

Compito del Corso di Sistemi di Elaborazione dell'Informazione

Anno Accademico 2003/2004

Appello 16 Dicembre 2003 - Parte Prima

Istruzioni

- Scrivere *Nome, Cognome e Matricola* su **ogni** foglio.
- Scrivere la risposta nello spazio bianco al di sotto della domanda; Non è possibile allegare fogli aggiuntivi, quindi cercate di essere chiari e non prolissi.
- In caso di errori indicate chiaramente quale parte della risposta deve essere considerata; annullate le parti non pertinenti.
- Assicuratevi che non manchi alcun foglio al momento della consegna.

Esercizio 1

- a) Descrivere in modo preciso l'algoritmo di ricerca generico, e dire come si istanzia per ottenere la ricerca Greedy. Dire anche le proprietà di cui gode la ricerca Greedy, motivandole in modo preciso;

- b) Descrivere in modo preciso il risolutore per vincoli sugli alberi (`unify(C)`) e mostrare come si applica al seguente vincolo

$$f(X, g(a, X), h(Q, a), s(s(Q))) = f(h(Q, Z), Y, Z, s(P))$$

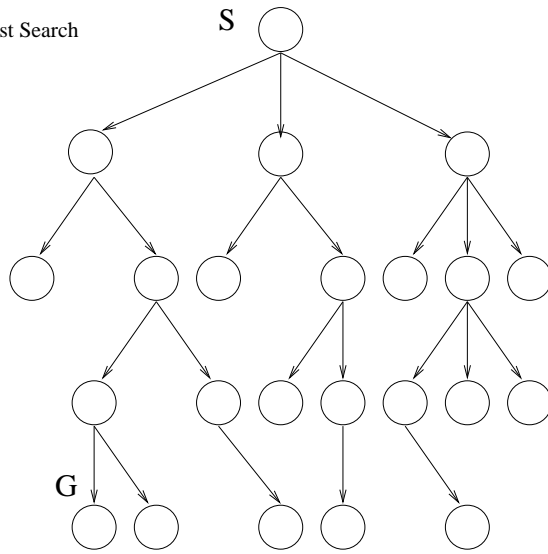
- c) Dare la definizione di errore ideale e di overfitting; dare un esempio pratico di overfitting, definendo uno spazio delle istanze, uno spazio delle ipotesi, un insieme di apprendimento con almeno 5 esempi;

d) Dare la definizione completa dell'algoritmo ID3, inclusa la formula completa del Gain;

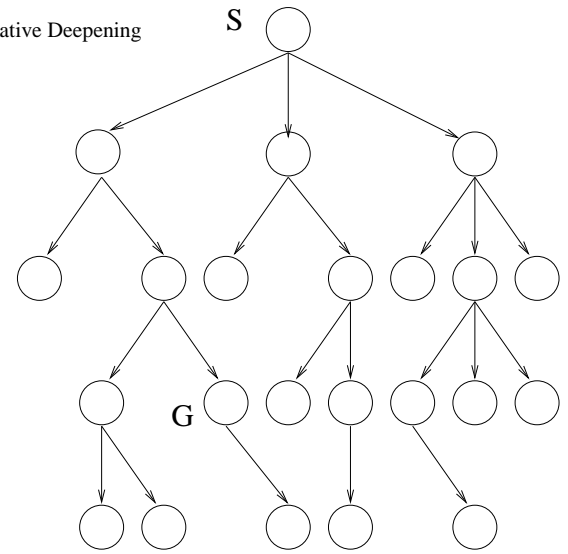
Esercizio 2

Dato l'albero di ricerca della figura allegata, numerare i nodi seguendo l'ordine di espansione delle seguenti strategie: breadth first, depth first, greedy, costo uniforme.

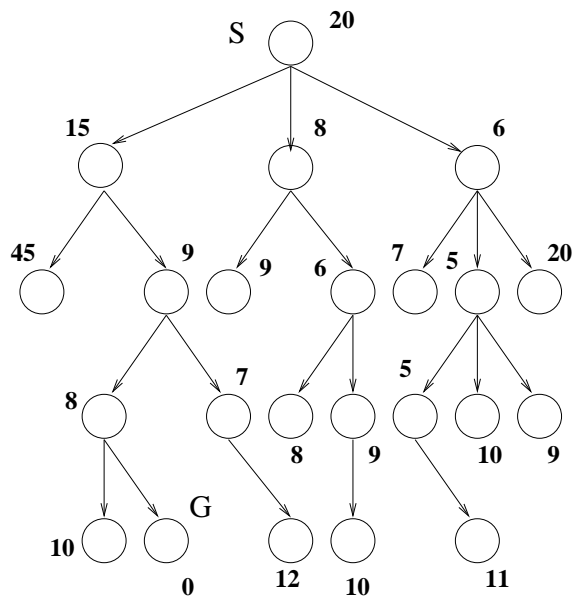
Breadth First Search



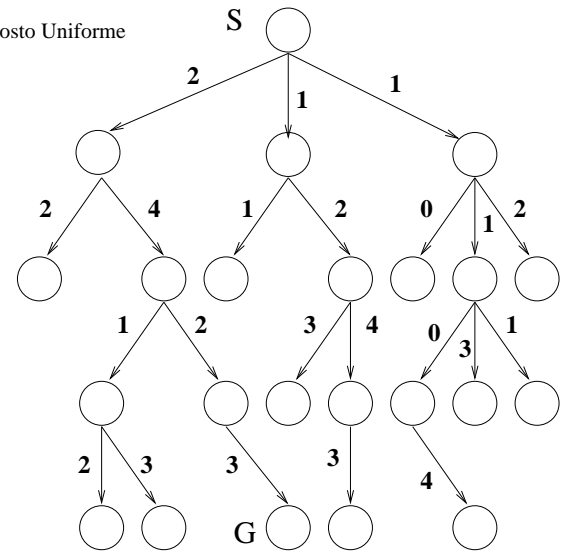
Iterative Deepening



Greedy



Costo Uniforme



Esercizio 3

Dato il seguente insieme di apprendimento

<i>Esempio</i>	<i>Target</i>	A_1 (5 val.)	A_2 (3 val.)	A_3 (4 val.)
1	+	v_1	v_1	v_2
2	-	v_1	v_3	v_2
3	-	v_3	v_1	v_1
4	-	v_2	v_2	v_2
5	-	v_2	v_2	v_1
6	+	v_3	v_1	v_2
7	-	v_4	v_3	v_2
8	+	v_2	v_3	v_4
9	+	v_5	v_2	v_4
10	+	v_5	v_3	-

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).

