

Esercizio del Corso di Sistemi per l'Elaborazione dell'Informazione

Apprendimento di Alberi di Decisione

Esercizio 1

a) dato il seguente insieme di apprendimento

<i>Esempio</i>	<i>Target</i>	<i>A</i> ₁ (5 val.)	<i>A</i> ₂ (3 val.)	<i>A</i> ₃ (4 val.)	<i>A</i> 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 $\text{GainRatio}(S,A)$) 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. (Ricordare che $\log(\frac{a}{b}) = \log(a) - \log(b)$ e $\log(a \cdot b) = \log(a) + \log(b)$; usare $\log_2(3) = 1.585$ e $\log_2(5) = 2.32$;)