

Soluzione Esercizi del Corso di Sistemi per l'Elaborazione dell'Informazione

Terza Parte

Esercizio 1

b) dato il seguente insieme di apprendimento

<i>Esempio</i>	<i>Target</i>	A_1 (5 val.)	A_2 (3 val.)	A_3 (4 val.)	A_4 (9 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_5
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_6
7	+	v_4	v_3	v_1	v_7
8	-	v_4	v_2	v_4	v_9
9	+	v_5	v_1	v_4	v_1
10	-	v_5	v_3	v_2	v_5

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

c) Mostrare cosa succede se si aggiunge l'esempio $(4,3,2,1)$ con target -

Esercizio 2

b) dato il seguente insieme di apprendimento

<i>Esempio</i>	<i>Target</i>	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 $\text{GainRatio}(S,A)$) costruisce l'albero di decisione corrispondente. Per ogni attributo calcolare $\text{GainRatio}(S,A)$ e giustificare la scelta dell'attributo utilizzato ad ogni nodo interno.

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

Esercizio 3

a) definire in modo completo $GainRatio(S,A)$ e dire perchè può essere preferito a $Gain(S,A)$;

b) dato il seguente insieme di apprendimento

Esempio	Target	A_1 (5 val.)	A_2 (3 val.)	A_3 (4 val.)	A_4 (9 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_5
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_6
7	+	v_4	v_3	v_1	v_9
8	-	v_4	v_2	v_4	v_9
9	-	v_5	v_1	v_4	v_1
10	+	v_5	v_3	v_2	v_5

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.

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

Esercizio 4

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.

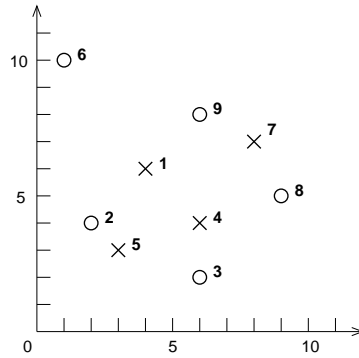
c) Mostrare cosa succede se si aggiunge l'esempio $(4,2,2,1)$ con target -

Esercizio 5

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



Esercizio 6

b) dato il seguente insieme di apprendimento

<i>Esempio</i>	<i>Target</i>	<i>Out</i>	<i>Temp</i>	<i>Hum</i>	<i>Wind</i>
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 $\text{Gain}(S,A)$) costruisce l'albero di decisione corrispondente. Per ogni attributo calcolare $\text{Gain}(S,A)$ 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.