5:[+,-] 2 0

Gain: 0.570951 Split: 2.321928 Ratio: 0.245895

Attributo A_2

valore 1:[+,-] 3 1, valore 2:[+,-] 0 3, valore 3:[+,-] 3 0

Gain: 0.646439 Split: 1.570951 Ratio: 0.411496

Attributo A_3

valore 1:[+,-] 1 1, valore 2:[+,-] 3 0, valore 3:[+,-] 0 0, valore 4:[+,-] 2 3

Gain: 0.285475 Split: 1.485475 Ratio: 0.192178

Attributo A_4

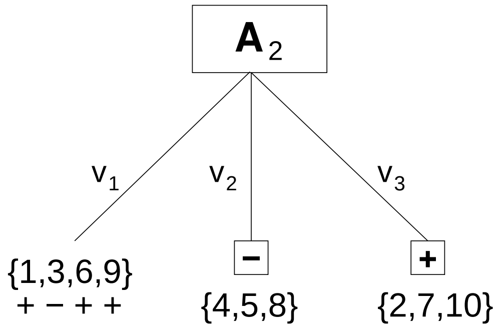
valore 1:[+,-] 3 1, valore 2:[+,-] 3 0, valore 3:[+,-] 0 2, valore 4:[+,-] 0 1

Gain: 0.646439 Split: 1.846439 Ratio: 0.350101

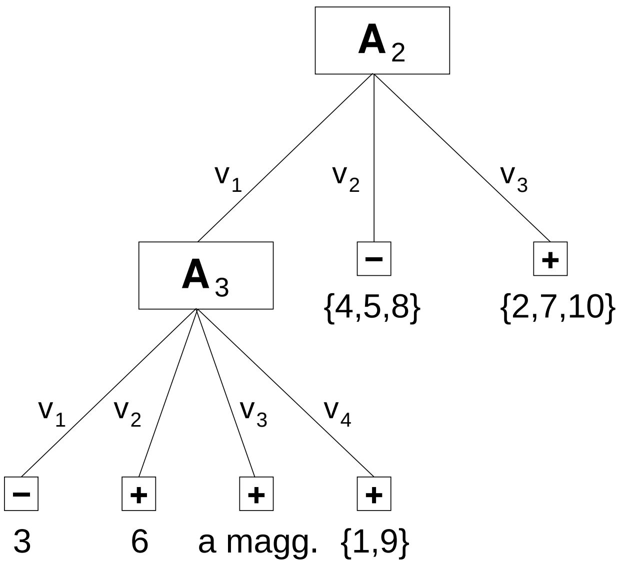
Quindi si sceglie A_2 per la radice:

S_1 = {1[+],3[-],6[+],9[+]}, S_2 = {4[-],5[-],8[-]}, S_3 = {2[+],7[+],10[+]}

e l'albero parziale è:



In particolare si deve ancora elaborare l'insieme S_1 = {1[+],3[-],6[+],9[+]}, che risulta essere classificato correttamente da A_1, o A_3. Supponendo di scegliere l'attributo che assume il numero minimo di valori distinti, cioé A_3, otteniamo l'albero finale:



c) Mostrare cosa succede se si aggiunge l'esempio *(4,2,2,1)* con target **-**

Risposta:
Entropia: 0.994030

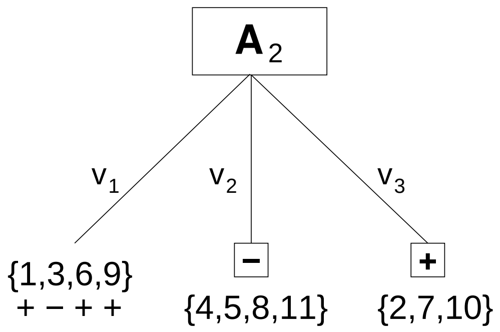
Attributo A_1
valore 1:[+,-] 2 0, valore 2:[+,-] 0 2, valore 3:[+,-] 1 1, valore 4:[+,-] 1 2, valore 5:[+,-] 2 0
Gain: 0.561768 Split: 2.299896 Ratio: 0.244258

Attributo A_2
valore 1:[+,-] 3 1, valore 2:[+,-] 0 4, valore 3:[+,-] 3 0
Gain: 0.699020 Split: 1.572624 Ratio: 0.444493

Attributo A_3
valore 1:[+,-] 1 1, valore 2:[+,-] 3 1, valore 3:[+,-] 0 0, valore 4:[+,-] 2 3
Gain: 0.075861 Split: 1.494919 Ratio: 0.050746

Attributo A_4
valore 1:[+,-] 3 2, valore 2:[+,-] 3 0, valore 3:[+,-] 0 2, valore 4:[+,-] 0 1
Gain: 0.552689 Split: 1.789929 Ratio: 0.308777

Quindi si sceglie di nuovo A_2 per la radice:
$S_1 = \{1[+],3[-],6[+],9[+]\}$, $S_2 = \{4[-],5[-],8[-],11[-]\}$, $S_3 = \{2[+],7[+],10[+]\}$
e l'albero parziale è:

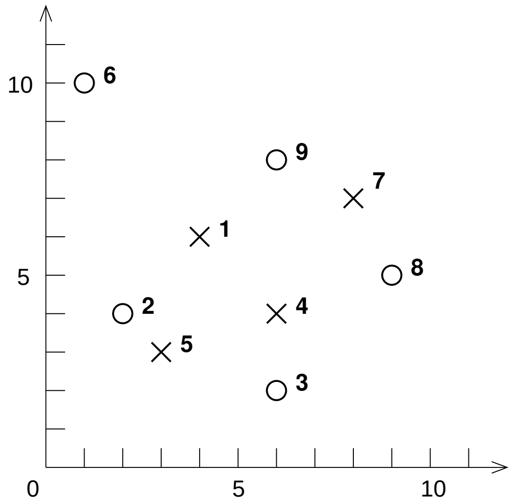


e poiché non cambia l'insieme S_1, si ottiene di nuovo l'albero finale precedente.

Esercizio 3

b) sia dato uno spazio delle istanze costituito da punti sul piano a coordinate intere nell'intervallo [1,10], e uno spazio delle ipotesi dato da rettangoli con lati paralleli agli assi e vertici a coordinate intere (sempre nell'intervallo [1,10]), come visto a lezione.

Dato il seguente insieme di esempi



(dove le ics costituiscono istanze positive e i tondini istanze negative).

Calcolare i valori di *Ratio(S,A)* e *GainRatio(S,A)* per i due attributi (cioè la coordinata x e la coordinata y). Quindi usare *Ratio(S,A)* per mostrare come ID3 costruisce l'albero di decisione.

Risposta:

Entropia: 0.991076

Attributo A_x

valore 1:[+,-] 0 1, valore 2:[+,-] 0 1, valore 3:[+,-] 1 0, valore 4:[+,-] 1 0,

valore 5:[+,-] 0 0, valore 6:[+,-] 1 2, valore 7:[+,-] 0 0, valore 8:[+,-] 1 0,

valore 9:[+,-] 0 1, valore 10:[+,-] 0 0

Gain: 0.684977 Split: 2.641604 Ratio: 0.259304

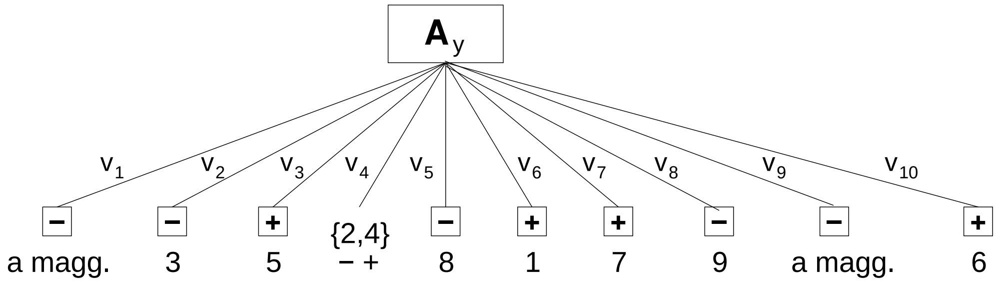
Attributo A_y

valore 1:[+,-] 0 0, valore 2:[+,-] 0 1, valore 3:[+,-] 1 0, valore 4:[+,-] 1 1,

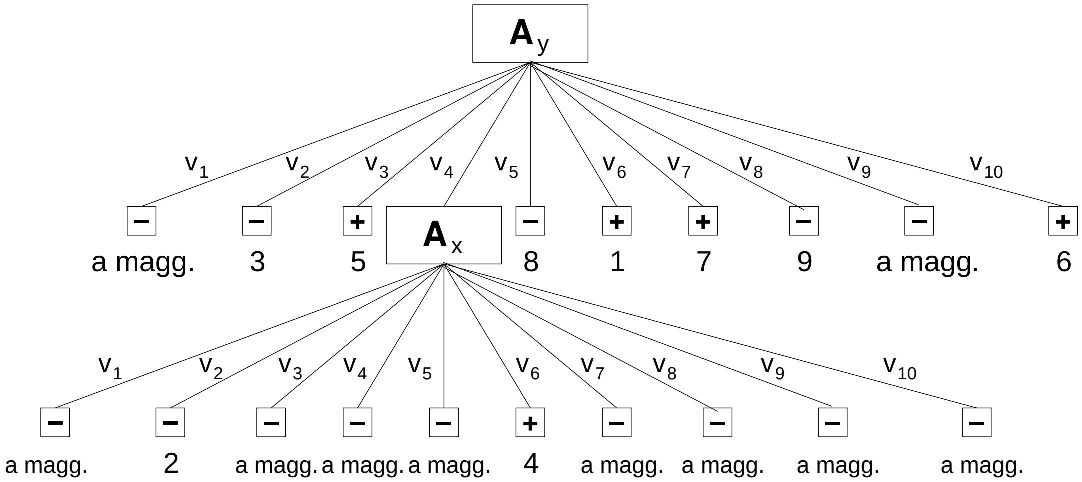
valore 5:[+,-] 0 1, valore 6:[+,-] 1 0, valore 7:[+,-] 1 0, valore 8:[+,-] 0 1, valore 9:[+,-] 0 0, valore 10:[+,-] 0 1

Gain: 0.768854 Split: 2.947703 Ratio: 0.260832

Quindi si sceglie A_y per la radice:



e S_4 = {2[-],4[+]} è infine classificato correttamente da A_x:



dove le decisioni a maggioranza sono prese risalendo alla radice a causa della parità riscontrata al livello del nodo.

Esercizio 3

b) dato il seguente insieme di apprendimento

$Esempio$	$Target$	Out	$Temp$	Hum	$Wind$
1	$-$	S	H	H	W
2	$-$	S	H	H	S
3	$+$	O	_	H	W
4	$+$	R	M	H	W
5	$+$	_	C	N	_
6	$-$	R	C	N	S
7	$+$	O	C	N	S
8	$-$	S	M	H	W
9	$+$	S	C	N	W
10	$+$	R	M	N	W
11	$+$	S	M	N	S
12	$+$	_	M	H	S
13	$+$	O	H	N	W
14	$-$	R	M	_	S

dove "_" indica un dato mancante. Mostrare come ID3 (con Gain(S,A)) costruisce l'albero di decisione corrispondente. Per ogni attributo calcolare $Gain(S, A)$ e giustificare la scelta dell'attributo utilizzato ad ogni nodo interno. Per il trattamento dei dati mancanti utilizzare l'approccio del valore più frequente.

Risposta:

Di seguito viene riportato il numero di occorrenza dei valori dei vari attributi:

Out	$Temp$	Hum	$Wind$
$\#``S" = 5$	$\#``H" = 3$	$\#``H" = 6$	$\#``W" = 7$
$\#``O" = 3$	$\#``M" = 6$	$\#``N" = 7$	$\#``S" = 6$
$\#``R" = 4$	$\#``C" = 4$		

Scegliendo il valore con il maggiore numero di occorrenze per ogni attributo, l'insieme di apprendimento diventa:

Esempio	*Target*	*Out*	*Temp*	*Hum*	*Wind*
$1'$	$-$	S	H	H	W
$2'$	$-$	S	H	H	S
$3'$	$+$	O	M	H	W
$4'$	$+$	R	M	H	W
$5'$	$+$	S	C	N	W
$6'$	$-$	R	C	N	S
$7'$	$+$	O	C	N	S
$8'$	$-$	S	M	H	W
$9'$	$+$	S	C	N	W
$10'$	$+$	R	M	N	W
$11'$	$+$	S	M	N	S
$12'$	$+$	S	M	H	S
$13'$	$+$	O	H	N	W
$14'$	$-$	R	M	N	S

Entropia: 0.940286

Attributo *Out*

valore S:[+,-] 4 3, valore O:[+,-] 3 0, valore R:[+,-] 2 2

$\boxed{\text{Gain: 0.161958}}$ ← **valore massimo**

Attributo *Temp*

valore H:[+,-] 1 2, valore M:[+,-] 5 2, valore C:[+,-] 3 1

$\boxed{\text{Gain: 0.080154}}$

Attributo *Hum*

valore H:[+,-] 3 3, valore N:[+,-] 6 2

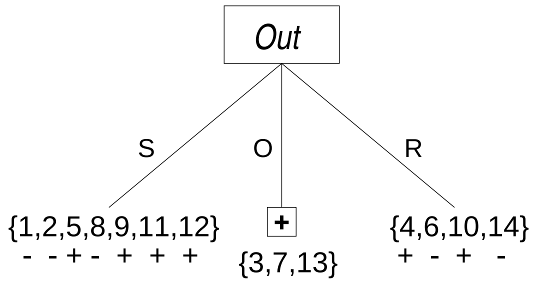
$\boxed{\text{Gain: 0.048127}}$

Attributo *Wind*

valore W:[+,-] 6 2, valore S:[+,-] 3 3

$\boxed{\text{Gain: 0.048127}}$

Quindi si sceglie *Out* come attributo per la radice:



Consideriamo adesso S_S = {1[-],2[-],5[+],8[-],9[+],11[+],12[+]}:

$Esempio$	$Target$	Out	$Temp$	Hum	$Wind$
1	$-$	S	H	H	W
2	$-$	S	H	H	S
5	$+$	_	C	N	_
8	$-$	S	M	H	W
9	$+$	S	C	N	W
11	$+$	S	M	N	S
12	$+$	_	M	H	S

Ricordando che non possiamo più utilizzare l'attributo Out, l'unico attributo che presenta un valore mancante è $Wind$ (esempio 5). Quindi calcoliamo il numero di occorrenze per i valori di $Wind$ relativamente ad S_S:

$Wind$
$\#``W'' = 3$
$\#``S'' = 3$

Poiché si presenta un caso di parità, scegliamo il valore più frequente a livello superiore (radice), cioè "W":

Entropia: 0.985228

Attributo $Temp$

valore H:[+,-] 0 2, valore M:[+,-] 2 1, valore C:[+,-] 2 0

$\boxed{\text{Gain: } 0.591673}$ ← **valore massimo**

Attributo Hum

valore H:[+,-] 1 3, valore N:[+,-] 3 0

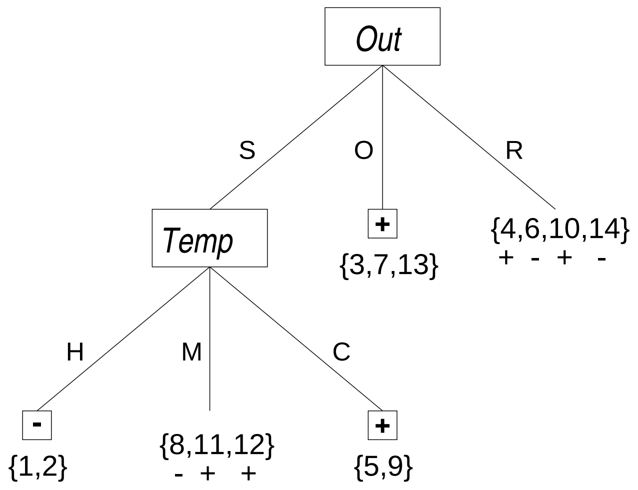
$\boxed{\text{Gain: } 0.521641}$

Attributo $Wind$

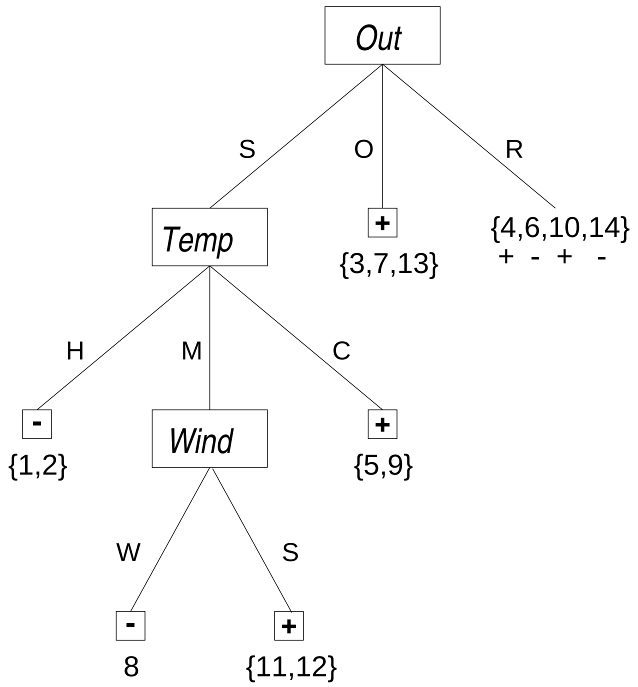
valore W:[+,-] 2 2, valore S:[+,-] 2 1

$\boxed{\text{Gain: } 0.020244}$

Quindi si sceglie $Temp$ come attributo per la radice:



ed infine $S_M = \{8[-],11[+],12[+]\}$ è classificato correttamente (e solo) dall'attributo $Wind$:



Infine consideriamo $S_R = \{4[+],6[-],10[+],14[-]\}$:

$Esempio$	$Target$	Out	$Temp$	Hum	$Wind$
4	$+$	R	M	H	W
6	$-$	R	C	N	S
10	$+$	R	M	N	W
14	$-$	R	M	_	S

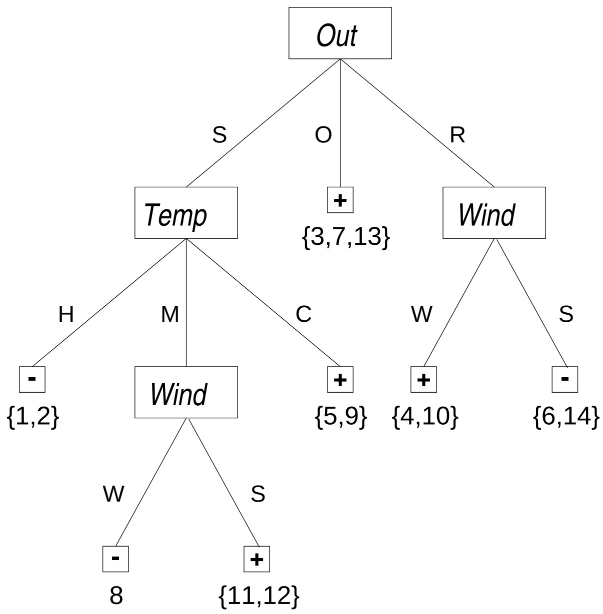
L'unico attributo che presenta un valore mancante è Hum (esempio 14). Quindi calcoliamo il numero di occorrenze per i valori di Hum relativamente ad S_R:

$Wind$
$\#``H'' = 1$
$\#``N'' = 2$

e scegliamo il valore più frequente, cioè "N":

$Esempio$	$Target$	Out	$Temp$	Hum	$Wind$
4	$+$	R	M	H	W
6	$-$	R	C	N	S
10	$+$	R	M	N	W
14	$-$	R	M	N	S

Si vede subito che $Wind$ classifica correttamente S_R:



Esercizio 3

b) dato il seguente insieme di apprendimento

$Esempio$	$Target$	Out	$Temp$	Hum	$Wind$
1	$-$	S	H	H	W
2	$-$	S	H	H	S
3	$+$	O	_	H	W
4	$+$	R	M	H	W
5	$+$	_	C	N	_
6	$-$	R	C	N	S
7	$+$	O	C	N	S
8	$-$	S	M	H	W
9	$+$	S	C	N	W
10	$+$	R	M	N	W
11	$+$	S	M	N	S
12	$+$	_	M	H	S
13	$+$	O	H	N	W
14	$-$	R	M	_	S

dove "_" indica un dato mancante. Mostrare come ID3 (con Gain(S,A)) costruisce l'albero di decisione corrispondente. Per ogni attributo calcolare $Gain(S, A)$ e giustificare la scelta dell'attributo utilizzato ad ogni nodo interno. Per il trattamento dei dati mancanti utilizzare l'approccio del valore più frequente con stesso target.

Risposta:

Di seguito viene riportato il numero di occorrenza dei valori dei vari attributi per esempi con stesso target dell'esempio che presenta il valore mancante:

Out (+)	$Temp$ (+)	Hum (-)	$Wind$ (+)
$\#``S'' = 2$	$\#``H'' = 1$	$\#``H'' = 3$	$\#``W'' = 5$
$\#``O'' = 3$	$\#``M'' = 4$	$\#``N'' = 1$	$\#``S'' = 3$
$\#``R'' = 2$	$\#``C'' = 3$		

Scegliendo il valore con il maggiore numero di occorrenze per ogni attributo, l'insieme di apprendimento diventa:

$Esempio$	$Target$	Out	$Temp$	Hum	$Wind$
$1'$	$-$	S	H	H	W
$2'$	$-$	S	H	H	S
$3'$	$+$	O	M	H	W
$4'$	$+$	R	M	H	W
$5'$	$+$	O	C	N	W
$6'$	$-$	R	C	N	S
$7'$	$+$	O	C	N	S
$8'$	$-$	S	M	H	W
$9'$	$+$	S	C	N	W
$10'$	$+$	R	M	N	W
$11'$	$+$	S	M	N	S
$12'$	$+$	O	M	H	S
$13'$	$+$	O	H	N	W
$14'$	$-$	R	M	H	S

Entropia: 0.940286

Attributo Out

valore S:[+,-] 2 3, valore O:[+,-] 5 0, valore R:[+,-] 2 2

$\boxed{\text{Gain: 0.307804}}$ $\leftarrow$ **valore massimo**

Attributo $Temp$

valore H:[+,-] 1 2, valore M:[+,-] 5 2, valore C:[+,-] 3 1

$\boxed{\text{Gain: 0.080154}}$

Attributo Hum

valore H:[+,-] 3 4, valore N:[+,-] 6 1

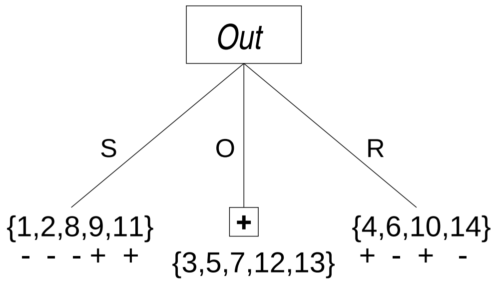
$\boxed{\text{Gain: 0.151835}}$

Attributo $Wind$

valore W:[+,-] 6 2, valore S:[+,-] 3 3

$\boxed{\text{Gain: 0.048127}}$

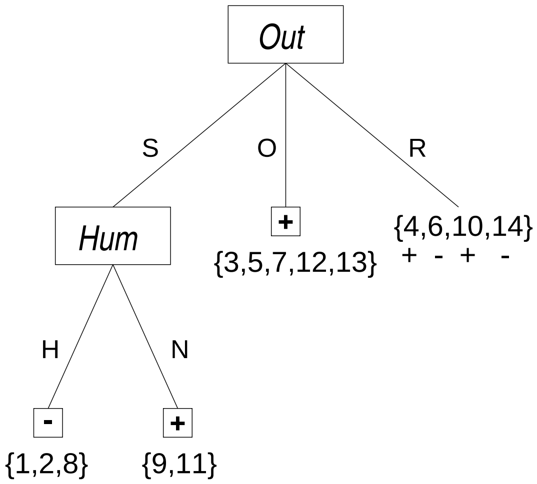
Quindi si sceglie Out come attributo per la radice:



Consideriamo adesso $S_S = \{1[-],2[-],8[-],9[+],11[+]\}$:

$Esempio$	$Target$	Out	$Temp$	Hum	$Wind$
1	$-$	S	H	H	W
2	$-$	S	H	H	S
8	$-$	S	M	H	W
9	$+$	S	C	N	W
11	$+$	S	M	N	S

Si vede subito che (solo) Hum riesce a classificare correttamente S_S:



Infine consideriamo $S_R = \{4[+],6[-],10[+],14[-]\}$:

$Esempio$	$Target$	Out	$Temp$	Hum	$Wind$
4	$+$	R	M	H	W
6	$-$	R	C	N	S
10	$+$	R	M	N	W
14	$-$	R	M	_	S

L'unico attributo che presenta un valore mancante è Hum (esempio 14). Quindi calcoliamo il numero di occorrenze per i valori di Hum relativamente ad esempi in S_R con target -:

$Wind$
$\#``H'' = 0$
$\#``N'' = 1$

e scegliamo il valore più frequente, cioè "N":

$Esempio$	$Target$	Out	$Temp$	Hum	$Wind$
4	$+$	R	M	H	W
6	$-$	R	C	N	S
10	$+$	R	M	N	W
14	$-$	R	M	N	S

Si vede subito che $Wind$ classifica correttamente S_R: