

**COMPITO DI INTELLIGENZA ARTIFICIALE**  
**9 dicembre 2003**  
**Parte II – Durata 45 minuti (totale su punti 16)**

**Esercizio 1 (punti 10)**

Si consideri il training set

Sky	Air Temp	Humid	Wind	EnjoySport
Sunny	Warm	Normal	Strong	Yes
Sunny	Cold	High	Strong	Yes
Rainy	Cold	High	Strong	No
Rainy	Warm	High	Strong	No
?	Cold	Normal	Weak	No

- a) qual è l'entropia del training set rispetto all'attributo EnjoySport?
- b) si calcoli il rapporto di guadagno per i quattro attributi Sky, Air Temp, Humid e Wind.
- c) si costruisca un albero decisionale ad un solo livello per il training set dato, indicando le etichette delle foglie (numero di esempi finiti nella foglia/numero di esempi finiti nella foglia non appartenenti alla classe della foglia).
- d) dato l'albero costruito, si classifichi l'istanza

Sky	Air Temp	Humid	Wind
?	Cold	Normal	Weak

**Esercizio 2 (punti 6)**

Si descriva come viene effettuata la pianificazione automatica non lineare (Partial Order Planning). In particolare, si descriva cosa sono i causal link e quali tecniche si adottano per risolvere eventuali minacce.

## SOLUZIONE

### Esercizio 1

a)  $\text{info}(T) = -3/5 \cdot \log_2 3/5 - 2/5 \cdot \log_2 2/5 = 0,971$

b) Per calcolare il guadagno dell'attributo Sky non si usa l'entropia calcolata su tutto il training set ma solo sugli esempi che hanno sky noto:

$\text{info}(F) = -2/4 \cdot \log_2 2/4 - 2/4 \cdot \log_2 2/4 = 1$

Per gli altri attributi invece si utilizza  $\text{info}(T)$

$\text{Gain}(\text{Sky}) = 4/5 \cdot (1 - 2/4 \cdot (-2/2 \cdot \log_2 2/2 - 0/2 \cdot \log_2 0/2)) - 2/4 \cdot (-0/2 \cdot \log_2 0/2 - 2/2 \cdot \log_2 2/2) = 0,8 \cdot (1 - 0,5 \cdot 0 - 0,5 \cdot 0) = 0,8$

$\text{Splitinfo}(\text{Sky}) = -2/5 \cdot \log_2 2/5 - 2/5 \cdot \log_2 2/5 - 1/5 \cdot \log_2 1/5 = 1,522$

$\text{Gainratio}(\text{Sky}) = 0,8 / 1,522 = 0,525$

$\text{Gain}(\text{Air Temp}) = 0,971 - 3/5 \cdot (-1/3 \cdot \log_2 1/3 - 2/3 \cdot \log_2 2/3) - 2/5 \cdot (-1/2 \cdot \log_2 1/2 - 1/2 \cdot \log_2 1/2) = 0,971 - 0,6 \cdot 0,918 - 0,4 \cdot 1 = 0,971 - 0,551 - 0,4 = 0,020$

$\text{Splitinfo}(\text{Air Temp}) = -3/5 \cdot \log_2 3/5 - 2/5 \cdot \log_2 2/5 = 0,971$

$\text{Gainratio}(\text{Air Temp}) = 0,020 / 0,971 = 0,021$

$\text{Gain}(\text{Humid}) = 0,971 - 2/5 \cdot (-1/2 \cdot \log_2 1/2 - 1/2 \cdot \log_2 1/2) - 3/5 \cdot (-2/3 \cdot \log_2 2/3 - 1/3 \cdot \log_2 1/3) = 0,971 - 0,4 \cdot 1 - 0,6 \cdot 0,918 = 0,020$

$\text{Splitinfo}(\text{Humid}) = -3/5 \cdot \log_2 3/5 - 2/5 \cdot \log_2 2/5 = 0,971$

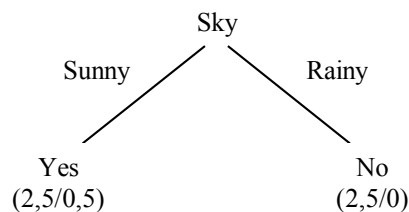
$\text{Gainratio}(\text{Humid}) = 0,020 / 0,971 = 0,021$

$\text{Gain}(\text{Wind}) = 0,971 - 4/5 \cdot (-2/4 \cdot \log_2 2/4 - 2/4 \cdot \log_2 2/4) - 1/5 \cdot (-0/1 \cdot \log_2 0/1 - 1/1 \cdot \log_2 1/1) = 0,971 - 0,8 \cdot 1 - 0,2 \cdot 0 = 0,171$

$\text{Splitinfo}(\text{Humid}) = -4/5 \cdot \log_2 4/5 - 1/5 \cdot \log_2 1/5 = 0,722$

$\text{Gainratio}(\text{Humid}) = 0,171 / 0,722 = 0,236$

c)



d) l'istanza ha l'attributo Sky sconosciuto, quindi viene suddivisa in due parti: una parte, con peso  $2,5 / (2,5 + 2,5) = 0,5$ , va nella foglia Yes e una parte, con peso  $2,5 / 5$  va nella foglia No. Quindi la classificazione risultante è Yes con probabilità  $2,5/5 \cdot 2/2,5 = 0,5 \cdot 0,8 = 0,4 = 40\%$  e No con probabilità  $2,5/5 \cdot 2,5/2,5 + 2,5/5 \cdot 0,5/2,5 = 0,5 \cdot 1 + 0,5 \cdot 0,2 = 0,6 = 60\%$ .